



Universidade do Minho
Escola de Engenharia

Gonçalo José Santiago Pereira

**Monitorização de dados de produção numa
empresa produtora de pneus, no contexto
da indústria 4.0**



Universidade do Minho

Escola de Engenharia

Gonçalo José Santiago Pereira

**Monitorização de dados de produção numa
empresa produtora de pneus, no contexto
da indústria 4.0**

Dissertação de Mestrado

Mestrado em Engenharia e Gestão Industrial

Trabalho efetuado sob a orientação de

Professor Doutor Rui Manuel Alves da Silva e Sousa

Professor Doutor Jorge Vaz de Oliveira e Sá

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



Atribuição-NãoComercial
CC BY-NC

<https://creativecommons.org/licenses/by-nc/4.0/>

AGRADECIMENTOS

Sempre acreditei que seria fácil escrever esta secção da Dissertação. Pois bem, ao refletir enquanto escrevo, apercebo-me da quantidade de pessoas que, direta ou indiretamente, de um modo mais ou menos determinante e continuado, contribuíram para eu poder chegar aqui! Do quão difícil é encontrar palavras que consigam fazer jus a tudo o que fizeram para me acompanhar.

Ao meu Orientador, Professor Rui Sousa, agradeço a confiança que depositou em mim quando aceitou ser meu orientador, ciente de que eu iria realizar uma dissertação num âmbito um tanto diverso do que seria expectável. A dedicação e o apoio que me prestou durante todo este processo foi imprescindível. Ao meu Coorientador, Professor Jorge Sá, estou grato por toda a disponibilidade que encontrou para poder partilhar comigo os conhecimentos e ferramentas técnicas, que foram inestimáveis, durante o meu Projeto.

Aos Supervisores Simão Castro e Ricardo Ferrão, quero agradecer a total colaboração que me foi prestada, todos os dias, na realização do meu trabalho. Dispor de tempo da vossa agenda, sobrecarregada, para me poderem ajudar na construção do meu saber é algo que eu reconheço imensamente. Um enorme Bem-haja.

Aos meus amigos, obrigado pela lealdade, o apoio e a sinceridade. Os desafios que encontrámos, as vitórias e derrotas que vivenciámos, acredito que fizeram de nós uma geração capaz de assegurar que o futuro estará em boas mãos!

Aos meus pais, agradeço o carinho, a educação e, acima de tudo, os sacrifícios para que eu conseguisse chegar ao “topo da montanha”. Ao meu irmão, o meu melhor amigo, agradeço-lhe por ter sido, sempre, o companheiro e a inspiração para o homem em que eu me quero tornar. À restante família, estou agradecido por todo o encorajamento e união que me foi proporcionada, durante os momentos importantes da minha vida.

Catarina, a ti, eu agradeço-te por todo o amor, ajuda e confiança que depositas em mim. Por todas os dias e noites em que me faltava a força para trabalhar e tu, com o teu toque divinalmente especial, conseguias que eu encontrasse uma réstia de energia para alcançar as minhas metas. Obrigado, por seres a razão de eu poder sentir que, aconteça o que acontecer, o ter-te conhecido, ao longo desta viagem, é alento para conseguir o que me propus conquistar.

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio, nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados, em nenhuma das etapas conducentes à sua elaboração.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

RESUMO

A presente dissertação foi desenvolvida no âmbito da Continental Mabor, uma empresa de produção de pneus, em resposta aos níveis significativos de sucata (*scrap*) gerados, diariamente. A extrusora, máquina responsável pela construção dos perfis que compõe parte do pneu e que influencia diretamente a qualidade do pneu, produz dados diariamente. A ausência de estruturação por parte destes dados leva à inexistência de monitorização e alarmísticas associadas a este segmento do processo. O objetivo principal do projeto retratado na dissertação consistiu em desenvolver um processo de tratamento dos dados gerados pela extrusora e um sistema abrangente de monitorização e controlo de qualidade para os perfis de paredes e pisos utilizados na construção dos pneus. O sistema permite ao departamento de qualidade monitorizar e analisar os indicadores de chave de desempenho (KPI) relacionados com os perfis, por forma a que se tomem decisões estratégicas baseadas em dados concretos que permitam reduzir o *scrap* e melhorar a qualidade dos produtos finais.

Para atingir esse objetivo, o trabalho foi dividido em três etapas principais. Inicialmente, foi implementado um processo de Extração, Transformação e Armazenamento (ETL) utilizando Python e o SQL. Esse processo possibilitou a extração, transformação e armazenamento de dados inicialmente não estruturados provenientes das máquinas extrusoras responsáveis pela produção dos perfis. Com a conclusão do ETL a principal causa para a ausência de um sistema eficaz de monitorização e de alarmes é suprimida. Posteriormente, um método de hierarquização de indicadores de desempenho foi implementado, eliminando o excesso de relatórios. A partir dos indicadores de maior valor desenvolveram-se relatórios interativos utilizando a ferramenta Power BI, permitindo a monitorização e análise complexa dos KPI requeridos pelo departamento de qualidade. Por fim, um sistema de alarme foi implementado que permite avisar por *email* e *Short Message Service* (SMS) os colaboradores sempre que um estado crítico relacionado com os perfis é encalçado, nomeadamente desvios significativos nos KPI monitorizados ou erros durante a execução do ETL.

Quanto a trabalhos futuros, pretende-se agendar reuniões com a empresa responsável pelo desenvolvimento das extrusoras, a fim de obter informações adicionais sobre os ficheiros. Planeia-se expandir a implementação do ETL para outras máquinas. Para além disso, pretende-se otimizar os valores a usar na configuração dos parâmetros que disparam os alertas.

PALAVRAS-CHAVE: Monitorização da produção; KPI; ETL; Alertas em tempo real; Big Data;

ABSTRACT

The following dissertation was developed at Continental Mabor, a tire manufacturing company, in response to unsatisfactory levels of daily scrap generated. The extruder, the machine responsible for constructing the tire profiles, which directly influence tire quality, produces unstructured data on a daily basis. The lack of data structuring in this process segment leads to a lack of monitoring and alarm systems. The main objective of the project depicted in this dissertation was to develop a data treatment process for the extruder-generated data and a comprehensive monitoring and quality control system for the wall and tread profiles used in tire construction. The system enables the quality department to monitor and analyze key performance indicators related to the profiles, allowing strategic decisions based on concrete data to reduce scrap and improve the quality of the final products.

To achieve this objective, the work was divided into three main stages. Initially, an Extraction, Transformation, and Loading process was implemented using Python and SQL. This process facilitated the extraction, transformation, and storage of initially unstructured data from the extruder machines responsible for profile production. With the completion of the ETL, the main cause for the absence of an effective monitoring and alarm system is addressed. Subsequently, a performance indicator hierarchization method was implemented, eliminating excessive reporting. Interactive reports were developed using the Power BI tool based on the most valuable indicators, enabling complex monitoring and analysis of the KPIs required by the quality department. Finally, an alarm system was implemented to notify employees via *email* and SMS whenever a critical state related to the profiles is detected, such as significant deviations in the monitored KPIs or errors during the ETL execution.

As for future work, meetings are planned with the company responsible for the development of the extruders to obtain additional information about the files. The expansion of the ETL implementation to other machines is also intended. Additionally, optimization of the parameter configuration values that trigger the alerts is desired.

KEYWORDS: Production monitoring; KPI; ETL; Real-time alerts; Big Data.

ÍNDICE

Agradecimentos.....	iii
Resumo.....	v
Abstract.....	vi
Índice.....	vii
Lista de Abreviaturas, Siglas e Acrónimos	xi
Índice de Figuras.....	xiii
Índice de Tabelas	xv
1. Introdução	1
1.1 Enquadramento e Motivação.....	1
1.2 Objetivos e Resultados Esperados	3
1.3 Metodologia de Investigação.....	4
1.4 Estrutura da Dissertação	5
2. Estado da Arte	7
2.1 Indústria 4.0.....	7
2.2 Big Data	8
2.2.1 Características da Big Data.....	9
2.2.1.1 5V da Big Data	10
2.2.1.2 Dados Estruturados vs Dados Semiestruturados vs Dados Não Estruturados	14
2.2.2 O Poder da Big Data	15
2.2.3 Desafios da Big Data.....	17
2.2.3.1 Volume	18
2.2.3.2 Variedade.....	19
2.2.3.3 Velocidade	20
2.2.3.4 Veracidade e Valor.....	21
2.3 Armazenamento de Dados	21
2.3.1 Base de Dados Relacional.....	21
2.3.2 Data Warehouse	23
2.3.2.1 Base de Dados Operacional vs Base de Dados Analítica	24
2.3.2.2 Extração Transformação e Armazenamento dos Dados	27

2.3.2.3	Arquitetura de uma Data Warehouse	28
2.3.2.4	Esquema de Estrela	30
2.4	Monitorização do Chão de Fábrica.....	32
2.5	Classes de Indicadores de Desempenho	34
2.5.1.1	Sistemas de Medição de Desempenho.....	35
2.5.1.2	Sistema de Monitorização de Desempenho Pessoal	36
2.6	Processo de Hierarquização Analítica	37
2.6.1	Objetivo.....	37
2.6.2	Funcionamento.....	37
3.	Contexto Empresarial da Dissertação.....	40
3.1	Continental Mabor	40
3.2	Processo Produtivo – <i>Passenger Car & Light Truck Tires</i>	41
3.3	Máquinas Extrusoras.....	42
3.3.1	Paredes e Piso.....	42
3.3.2	Extrusão	43
3.3.3	Perfilômetro.....	45
4.	Diagnóstico de Problemas e Proposta de Soluções	46
4.1	Estado Inicial	46
4.2	Desafios Associados	47
4.2.1	Dados Não Estruturados	47
4.2.2	Ineficácia e Limitações do Software de BI.....	48
4.2.3	Inexistência de Monitorização e Alarmísticas	52
4.3	Soluções Propostas.....	52
5.	Desenvolvimento do ETL	53
5.1	Extração	54
5.2	Transformação – 1ª parte	56
5.2.1	Etapa 1 – Estruturação dos Dados.....	56
5.2.1.1	Engenharia Inversa.....	56
5.2.1.2	Construção e Configuração dos Dataframes.....	60
5.2.2	Implementação do Processo de Hierarquização Analítica.....	60
5.2.3	Etapa 2 – Processamento dos Dados.....	61
5.2.3.1	Chaves dos Dataframes.....	62

5.2.3.2	KPI – Espessura.....	64
5.2.3.3	KPI – Área	67
5.2.3.4	KPI – Assimetria.....	69
5.2.3.5	KPI – Peso	71
5.2.3.6	Volume dos Dados	72
5.2.3.7	Sobreposição das Espessuras.....	73
5.2.3.8	Descentralização do ETL.....	74
5.3	Armazenamento – 1ª parte	75
5.3.1.1	Armazenamento dos Dados.....	75
5.3.1.2	Transformação e Armazenamento com ou sem Processamento	76
5.4	Transformação e Armazenamento – 2ª parte	77
5.5	Conversão do ETL num Executável.....	79
5.6	Rotina do ETL.....	79
6.	Desenvolvimento do Relatório Interativo.....	82
6.1	Seleção do <i>Software</i>	82
6.2	Relatório – Last 15 Days	82
6.2.1	Dados Importados	82
6.2.2	Arquitetura do Relatório	82
6.2.2.1	Introdução.....	83
6.2.2.2	Espessura	83
6.2.2.3	Área.....	84
6.2.2.4	Assimetria	84
6.3	Relatório – <i>Beginning Middle End</i>	85
6.3.1	Dados Importados	85
6.3.2	Arquitetura do Relatório	86
6.3.2.1	Introdução.....	86
6.3.2.2	Espessura	86
6.3.2.3	Área.....	86
6.3.2.4	Assimetria	87
7.	Desenvolvimento da Alarmística	88
7.1	Seleção dos KPI a Monitorizar	88
7.2	Comparação e Seleção do <i>Software</i>	88

8. Análise de Resultados	92
8.1 Hierarquização dos KPI.....	92
8.2 Processo de ETL.....	92
8.3 Relatório Interativos	93
8.4 Sistema de Alarmísticas.....	94
9. Conclusão.....	95
Referências Bibliográficas	97
Anexo A.....	105
Apêndice A – Esquema das Tabelas	106
Apêndice B – Relatórios em Power BI	114

LISTA DE ABREVIATURAS, SIGLAS E ACRÓNIMOS

A

- ACID** Atomicidade, Consistência, Isolamento e Durabilidade
AHP Processo de Hierarquização Analítica

B

- BI** *Business Intelligence*
BME *Beginning Middle End*

C

- CPS** *Cyber Physical Systems*
CSV *Comma Separated Values*

D

- DW** *Data Warehouse*

E

- ETL** Extração Transformação Armazenamento

G

Global Positioning System

H

- HMI** *Human Machine Interface*

I

- IA** Inteligência Artificial

K

- KPI** *Key Performance Indicators*

M

- MCDM** *Multicriteria Decision-Making*
MRP *Material Requirement Planning*

O

- OEE** *Overall Equipment Effectiveness*
- OLAP** *Online Analytical Processing*
- OLTP** *Online Transaction Processing*
- OPC-UA** *Open Platform Communications Unified Architecture*

P

- PLC** *Programmable Logic Controller*
- PLT** *Passenger Car & Light Truck Tires*
- PMS** *Performance Measurement System*

R

- RDBMS** *Relational Database Management Systems*

S

- SMS** *Short Message Service*
- SQL** *Structured Query Language*

ÍNDICE DE FIGURAS

Figura 1 – Características da Big Data. Retirado de (Chandarana & Vijayalakshmi, 2014).	11
Figura 2 – Volume de dados criados e replicados mundialmente. Adaptado de (IDC, 2021).	11
Figura 3 – Data Pipelines. Adaptado de (StreamSets, 2021).	13
Figura 4 – Valor da Big Data para as Fortune 1000. Retirado de (Bean, 2017).	14
Figura 5 – Perda de valor de negócio face à latência. Retirado de (Hackathorn, 2004).	19
Figura 6 – Base de dados relacional. Retirado de (Jukić et al., 2015).	22
Figura 7 – Propriedades de uma DW.	23
Figura 8 – Contrastes entre OLTP e OLAP. Adaptado de (Franco, 2019).	26
Figura 9 – Arquitetura minimalista de uma <i>Data Warehouse</i> . Adaptado de (Rostek, 2010).	28
Figura 10 - Arquitetura de uma DW. Retirado de: (Shi et al., 2001).	28
Figura 11 – Fluxo de informação. Retirado de (Borges, 2020).	31
Figura 12 – Exemplo de um esquema de estrela. Retirado de (Chaudhuri & Dayal, 1997).	31
Figura 13 - Termos chave de um KPI. Retirado de: (Alonso, 2022)	34
Figura 14 – Classes de KPIs. Adaptado de (Cristea & Cristea, 2021).	35
Figura 15 – Sistema de Medição de Desempenho. Retirado de (Wohlers et al., 2020).	35
Figura 16 – Monitorização de desempenho pessoal. Retirado de (Prathima et al., 2020).	37
Figura 17 - Corte de secção transversal de um pneu. Adaptado de: (Király et al., 2022)	43
Figura 18 - (a) Corte de secção transversal do piso extrudido; (b) Corte de secção transversal de um pneu. Adaptado de:(Caetano, 2019)	43
Figura 19 - Linha de co-extrusão triplex. Adaptado de: (Costa, 2022)	44
Figura 20 - Fieira utilizada para a produção de perfis de pisos. Retirado de: (Costa, 2022)	44
Figura 21 - (A) Perfil já finalizado a ser transportado por uma passadeira; (B) Perspetiva transversal de um perfil de paredes. Adaptado de: (NEXEN TIRE, 2017)	45
Figura 22 - Esboço figurativo de um perfilômetro utilizado em uma linha de co-extrusão.	45
Figura 23 - Arquitetura dos ficheiros gerados pela extrusora.	47
Figura 24 - Segmento pixelizado retirado de um ficheiro TSX real.	48
Figura 25 - Exemplos concretos da 1ª e 2ª funcionalidade do <i>software</i>	49
Figura 26 - Exemplos concreto da 1ª e 3ª funcionalidade do <i>software</i> , para um perfil de paredes.	50
Figura 27 - Exemplos concreto da 1ª e 3ª funcionalidade do <i>software</i> , para um perfil de pisos.	51
Figura 28 - Fluxograma do ETL.	53
Figura 29 - Processamento associado a cada tipo de lista.	54
Figura 30 - Exemplo de uma secção do ficheiro TSX em formato de texto bruto.	55

Figura 31 - Lista tratada da secção <i>date</i>	56
Figura 32 - Secções em que a engenharia inversa não teve sucesso.	60
Figura 33 - Número de pontos por 5 milímetros.....	65
Figura 34 - Discrepância entre a largura do perfil real e nominal.	66
Figura 35 - Adição de 0 para igualar número de pontos.	66
Figura 36 - Ilustração dos diversos trapézios.....	68
Figura 37 - Exemplos de par de pontos para um perfil nominal de um piso.	70
Figura 38 - Nomenclatura da função <i>connect</i> proveniente da livraria <i>pyodbc</i>	76
Figura 39 - Esquema estrela da base de dados.....	80
Figura 40 - Arquitetura da pasta local onde se encontra o executável.	80
Figura 41 - Dashboard desenvolvido pelo Grafana.....	89
Figura 42 - Alerta Grafana, <i>PERF_TOTAL_AREA_DIFF</i>	90
Figura 43 – Alerta IT Soft., <i>PERF_Area_TOTAL_EXE09</i>	90
Figura 44 – Alerta IT Soft., <i>PERF_Area_LEFT_EXE09</i>	91
Figura 45 – Alerta IT Soft., <i>PERF_Area_RIGHT_EXE09</i>	91
Figura 46 – Alerta IT Soft., <i>PERF_Logs_Alert_EXE09</i>	91
Figura 47 – <i>Introdução, Last 15 Days</i>	114
Figura 48 - <i>Espessura parte 1, Last 15 Days</i>	115
Figura 49 - <i>Espessura parte 2, Last 15 Days</i>	116
Figura 50 - <i>Espessura parte 3, Last 15 Days</i>	117
Figura 51 - <i>Área, Last 15 Days</i>	118
Figura 52 - <i>Assimetria, Last 15 Days</i>	119
Figura 53 - <i>Introdução, BME</i>	120
Figura 54 – <i>Espessura parte 1, BME</i>	121
Figura 55 - <i>Espessura parte 2, BME</i>	122
Figura 56 - <i>Área, BME</i>	123
Figura 57 - <i>Assimetria, BME</i>	124

ÍNDICE DE TABELAS

Tabela 1 – Objetivos da presente dissertação	4
Tabela 2 – Operações de Controlo da Qualidade e Desempenho. Informação de (Bosi et al., 2019; Godinho & Piedade, 2022; Jukić et al., 2015; Prathima et al., 2020; Victor, 2018; Wohlers et al., 2017, 2020)	16
Tabela 3 – Operações de <i>Data Mining</i> . Informação de (Mishra, 2017; Nagorny et al., 2017; Rostek, 2010).....	17
Tabela 4 – Operações de <i>Machine Learning</i> . Informação de (Davies et al., 2007; Teti et al., 2022)	18
Tabela 5 – Base de Dados Operacional vs Base de Dados Analítica.....	27
Tabela 6 - Soluções implementadas para cada problema identificado no diagnóstico inicial.	52
Tabela 7 - Leque inicial de KPI em estudo no AHP.	61
Tabela 8 - Leque de KPI selecionados após o AHP.	61
Tabela 9 - Parâmetros configuráveis do ETL.....	64
Tabela 10 - Comparação entre ETL com e sem processamento.	77
Tabela 11 - Características principais do ETL.....	81
Tabela 12 - KPI e filtros associados à página Espessura.....	83
Tabela 13 - KPI e filtros associados à página Área.....	84
Tabela 14 - KPI e filtros associados à página Assimetria.....	85
Tabela 15 - KPI e filtros associados à página Espessura.....	86
Tabela 16 - KPI e filtros associados à página Área.....	87
Tabela 17 - KPI e filtros associados à página Assimetria.....	87
Tabela 18 - Resultados obtidos a partir da implementação de relatórios usando o Power Bi.....	94
Tabela 19 - Escala Valorativa do AHP. Retirado de: (Cristea & Cristea, 2021)	105
Tabela 20 - Esquema da tabela <i>Header</i>	106
Tabela 21 - Esquema da tabela <i>Contour</i>	108
Tabela 22 - Esquema da tabela <i>Date</i>	109
Tabela 23 - Esquema da tabela <i>Dimensions</i>	109
Tabela 24 - Esquema da tabela <i>Nominal</i>	110
Tabela 25 - Esquema da tabela <i>Section</i>	111
Tabela 26 - Esquema da tabela <i>Overlay</i>	111
Tabela 27 - Esquema da tabela <i>Assimetry_Differences</i>	112
Tabela 28 - Esquema da tabela <i>Assimetry_Differences_BME</i>	113

1. INTRODUÇÃO

1.1 Enquadramento e Motivação

Ao contrário da Indústria 3.0, que se baseava principalmente na automação e na introdução de sistemas computacionais nas fábricas, a Indústria 4.0 leva a automação a um novo nível, incorporando tecnologias como a inteligência artificial, computação em nuvem e análise de dados avançada. A Indústria 4.0 é uma nova realidade que está a influenciar todo o meio industrial. Os métodos e tecnologias que a suportam são complexos, porém o seu objetivo é simples: reduzir o *gap* entre os sistemas físicos e digitais de uma empresa de modo a não só minorar a intervenção humana como também alcançar uma maior eficiência, produtividade e competitividade (Prathima et al., 2020).

Acoplada à modernização da indústria, a quantidade e tipo de dados gerados aumenta drasticamente devido à implementação de *Cyber Physical Systems* (CPS), considerada por muitos o ponto de partida da transição para a Indústria 4.0 (Prathima et al., 2020; Stock & Seliger, 2016). Os CPS permitem monitorizar os ativos físicos de uma empresa e os seus processos, expondo num formato digital o seu sistema físico. A monitorização é possível devido aos sensores incluídos nas máquinas ou infraestruturas das células em chão de fábrica. Os vários sensores captam a informação e conectam-se, a partir de servidores digitais, com a base de dados centralizada da empresa; a *cloud*, onde é armazenada essa informação. *A posteriori*, os dados serão normalizados e analisados, permitindo extrair conhecimento pertinente ao negócio, garantindo valor à empresa (Prathima et al., 2020).

Os *Performance Measurement Systems* (PMS), suportados pelos sistemas de sensorização, permitem medir e controlar todos os indicadores de desempenho relevantes, promovendo a proatividade (Wohlers et al., 2020). Na indústria podem realizar-se dois tipos de medições: medição direta e medição indireta. Na medição direta, a medida de interesse é registada diretamente a partir de um sensor, diminuindo a incerteza do valor captado. No entanto, a sua implementação está dependente da instalação de um sensor compatível com o processo, tornando esta medição mais complexa a nível da instalação. Na medição indireta, são utilizadas medidas externas à medida que se pretende determinar. As medidas de interesse são adquiridas a partir de correlações empíricas entre as medidas externas. Este método de mensuração é menos preciso, contudo, é bastante exequível em contexto industrial (Teti et al., 2022).

A digitalização dos processos em chão de fábrica, impulsionada pelos métodos e tecnologias de monitorização, permite gerar um grande volume de dados heterogéneo, com diferentes focos e objetivos,

relacionados com cada processo e ativo do sistema, como as máquinas de manufatura ou os próprios colaboradores. A este aglomerado de dados dá-se o nome de Big Data. A Big Data despoleta novas oportunidades estratégicas e operacionais, como, por exemplo, a monitorização em tempo real dos ativos ou a implementação de controlo da qualidade, que garantem à empresa vantagem competitiva no mercado. Consegue-se, assim, alcançar um dos objetivos mais desejados pela indústria, garantir que os possíveis problemas da empresa se tornem visíveis, desencadeando ações preventivas.

Todavia, a Indústria 4.0 não trouxe apenas vantagens para o mundo empresarial, trouxe também bastantes obstáculos (Prathima et al., 2020; Teti et al., 2022; Wohlers et al., 2020). Um dos principais problemas originados pela Indústria 4.0, consiste na dificuldade em identificar e extrair valor dos dados fornecidos em chão de fábrica (Jukic et al., 2015). Esta dificuldade é causada, principalmente, por dois fatores:

1. A heterogeneidade entre a sintaxe dos diversos dados – a sintaxe consiste na representação estrutural dos dados, estrutura esta que difere consoante a fonte de onde os dados são oriundos. Por exemplo, dados originados a partir de sensores de força de corte são estruturados porque podem ser armazenados num formato tabular. Por outro lado, dados fornecidos a partir de câmaras de alta velocidade, como vídeos e imagens, são dados não estruturados, porque não possuem um modelo comum de armazenamento. A diversidade na sintaxe requer a realização de diferentes abordagens ao analisar o conjunto de dados, é disso exemplo o uso de um método analítico para analisar o estado da máquina a partir de dados que refletem a vibração dos componentes, enquanto para realizar controlo de qualidade se utiliza um método distinto que se foca no estudo do rácio entre produtos conformes e não conformes produzidos diariamente (Bontempo & Zagelow, 1998; Chandarana & Vijayalakshmi, 2014; Teti et al., 2022).
2. O grande volume e complexidade de dados fornecidos diariamente – este par de fatores leva a que seja mais exigente para a empresa, não só a sua limpeza e processamento (visto que cada classe de dados possui um tratamento diferente), como também o acesso aos dados já processados com o intuito de os analisar. Esta exigência é um entrave ao sucesso da empresa, já que, para os dados assumirem algum valor eles têm de ser identificáveis, acessíveis e aplicáveis, o que não acontece na ausência de processos de normalização de dados (Bontempo & Zagelow, 1998; Chandarana & Vijayalakshmi, 2014; Jukic et al., 2015).

Outro desafio comum no mundo empresarial consiste na incapacidade, por parte das empresas, de conseguir gerir e analisar regularmente toda a informação produzida, processada e armazenada internamente. Existe esta ideia de que quanto maior a quantidade de informação armazenada, maior a vantagem competitiva e, portanto, as empresas optam por gerar o máximo de dados possíveis. Porém, apesar da informação estar presente nas empresas, esta, geralmente não é completamente organizada ou entendida. As empresas acabam por não saber que informação possuem, onde e como está armazenada, qual a sua utilidade e como deve ser analisada para gerar valor (Ferraioli & Burke). A Continental Mabor não é exceção, com a digitalização das operações gera-se continuamente um grande volume de dados, que posteriormente é processado originando inúmeras classes de KPI, usadas para criar vários relatórios, com objetivos distintos. Não obstante, tais indicadores de desempenho não possuem uma hierarquização, sendo a valoração de cada um igual aos demais, logo, todos merecem a mesma atenção. A inexistência de priorização e o grande volume de KPI leva a que todos sejam utilizados para gerar relatórios, despoletando um grande problema, um excesso de relatórios na empresa. Este excesso impossibilita a análise de cada um, regularmente, tornando-se a gestão e análise destes impraticável (Cristea & Cristea, 2021).

Esta dissertação, enquadra-se no âmbito de um estágio curricular, numa colaboração entre a Continental Mabor, empresa da indústria de pneus do grupo *Continental Aktiengesellschaft GmbH*, e a Universidade do Minho. O estágio curricular assenta no desenvolvimento de um projeto interno à empresa, denominado *Business Intelligence Manufacturing Implementation*.

1.2 Objetivos e Resultados Esperados

Antes de se enumerar qualquer um dos objetivos concretos da dissertação é relevante entender o objetivo global: impulsionar a extração de conhecimento a partir dos dados gerados diariamente em chão de fábrica na Continental Mabor.

Face aos objetivos concretos desta dissertação, estes subdividem-se pelos dois problemas anteriormente citados no subcapítulo 1.1. Face ao primeiro problema, o objetivo é simples, desenvolver e implementar uma rotina digital de tratamento de dados que permita extrair, normalizar e armazenar os dados gerados em chão de fábrica (Inmon, 1992; Jukic et al., 2015).

Relativamente ao excesso de relatórios, pretende-se desenvolver um sistema de alarmes para as máquinas de produção, que permita aos operadores configurar os alarmes e, conseqüentemente, serem notificados quando algum sistema em chão de fábrica está prestes a alcançar um estado crítico. Ainda face a este problema, pretende-se implementar a metodologia Processo de Hierarquização Analítica

(AHP) que visa criar uma ordem hierárquica, que exprima a priorização dos KPI existentes face às opiniões e objetivos dos colaboradores. Realizada a hierarquização, criar-se-ão relatórios com os KPI do topo da hierarquia. Deste modo, existe um *cluster* de relatórios com grande valor para a empresa, que será regularmente analisado (Cristea & Cristea, 2021).

O objetivo global do projeto é, por conseguinte, subdividido em quatro tarefas, representadas na Tabela 1. Pretende-se com a conclusão destas tarefas garantir um aumento significativo na eficácia do modo de monitorização dos processos produtivos e um aumento da eficiência das metodologias de BI.

O tema da dissertação enquadra-se no projeto a desenvolver em contexto empresarial e será desenvolvido em paralelo com a dissertação, ao longo do estágio curricular.

1.3 Metodologia de Investigação

A metodologia de investigação a utilizar para a concretização da presente dissertação é a *Action-Research*, método em que as pessoas envolvidas no problema colaboram, em conjunto, de forma ativa na busca de uma solução, juntamente com o apoio do(s) investigador(es) (Ventura, 2012). Como a dissertação será realizada em paralelo com a participação num projeto real em contexto empresarial e baseada no mesmo, não faria muito sentido adotar o objetivismo, metodologia em que o investigador apenas atua como espectador. O investigador, neste caso, como colaborador da empresa terá de participar na busca ativa da solução, cimentando a frase "*learning by doing*" com que (O'Brien, 1998) define a metodologia. Este método de investigação tem por base 5 fases: o diagnóstico, o planeamento de ações, a implementação de ações, a avaliação das ações e a especificação da aprendizagem (Susman, 1978). A sequência destas fases representa um processo iterativo e cíclico.

A metodologia *Action-Research* possui dois objetivos distintos, divididos pela área conceptual com que se relacionam (Ventura, 2012):

Tabela 1 – Objetivos da presente dissertação

Tarefas	Objetivos
Desenvolver e Implementar uma rotina digital de tratamento de dados	Reduzir a quantidade de dados não processados
	Normalizar a semântica de vários clusters de dados
Desenvolver um sistema de alarmes para as máquinas de produção	Prevenir que os sistemas e células em chão de fábrica atinjam algum estado crítico
Hierarquizar os KPI	Eliminar o excesso de reports
	Permitir a análise regular dos KPI mais importantes
	Reduzir o tempo alocado à análise de dados
Desenvolver relatórios iterativos	Permitir uma análise interativa e visual dos KPI.

1. Praticabilidade: o investigador tem a responsabilidade de contribuir para a resolução do problema em estudo, a partir de uma cooperação contínua com a equipa responsável;
2. Conhecimento: como investigador possui um papel ativo num projeto prático de resolução de problemas. Independentemente da área, será possível com maior facilidade e, até mesmo qualidade, obter conhecimento através dessa experiência, podendo partilhá-lo com a comunidade científica.

Por forma a assegurar a qualidade da dissertação, os objetivos da mesma têm de ser delineados *a priori*, desde o começo da investigação, mesmo sendo necessário alterá-los com o decorrer do projeto.

1.4 Estrutura da Dissertação

Neste capítulo introdutório, são apresentados os elementos essenciais da dissertação. Inicia-se com o enquadramento e a motivação que conduziram ao desenvolvimento deste trabalho. De seguida, são traçados os objetivos e os resultados esperados, delineando o propósito e a direção da pesquisa. A metodologia de investigação adotada também é discutida, abordando as técnicas utilizadas para realizar o estudo. O capítulo seguinte apresenta o estado da arte, fornecendo uma visão geral dos conceitos relevantes e de pesquisas anteriormente desenvolvidas relacionadas com o tema. São abordados tópicos como Indústria 4.0, Big Data e armazenamento de dados, além da monitorização do chão de fábrica e indicadores de desempenho.

No capítulo três, é apresentado o contexto empresarial da dissertação, com enfoque na Continental Mabor. São objeto de discussão o processo produtivo de pneus, a importância das extrusoras, bem como os perfis de paredes e pisos utilizados. Também é abordado o perfilômetro e sua relevância para a monitorização dos dados presentes em chão de fábrica. No capítulo seguinte é realizado o diagnóstico do estado inicial do problema em estudo. São identificados os desafios associados; a presença de dados não estruturados, a ineficácia do *software* de *Business Intelligence* (BI) que analisa estes dados e a ausência de qualquer monitorização ou alarmística. Além disso, são apresentadas as soluções propostas para lidar com esses desafios.

Os capítulos cinco e seis detalham o desenvolvimento do processo de Extração, Transformação e Armazenamento (ETL) e a criação dos relatórios relevantes, utilizando a ferramenta Power BI, respetivamente. No capítulo sete, é apresentado o desenvolvimento do sistema de alarmística, incluindo a seleção dos KPI a serem monitorizados e a comparação de diferentes *softwares* disponíveis.

Os resultados obtidos são discutidos no capítulo oito, seguidos da conclusão da dissertação, no capítulo nove. Por fim, são apresentadas as referências bibliográficas e, um anexo e dois apêndices que contêm informações de relevo face ao conteúdo da dissertação.

2. ESTADO DA ARTE

2.1 Indústria 4.0

Na última década é reconhecido por toda a comunidade científica que os sistemas de manufatura foram impulsionados a partir da implementação de melhorias tecnológicas na sua arquitetura (Nakayama et al., 2020). Os investigadores defendem que tais melhorias, resultantes da integração de vários sistemas tecnológicos digitais, independentes uns dos outros, despoletaram a Quarta Revolução Industrial. Esta revolução, composta pela transformação tecnológica de vários setores da indústria tais como: sistemas de produção; logística; modelos de BI, resultaram no princípio de uma nova era industrial, denominada Indústria 4.0 (Oliveira et al., 2022).

A Indústria 4.0 reduz o *gap* entre os sistemas físicos e digitais de uma empresa, alavancando a automação de maneira a reduzir a intervenção humana (Prathima et al., 2020). A automação, impulsionada pela implementação de tecnologias digitais, promove mudanças:

1. Nas fases de desenvolvimento – a mão de obra outrora essencial na maioria da extensão temporal do processo produtivo, vê-se agora substituída pela capacidade automática da máquina produtiva ao realizar as mesmas operações, com menor tempo de ciclo, maior qualidade e menor interdependência face a outras operações;
2. Na flexibilidade na produção – a capacidade das máquinas se adaptarem rapidamente ao tipo de produto a desenvolver, permite reduzir tempos de *setup* e, consequentemente, *work in progress* e tempos de espera;
3. Na eficiência no consumo de recursos – a digitalização e monitorização dos processos face aos vários tipos de produtos, permite registar as suas métricas de desempenho possibilitando realizar, por exemplo, o *Material Requirement Planning* (MRP), capaz de, em segundos, delinear o planeamento da produção para uma semana;
4. Na digitalização e monitorização da organização – a implementação de CPS, sistemas a explicar *a posteriori*, resulta num drástico aumento de dados gerados. Este procedimento permite à empresa aceder a conhecimento, anteriormente inacessível, a partir da implementação de métodos distintos, nomeadamente, a monitorização em tempo real dos ativos da empresa ou a implementação de métodos de *Data Mining*.

Estas mudanças culminam no nascimento das *Smart Factories*, fábricas que integram esta nova metodologia de produção, que caracteriza os *Intelligent Manufacturing Systems*, um sistema produtivo

mais flexível, eficiente e digitalizado do que os sistemas produtivos tradicionais (Hozdić, 2015; Shi et al., 2020; Wang et al., 2016).

A introdução de CPS na indústria é considerada por muitos o ponto de partida da transição para a Indústria 4.0 (Prathima et al., 2020; Stock & Seliger, 2016). Os CPS são sistemas capazes de monitorizar os sistemas físicos de uma empresa e respetivos processos, com o propósito de, a partir da informação originada por estes, elaborar uma cópia digital detalhada do sistema físico da organização. Este espelhamento computacional contém um grande volume de dados e face à sua representação permite retratar todo um sistema. Existem vários *clusters* de dados com diferentes objetivos e significados. Apesar do volume e variedade de dados, muitos destes acabam por ser inúteis à empresa, quer pelo objetivo que representam não possuir relevância para o negócio atualmente, quer pelo facto de os dados não estarem processados. A ausência de uma estrutura organizacional por parte destes reflete-se pelo facto de não terem sido processados por qualquer tipo de algoritmo, o que lhes garantirá um valor estrutural capaz de ser analisado *a posteriori*, onde se gerará informação útil ao negócio da empresa (Prathima et al., 2020). Este conjunto de dados estruturados permitirá aceder a dados previamente “invisíveis” e analisá-los, permitindo à empresa responder diariamente a perguntas a que anteriormente não conseguiria responder, tomando, assim, decisões eficientes e eficazes com base no conhecimento que se extrai dos dados. É este conceito de tomada de decisões, suportada em informação proveniente de um conjunto de dados, que garante às empresas uma vantagem competitiva no mercado (Alcacer & Cruz-Machado, 2019; Mourtzis et al., 2018; Prathima et al., 2020).

Qualquer empresa que anseia tornar-se numa *Smart Factory* precisa de gerar continuamente dados. É um requisito fundamental devido às melhorias que garante à indústria o baixo custo de extração e a aparente inesgotabilidade, ou o valor que gera aos vários departamentos a partir da análise de dados (Klingenberg et al., 2021). Shi et al. (2020) afirmam que, para além de outros requisitos, o estatuto de *Smart Factory* é alcançado quando soluções de problemas visíveis à empresa são conseguidos com o uso de inteligência artificial (IA). Portanto, facilmente se entende o quão importante é desenvolver internamente um paradigma orientado a dados no contexto da Indústria 4.0.

2.2 Big Data

Na última década as empresas têm sido obrigadas a lidar com uma grande quantidade e complexidade de dados, muito devido à Indústria 4.0. Para além dos sistemas de armazenamento tradicionais, como as bases de dados operacionais ou *Data Warehouses* (DW), as empresas veem hoje a necessidade de se envolver em novos objetivos e ambições relacionadas com a análise e

processamento da tão aclamada Big Data. Inúmeros estudos suportam este movimento demonstrando que esta nova fonte de dados, quando bem processada e analisada, é preciosa, face a toda a nova informação que gera e a todos os novos conhecimentos, úteis às decisões de negócio a realizar no dia a dia de uma organização (Jukic et al., 2015).

Big Data é um dos termos mais usados e relatados pela comunidade científica ao longo dos últimos dez anos. Aos olhos de muitas pessoas, o mundo está a transformar-se, digitalmente, a partir da explosão de dados impulsionada pelo desenvolvimento de cada vez mais tipos de oportunidades de captura dos mesmos, distintas umas das outras (Höller, 2014). Todas estas novas fontes de informação do século XXI, como os *smartphones*, os *websites* de compras, as redes sociais ou os *Global Positioning System* (GPS), promovem, a partir do seu simples funcionamento operacional, a produção de um grande volume de dados (McAfee & Brynjolfsson, 2012). É esta “bola de neve” de criação de dados que incentiva todas as organizações, academias científicas, consultores e investigadores a colocar o seu foco e tempo no desenvolvimento de metodologias e tecnologias de processamento e análise de Big Data (Goes, 2014).

2.2.1 Caraterísticas da Big Data

O termo Big Data identifica um tipo de *dataset* específico, constituído maioritariamente por dados não estruturados (conceito explicado na secção 2.2.1.2) que se encontram presentes em aplicações específicas computacionais dos dias de hoje, desde a análise de dados mais simples à técnica mais complexa de *machine learning* (Cuzzocrea et al., 2013).

Os dados que compõem estes Big Data *datasets* podem ser bastante distintos, quer na sua estrutura quer no seu objetivo. Porém, para serem denominados de Big Data têm de possuir determinadas caraterísticas (Cuzzocrea et al., 2011):

1. Grande escala de dados – cada um dos repositórios que acolhe um certo *cluster* de dados possui uma escala significativamente volumosa e heterogénea de dados;
2. Problemas de escalabilidade – as aplicações computacionais que usufruem dos dados têm grandes dificuldades em suportar eficaz e eficientemente o processamento das respetivas tarefas computacionais, devido ao volume de dados em análise que aumenta continuamente face ao fluxo contínuo de *inputs* de dados gerados;
3. Suportam processos de Extração, Transformação e Armazenamento – face à heterogeneidade tanto das fontes dos dados, como da sua própria natureza, é vital o uso de processos que permitam extrair os dados dessas diversas fontes, transformá-los, de modo a assegurar a sua normalização e, posteriormente, armazená-los.

4. Permitir o desenvolvimento de modelos analíticos simples e interpretáveis – o principal objetivo da Big Data é assegurar que a partir da análise dos dados qualquer organização consiga obter informação valiosa para o seu modelo de negócio.

Jukic et al. (2016), de uma maneira mais simplista, definem Big Data como *datasets* que possuem um enorme volume de dados diversificados, que se criam a um ritmo acelerado e que não são modelados formalmente. Isto é, não existe uma estrutura de dados predefinida que todos necessitam de respeitar. Krish Krishnan (2013) complementa a definição anterior afirmando que: face à velocidade a que são criados os dados na fonte, aos diferentes níveis de complexidade e à possível ambiguidade dos dados, originada pela heterogeneidade dos mesmos, esses não são passíveis de ser processados e analisados utilizando *legacy systems*, sistemas constituídos por tecnologias e metodologias tradicionais de processamento de dados.

A literatura científica, suportada pelos diversos investigadores, defende que a Big Data pode ser descrita por 5V: volume, variedade, velocidade, veracidade e valor (Biehn, 2015; Doug, 2001; Ellars, 2013; Goes, 2014; Jukic et al., 2015).

2.2.1.1 5V da Big Data

Os 5V: volume; velocidade; variedade; veracidade; valor, podem ser interpretados como umas “lentes pelas quais nós podemos olhar de forma a entender a natureza da Big Data e a plataforma disponível onde esta pode ser explorada” (Chandarana & Vijayalakshmi, 2014). Na Figura 1 encontram-se representados os 5V, as suas interdependências e parte das propriedades de cada um.

Volume

Esta dimensão caracteriza a vasta quantidade de dados presente no *dataset* utilizado nas várias aplicações da Big Data. Oliveira et al. (2022) defendem que para uma determinada base de dados ser considerada Big Data precisa que a medida que reflete o seu volume se compreenda entre a dimensão dos N *petabytes* e dos N *exabytes*. Ainda assim, não existe um valor ótimo estabelecido que seja representativo do peso que a base de dados deva tomar. Com o passar dos meses, o volume de dados presente no mercado industrial aumenta a um rácio de 1000 *Exabytes* por ano. Face ao desenvolvimento tecnológico empresarial impulsionado pela Indústria 4.0, a capacidade por parte das empresas de

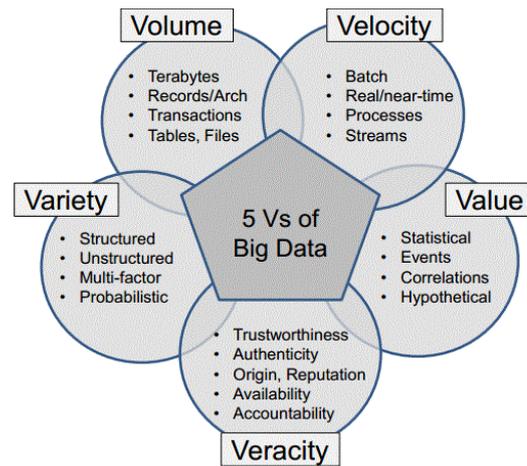


Figura 1 – Características da *Big Data*. Retirado de (Chandarana & Vijayalakshmi, 2014).

conseguir monitorizar e digitalizar todos os seus recursos e processos leva a que exista um crescimento considerável do volume de dados a circular no sistema da empresa (Wan et al., 2015). Kagermann et al. (2013) afirmam mesmo que “a Big Data nos dias de hoje é incapaz de ser processada e gerida num só dispositivo computacional. É necessário, portanto, uma arquitetura distribuída para um Intelligence Manufacturing System”.

No gráfico presente na Figura 2 encontra-se representado o crescimento exponencial da quantidade de dados criada e replicada à escala global, ao longo de uma década.

Em 2020 a International Data Corporation previu que serão gerados 181 *zettabytes* em 2025, experienciando um aumento de 182% face ao ano de 2020.

Velocidade

A velocidade representa dois módulos distintos da Big Data: o ritmo com que os dados são gerados a partir da fonte; a rapidez com que o processamento dos dados ocorre (Fatima & Wasnik,

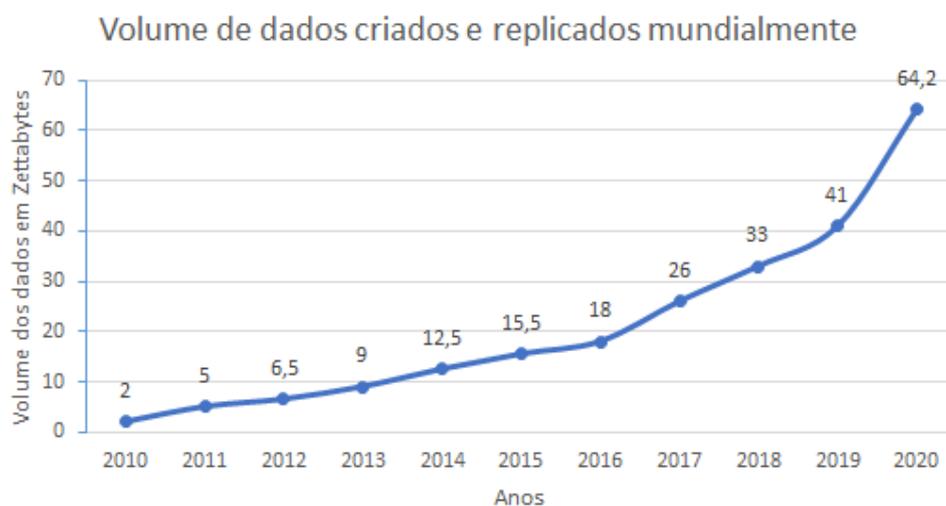


Figura 2 – Volume de dados criados e replicados mundialmente. Adaptado de (IDC, 2020).

2016). A digitalização dos processos e sistemas veio alterar completamente o ritmo com que os dados são criados e armazenados. Em sistemas tradicionais os dados são gerados em lotes a partir de uma certa operação. Por exemplo, o colaborador de uma fábrica no fim de um turno de trabalho analisa a quantidade de produtos produzidos sem defeito numa determinada célula do chão de fábrica e regista, manualmente, esse número numa base de dados. Por conseguinte, estes dados ao serem criados e armazenados, manualmente, em grandes quantidades e de uma só vez, levam a que a velocidade do processamento seja lenta e irregular. Atualmente, em sistemas Big Data, o processo é completamente diferente, os procedimentos em chão de fábrica são monitorizados em tempo real, a partir da criação contínua e interativa dos dados, proporcionada por sensores eletrónicos. O seu processamento e armazenamento é contínuo (*streaming*) ou em lotes, recorrendo ao uso de uma *data pipeline*, i.e., uma arquitetura digital que permite transferir os dados desde um sistema para outro, podendo, ainda, suportar a extração dos dados de uma fonte, o seu processamento e armazenamento, como é o caso de uma *pipeline* ETL, que permite tratar os dados para que estes possam ser analisados facilmente *a posteriori*. São estas novas metodologias que permitem assegurar uma velocidade constante e rápida, tanto ao nível do processamento como da criação de dados (Bauer et al., 2015; Fatima & Wasnik, 2016; Jukic et al., 2015; Oliveira et al., 2022; Rubmann & Lorenz, 2015; Watson et al., 2001).

A arquitetura dos *data pipelines* pode divergir consoante o seu objetivo, como é demonstrado pela Figura 3, em que num só fluxo de informação são usados três tipos de *Data Pipelines: Streaming; Batch; e ETL*.

Veracidade

A presente dimensão remete para os vários problemas associados à qualidade dos dados gerados que prejudicam a sua fiabilidade (Ellars, 2013; Jukic et al., 2015; Oliveira et al., 2022). Este problema pode dividir-se consoante a sua natureza (Han et al., 2011):

1. Subjetividade – a mesma família de dados pode representar diferentes objetivos para diferentes departamentos da empresa. Exemplificando, as representações gráficas da amplitude das vibrações dos componentes de uma máquina são úteis ao departamento de qualidade, face à capacidade desta métrica em representar o estado de uma máquina. Contudo, não é útil ao departamento do planeamento da produção.
2. Fraude – pode existir irresponsabilidade accidental, no ato operativo que gera os dados, levando a que sejam provocados equívocos.

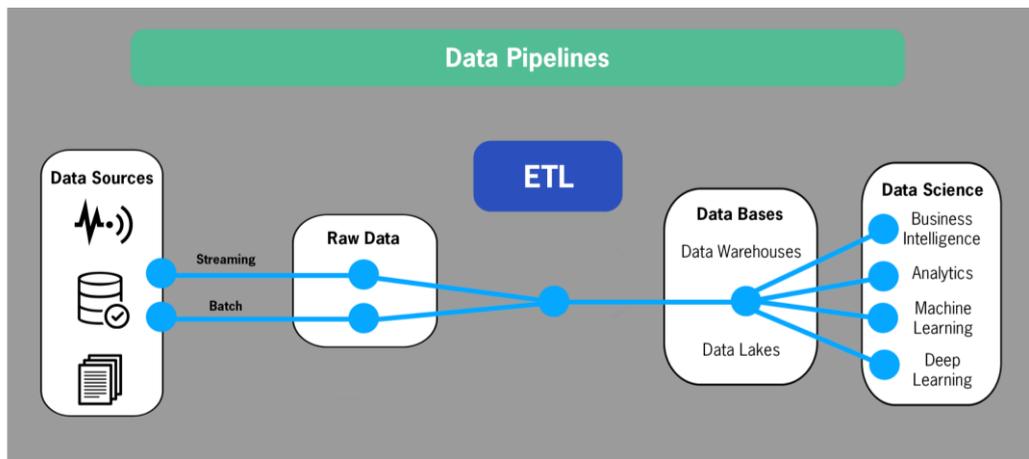


Figura 3 – Data Pipelines. Adaptado de (StreamSets, 2021).

3. Integridade – face à complexidade e *high-maintenance* das fontes de informação em chão de fábrica, é expectável que, na presença de alguma inconsistência laboral por parte dos operadores, a integridade dos dados possa estar em risco.

Todas as características da Big Data, tal como a sua heterogeneidade estrutural, a sua complexidade ou até a sua subjetividade, levam a que as métricas tradicionais da qualidade de dados, como a precisão, a conformidade e a consistência, sejam naturalmente inferiores, comparativamente a um sistema de armazenamento tradicional, tabelado e modelado, onde a sua estrutura tabular é homogênea entre os vários *datasets* (Jukic et al., 2015).

Valor

Extrair informação a partir dos dados gerados internamente, recorrendo a técnicas de análise de dados, como o *deep learning* que permite construir modelos preditivos ou métodos de *unsupervised learning*, capazes de formar *clusters* de dados, possibilita à empresa obter todo um novo conhecimento do que se passa internamente nos seus processos. O conhecimento é poder, na medida em que permite à organização tomar decisões com base em algo concreto e explícito, em vez de tomar as decisões na convicção de que se obterá um bom resultado. É aqui que surge a dimensão valor, a capacidade que a empresa tem de, a partir da análise de dados vencer no mercado pela implementação de decisões táticas e estratégicas suportadas pela Big Data. No setor industrial estas vitórias refletem-se, principalmente, na redução dos custos alocados aos processos de conceção da diversa gama de produtos de uma empresa (Oliveira et al., 2022; Han et al., 2011; Rubmann & Lorenz, 2015).

A Figura 4 foi retirada de uma publicação na *Harvard Business Review* que demonstra quais foram os tipos de projetos que as Fortune 1000 selecionaram como projetos pioneiros nos quais se recorre à Big Data como recurso diário. A figura indica, também, a percentagem de empresas que reconheceram o valor da Big Data em cada um desses tipos de projeto.

How Fortune 1000 Executives Report Using Big Data

The projects they've started, and where they're finding value.

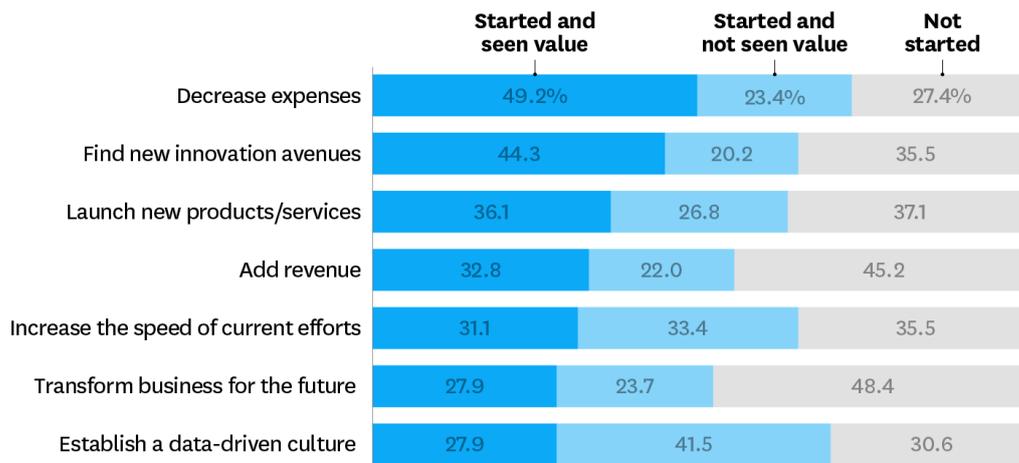


Figura 4 – Valor da Big Data para as Fortune 1000. Retirado de (Bean, 2017).

Variedade

A variedade é um dos V mais importantes uma vez que reflete a grande heterogeneidade dos dados, tanto na sua estrutura como no seu objetivo (Chandarana & Vijayalakshmi, 2014; Jukic et al., 2015). A diversidade de tipo de dados encontra-se diretamente relacionada com o tipo da fonte de onde estes são gerados, que diverge desde o uso de dispositivos inteligentes, como sensores de temperatura ou vibração, até ao uso de *web-logs*, páginas *web* onde é possível armazenar conteúdo de qualquer parte do espectro das áreas do conhecimento (Franks & Davenport, 2012; Jukic et al., 2016).

Esta classe é importante, pois representa tanto uma causa que garante enormes benefícios, nomeadamente, a capacidade de realizar diferentes métodos de *machine learning*, *i.e.*, modelos preditivos que nos permitem prever a procura de um cliente, como também, uma causa que impulsiona imensos desafios, entre os quais a necessidade de construir diferentes arquiteturas de processamento de dados, uma para cada tipo de estrutura de dados.

Face à sua estrutura, os dados podem ser subdivididos em diferentes classes seguindo uma certa perspectiva, a desenvolver no próximo subcapítulo.

2.2.1.2 Dados Estruturados vs Dados Semiestruturados vs Dados Não Estruturados

As bases de dados que armazenam Big Data não possuem um formato ou estrutura tipo, como um ficheiro *Comma Separated Values* (CSV) que armazena dados em formato tabular ou um ficheiro MP4 que armazena vídeos. Devido à diversidade de estruturas que os dados Big Data podem adotar estes subdividem-se em 3 tipos: estruturados, semiestruturados e não estruturados (Ali et al., 2019).

Dados Estruturados

Este tipo de dados respeita uma estrutura específica, previamente definida, como as bases de dados tabulares rigidamente estruturadas, em que a tabela é subdividida em vários atributos, cada um representado numa coluna em que os limites dos mesmos são bem definidos, quer o atributo seja numérico ou categórico (Chandarana & Vijayalakshmi, 2014; Fujishima et al., 2016; Moniruzzaman & Hossain, 2013; Oliveira et al., 2022).

Dados Não Estruturados

Este tipo de dados pode adotar todo um espectro de representações dentro de uma certa classe, face à inexistência de uma estrutura predefinida. Tais dados são armazenados no seu formato nativo até, mais tarde, serem processados para que se atinja uma normalização que os permita analisar. Como exemplo destes dados temos os documentos de texto, as imagens e os vídeos (Chandarana & Vijayalakshmi, 2014; Jukic et al., 2015; Lee, 2008; Moniruzzaman & Hossain, 2013; Oliveira et al., 2022).

Dados Semiestruturados

Esta classe de dados é armazenada em documentos compostos por um *template standard*, em que apesar da sua normalização existe uma porção de dados não estruturada. Tem-se como exemplo os *emails* presentes em qualquer contexto corporativo todos os dias e que respeitam um formato composto por secções que permite distinguir: o sujeito que enviou o *email*; o(s) sujeito(s) que recebeu o *email*; o assunto do *email*; corpo de texto do *email*. Ainda assim, o conteúdo do corpo do texto representa um conjunto de dados não estruturados, visto não possuir qualquer estrutura predefinida (Jukic et al., 2015; Lee, 2008; Oliveira et al., 2022).

2.2.2 O Poder da Big Data

Como referido anteriormente, todos os benefícios que a Big Data desencadeia numa empresa, resultam no aumento significativo da vantagem competitiva que a mesma possui comparativamente ao resto do mercado. Este *status* é obtido a partir do novo conhecimento que se extrai da Big Data, que permite tomar decisões estratégicas com base em algo concreto e explícito.

Este novo conhecimento é obtido a partir do *Knowledge Discovering Process*, um processo de análise complexa de dados, em que o utilizador identifica padrões, correlações e tendências com um certo potencial. Isto é, a empresa é capaz de identificar padrões que são recorrentes no ambiente

empresarial e, a partir da sua análise, retirar conhecimento que apoiará futuras decisões estratégicas. Este procedimento recorre a diferentes componentes científicos para o seu funcionamento, como a IA, a Computação, a Econometria e a Estatística Aplicada. Os dados que se analisam devem ser detalhados e recolhidos continuamente durante um período considerável. De notar que, quanto maior este período, maior a probabilidade por parte do utilizador de realizar uma análise mais eficaz e, conseqüentemente, obter resultados de maior qualidade. Para o sucesso do processo é necessário, também, uma intensa cooperação entre o utilizador e o sistema científico e tecnológico (Frawley et al., 1992; Rostek, 2010).

Da tabela 2 à 4, evidenciam-se vários exemplos de operações que podem compor certas etapas do processo de *Knowledge Discovering*. As operações são impulsionadas pela Big Data, garantindo, logicamente, valor às organizações.

Na tabela 2 retratam-se três operações de controlo da qualidade e desempenho regularmente adotadas no dia a dia de uma empresa: Monitorização do Estado; Determinar o *Overall Equipment Effectiveness* (OEE); Visualização Interativa.

Tabela 2 – Operações de Controlo da Qualidade e Desempenho. Informação de (Bosi et al., 2019; Godinho & Piedade, 2022; Jukic et al., 2015; Prathima et al., 2020; Victor, 2018; Wohlers et al., 2017, 2020)

Operações	
Tipo	Descrição
Monitorização do Estado	A monitorização do estado, também conhecida como <i>Condition Monitoring</i> , é um método que permite analisar continuamente os vários parâmetros que refletem o estado atual da máquina, permitindo avaliar tanto a saúde da mesma, como também a sua eficácia e eficiência. Estas dimensões são estudadas a partir de indicadores de desempenho; indicadores quantitativos que permitem analisar estatisticamente as funcionalidades da máquina a partir dos dados extraídos pelos seus sensores eletrónicos. Com todo um novo espectro de dados a ser gerado, diariamente, as empresas são capazes de analisar novos parâmetros, mais complexos, como por exemplo o OEE. Segundo (Amirat et al., 2009; Lu et al., 2009; Márquez et al., 2010) este método de monitorização pode ser aplicado num vasto "leque" de sistemas mecatrónicos onde, até à data, se obtiveram melhorias consideráveis nas dimensões da qualidade e desempenho.
Determinar o OEE	O OEE é um dos KPI mais utilizados em contexto empresarial e indica a eficiência de um dado equipamento. Este indicador de desempenho é calculado a partir de três dimensões: disponibilidade, velocidade e qualidade. A partir dos novos sistemas de monitorização, compostos pelos diversos sistemas de sensorização, a empresa é capaz de ter uma grande variedade de dados úteis ao cálculo de cada dimensão, permitindo determinar o OEE mais facilmente e com maior rigor, face à maior fiabilidade dos dados.
Visualização Iterativa	Apesar do subcapítulo 2.2.1.1 apenas retratar 5 V, foi referido que existem investigadores que consideram que a Big Data possui mais V. Um destes V representa a palavra Visualização, ou seja, a necessidade por parte das empresas de representar visualmente e interativamente os <i>sets</i> de Big Data, de tal modo que, a informação destes seja mais facilmente compreendida. Portanto, é da máxima relevância para as empresas implementar sistemas de visualização interativa de dados, como: <i>self-aware screens</i> ; <i>Human Machine Interface (HMI)</i> ; Dashboards (ex: Grafana; PowerBi).

Na tabela 3 são descritas quatro operações de *Data Mining*: *Clustering*; Análise Classificativa; Análise de Séries Temporais; Análise de Tendências e Desvios.

Tabela 3 – Operações de *Data Mining*: Informação de (Mishra, 2017; Nagorny et al., 2017; Rostek, 2010)

Operações	
Tipo	Descrição
<i>Clustering</i>	O <i>clustering</i> é um método de análise de dados que permite procurar e determinar classes através das quais os dados serão subdivididos com base nas suas características semelhantes. É um método de <i>unsupervised learning</i> , em que não são definidas quaisquer classes de dados <i>a priori</i> , tendo o modelo de, por si só, encontrar relações entre este conjunto de dados de modo a criar as várias classes.
Análise Classificativa	Esse método de análise permite classificar os dados consoante o seu tipo, face a uma gama de classes predefinidas a partir do uso de <i>model revealing techniques</i> . A análise classificativa é semelhante ao <i>clustering</i> . No entanto, neste caso, existe conhecimento das diferentes classes de dados e das suas características, <i>a priori</i> .
Análise de Séries Temporais	Este processo é composto por métodos de análise de séries temporais, que permitem identificar: tendências, correlações, anomalias, e ciclos sazonais.
Análise de Tendências e Desvios	Este diagnóstico permite detetar diferenças valorativas entre o valor atual em análise e os valores futuros, a partir do estudo da série temporal deste atributo.
<i>Machine Learning</i>	Atualmente, <i>Machine Learning</i> é um ramo da inteligência artificial que envolve o uso de um conjunto de técnicas computacionais (algoritmos e modelos estatísticos) que permitem melhorar o desempenho de um sistema informático numa dada tarefa, a partir do “consumo” de dados experimentais. Temos como famosos modelos de <i>Machine Learning</i> : regressão linear; árvores de decisão; redes neuronais. Focando no setor industrial o <i>Machine Learning</i> é visto como uma ferramenta complexa que permite alcançar um acréscimo de performance a partir do alcance de diversos objetivos como a redução da quantidade de <i>scrap</i> , do consumo de energia ou do <i>lead time</i> de um produto.

Finalmente, na Tabela 4 são indicadas três operações suportadas por algoritmos de *Machine Learning*: Previsão da Manutenção; Previsão da Procura; Detecção de Desperdícios.

Todas estas operações permitem à empresa adotar uma postura proativa, o que significa que não se reage ao incidente quando o mesmo ocorre, antecipa-se e corrige-se a causa antes que o mesmo aconteça.

2.2.3 Desafios da Big Data

A Big Data apresenta uma grande variedade de desafios, que diferem não só nas consequências que estes acarretam como também nas causas que os originam. Relativamente às suas causas os desafios podem ser categorizados de acordo com os 5V. Nos seguintes subcapítulos irão ser expressos

Tabela 4 – Operações de *Machine Learning*. Informação de (Davies et al., 2007; Teti et al., 2022)

Operações	
Tipo	Descrição
Previsão da Manutenção	A manutabilidade, facilidade com que se realiza a manutenção a uma máquina, é geralmente reduzida face à complexidade da arquitetura das novas <i>smart machines</i> . O desconhecimento de quando e qual o componente da máquina que irá falhar, associado à baixa manutabilidade, resulta na perda de um ou mais dias de trabalho. Para além do <i>lead time</i> da entrega da peça sobressalente e do tempo de substituição da mesma, ainda se perderá o que resta do turno de trabalho do dia em que ocorre a falha.
Previsão da Procura	Outro elemento volátil e dependente de imensos fatores é a procura associada aos clientes. Tais características justificam a implementação de modelos preditivos, que permitem às empresas ter uma previsão da procura para os vários produtos, categorizada por um erro previamente calculado durante a fase de teste do modelo. Tal conhecimento permite reduzir <i>stock</i> em armazenamento, otimizar o planeamento e nivelamento da produção e, maximizar o nível de serviço.
Deteção de Desperdícios	A incapacidade, por parte do operador ou da máquina, de detetar um defeito durante ou após o processo produtivo leva a que elevados níveis de <i>scrap</i> sejam produzidos anualmente. Toda este <i>scrap</i> representa uma grande quantidade de matéria-prima desperdiçada e um período produtivo perdido. Este nível de <i>scrap</i> é capaz de ser reduzido a partir da metodologia <i>lean</i> , <i>Jidoka</i> , que retrata a capacidade da máquina de cessar a produção, assim que é detetada uma incoerência no produto e sinalizar instantaneamente o respetivo colaborador para que a medida de correção necessária seja imposta o mais rápido possível. Este nível de automação que a <i>Jidoka</i> representa pode, por exemplo, ser alcançado a partir de metodologias de <i>machine learning</i> , suportadas por câmaras de alta velocidade.

alguns dos principais desafios da Big Data, organizados por cada um dos 5V. Esta subdivisão permite compreender com maior facilidade o desafio que se enfrenta.

2.2.3.1 Volume

Tomada de Decisão

Atualmente, é do senso comum, que qualquer empresa que pretenda estar na vanguarda da inovação e competição, no mercado industrial, necessita de deliberar e tomar todas as suas decisões de negócio com base em dados. Sem dúvida que com a ascensão da Big Data a quantidade e qualidade de novo conhecimento a extrair dos dados aumentou. Contudo, associado a este conhecimento surge uma espera longa designada de Latência (Figura 5). A Latência representa o tempo que uma empresa demora a realizar cada uma das etapas fundamentais do processo de tomada de decisão, e subdivide-se em 3 classes (Rostek, 2010):

1. Latência de preparação – tempo despendido na extração e tratamento dos dados;

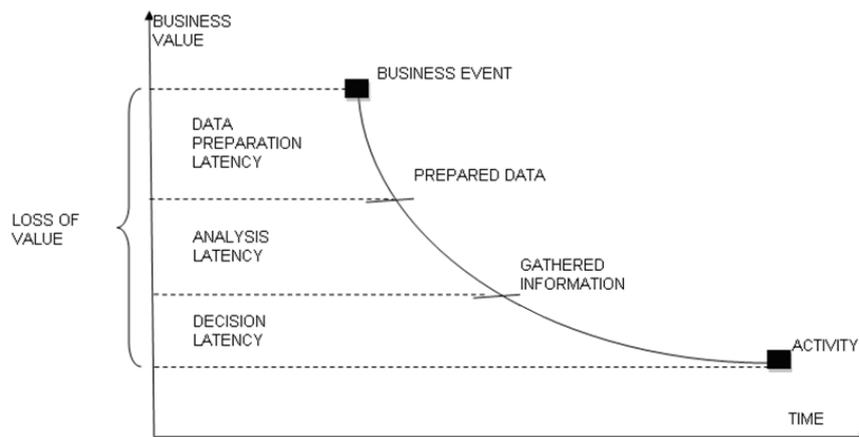


Figura 5 – Perda de valor de negócio face à latência. Retirado de (Hackathorn, 2004).

2. Latência de análise – tempo despendido na análise e transformação dos dados em informação útil e interpretável;
3. Latência de decisão – tempo despendido na análise dos dados, respetiva interpretação, discussão e tomada de decisão.

Face à considerável perda de valor de negócio, várias medidas foram postas em prática com o objetivo de desenvolver novos sistemas de suporte à decisão, que permitem agilizar a etapa de análise e transformação de dados e, desse modo, minimizar este desafio (Rostek, 2010).

Segurança e Privacidade

As organizações têm um dever moral e ético para com as entidades com as quais colaboram diariamente, tais como: os clientes, colaboradores e fornecedores. Qualquer empresa que pretenda trabalhar com Big Data terá de gerir uma enorme quantidade de dados que contém informação sensível a todas as entidades referidas. Daí a necessidade, por parte das organizações, de manter os dados privados e seguros a todo o instante. Qualquer brecha de informação que possibilite a partilha de conhecimento privado irá resultar numa perda de valor bastante substancial. Com o aumento exponencial da quantidade de dados gerada, como demonstrado pela Figura 2, torna-se um desafio para a empresa gerir e garantir a segurança de todos os dados em circulação e armazenamento (Chandarana & Vijayalakshmi, 2014; Jukic et al., 2015).

2.2.3.2 Variedade

Gestão e Partilha dos Dados

No presente, certas empresas optam por partilhar parte dos seus dados internos, continuamente, com os seus colaboradores, ao longo da cadeia de abastecimento, com o intuito de obter

uma vantagem competitiva face a outras cadeias. Esta vantagem é sustentada pela capacidade das várias empresas se autogerirem mais facilmente, sem a necessidade de, permanentemente, requisitar informação por meios tradicionais, como o *email*, aos seus colaboradores. Vários processos como o planeamento da produção e o controlo de *stock* acabam por se tornar mais eficazes face à partilha em tempo real dos dados entre organizações. Todavia, os dados só têm valor se forem acessíveis e interpretáveis, o que não é fácil de garantir devido à grande heterogeneidade dos mesmos e a consequente necessidade, por parte da empresa, de ter um processamento complexo para cada classe de dados. Existe, portanto, a necessidade de garantir um tratamento de dados que os permita normalizar, de modo a agilizar a análise por parte das outras empresas, bem como garantir que todas as normas de privacidade são atingidas, de modo que não se partilhe qualquer informação confidencial (Chandarana & Vijayalakshmi, 2014; Jukic et al., 2015).

Processamento de Dados Não Estruturados

Com todo o movimento da Indústria 4.0, especialmente no que toca ao desenvolvimento de tecnologias de monitorização e extração de dados, a tipologia de armazenamento em que toda a informação é armazenada em formato tabular e onde tudo é categorizado por atributos predefinidos, não representa a realidade atual. A desestruturação dos dados que se verifica leva a que toda a metodologia de processamento necessite de ser continuamente analisada e reinventada, de modo que qualquer classe de dados em circulação na empresa consiga ser tratada (Chandarana & Vijayalakshmi, 2014; Moniruzzaman & Hossain, 2013).

2.2.3.3 Velocidade

Tecnologia e Arquitetura

A velocidade a que os dados são gerados, diariamente, coloca imensa pressão nos fornecedores tecnológicos que têm de desenvolver e garantir que um *software* seja capaz de processar, armazenar e gerir toda a informação de um modo eficaz e eficiente. Com o ritmo, cada vez maior a que os dados são gerados, os sistemas que hoje são modernos o suficiente, facilmente se tornam ultrapassados, sendo incapazes de suportar o tratamento e análise dos dados (Chandarana & Vijayalakshmi, 2014).

2.2.3.4 Veracidade e Valor

Autenticidade

Assegurar a veracidade dos dados é um dos requisitos mais importantes por parte de uma empresa, face à necessidade de se trabalhar com dados reais, autênticos e de qualidade. Em contexto de Indústria 4.0, em que as decisões estratégicas são suportadas pelo conhecimento que advém da análise de dados, o uso de dados não conformes à realidade acaba por resultar em conclusões incorretas. Com toda a quantidade de dados a circular nos departamentos de uma organização, torna-se difícil para a empresa: identificar quais os dados que são úteis; assegurar a sua autenticidade; e estimar o seu valor (Chandarana & Vijayalakshmi, 2014; Moniruzzaman & Hossain, 2013).

2.3 Armazenamento de Dados

2.3.1 Base de Dados Relacional

Uma base de dados é formada a partir do armazenamento de um grupo de dados com um grau de valor considerável, resultantes de uma recolha lógica e coerente. É, a partir deste meio, que se permite que qualquer utilizador consiga aceder e extrair informação a qualquer instante (Navathe & Elmasri, 2005). A grande maioria desta informação é estruturada e armazenada numa base de dados tabular, denominada de base de dados relacional. Esta metodologia de armazenamento baseia a sua estrutura no conceito matemático de relação, em que a cada elemento (também denominado de tuplo) representa uma linha à qual estão associados diversos atributos (colunas), todos eles interrelacionados e representando um certo valor, bem definido, nos seus diversos domínios: nome; tipo de dados; formato; entre outros. Esta tipologia de base de dados tem permanecido, durante várias décadas, a opção padrão de armazenagem de dados (Oliveira et al., 2022; Sadalage & Fowler, 2013). Este esquema estrutural encontra-se representado na Figura 6, onde é ilustrado um conjunto de tabelas relacionais que compõem uma base de dados. Os atributos assinalados com um asterisco (*) representam as chaves, elemento a explicar com maior detalhe no subcapítulo 2.3.2.4.

Com o impulsionamento da frequência com que novos dados são gerados a cada instante, incentivado pela Indústria 4.0, é cada vez mais difícil gerir e armazenar tão grande volume de dados. Torna-se, assim, necessário recorrer ao uso de *Database Management Systems*, classes de ferramentas que permitem ao utilizador: armazenar, transformar e recuperar todo o tipo de dados previamente armazenados, assegurar a integridade dos mesmos, prevenindo acessos não autorizados e, suportar diferentes visões dos dados. Uma das classes mais relevantes e usadas globalmente são os *Relational*

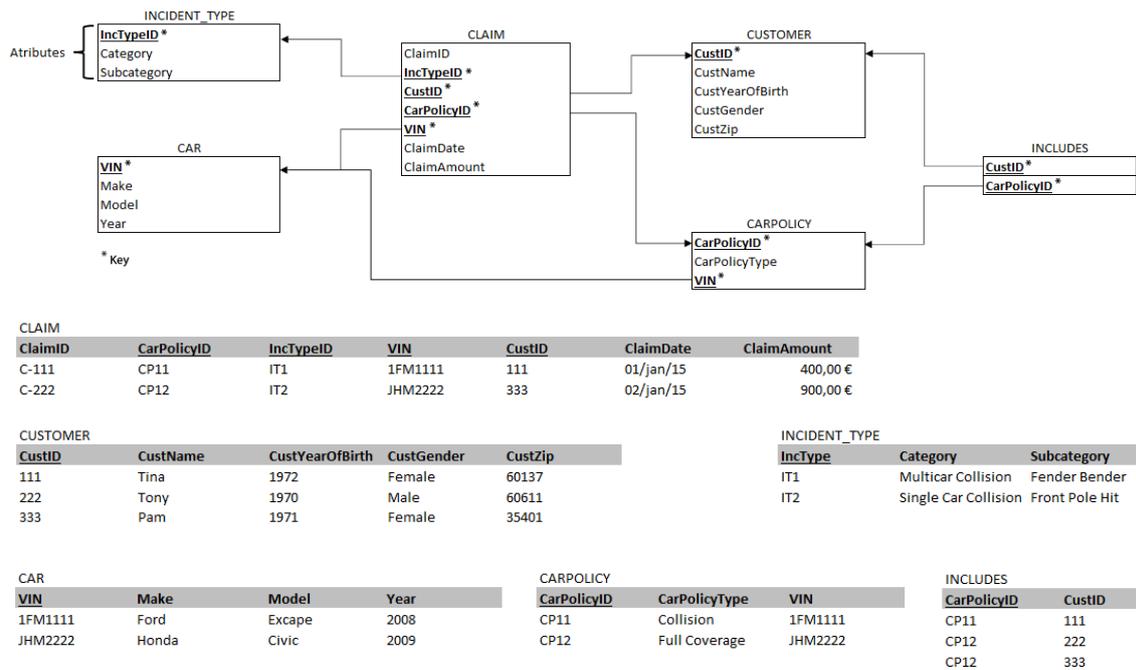


Figura 6 – Base de dados relacional. Retirado de (Jukic et al., 2015).

Database Management Systems (RDBMS), também conhecidos como sistemas de *Structured Query Language* (SQL), face à predominância da linguagem SQL nas *queries* utilizadas pelo sistema no desenvolvimento de modelos relacionais. Temos como exemplo de um RDBMS, o Microsoft SQL Server, sistema utilizado no presente projeto. As bases de dados geridas por esta ferramenta adotam como arquitetura de armazenamento o formato clássico suportado pela metodologia relacional (Elmasri & Navathe, 2000; Fatima & Wasnik, 2016; Ramakrishnan, 2008).

Os RDBMS têm a obrigação de respeitar propriedades que se relacionam com todas as transações que decorrem dentro de uma base de dados SQL, durante o acesso aos dados ou durante possíveis falhas dentro do sistema. Estas propriedades subdividem-se nos seguintes atributos: atomicidade; consistência; isolamento; durabilidade (ACID) (Ali et al., 2019; Fatima & Wasnik, 2016; Oliveira et al., 2022):

1. Atomicidade – qualquer transação tem de ser executada na íntegra, caso contrário não é executada. A atomicidade garante que caso ocorra alguma falha durante a transação, primeiramente a transação é interrompida e, conseqüentemente, qualquer tipo de mudança que tenha ocorrido no *dataset* será revertida.
2. Consistência – caso uma transação ocorra na totalidade, o sistema tem de assegurar que a base de dados parte de um estado consistente para outro. Isto é, o seu estado inicial e final tem de respeitar todas as restrições da estrutura da própria base de dados.

3. Isolamento – todas as transações são independentes; a execução de uma transação não pode sofrer interferência pela execução de outra que decorre em paralelo.
4. Durabilidade – qualquer transformação nos dados do *dataset* provocada pela conclusão de uma transação deve permanecer na base de dados.

As propriedades ACID são uma das características e atributos mais importantes das bases de dados relacionais (Fatima & Wasnik, 2016).

2.3.2 Data Warehouse

Uma das arquiteturas de bases de dados relacionais mais conhecidas e utilizadas atualmente é a DW, introduzida pela primeira vez em (Inmon, 1992) com o objetivo de suportar os sistemas de tomada de decisão que sentiam enorme dificuldade em gerir e processar o grande volume e complexidade de dados gerados diariamente. Shi et al. (2001) defendem que é a partir do uso de uma DW, uma plataforma segura onde grandes volumes de dados históricos estão integrados, que qualquer utilizador consegue processar e analisar informativa e analiticamente um conjunto de dados colhidos durante um espectro temporal. Permite que qualquer organização seja capaz de suportar as suas decisões estratégicas e operacionais a partir de conclusões analíticas resultantes de operações previamente realizadas como: desenvolvimento de relatórios; aplicação de *Data Mining*, entre outros (Inmon, 1996; Rostek, 2010). (Inmon, 1992) define genericamente uma DW como uma coleção de dados integrados, orientados a um certo tema, históricos e não voláteis (Figura 7).

Orientados para um tema

No ambiente comum de uma base de dados o foco dos dados passa por ser um único tipo de processo ou função, como por exemplo, a prensagem realizada por uma prensa em chão de fábrica. No entanto, numa DW os dados são processados de modo que todo um campo macroscópico seja analisado;

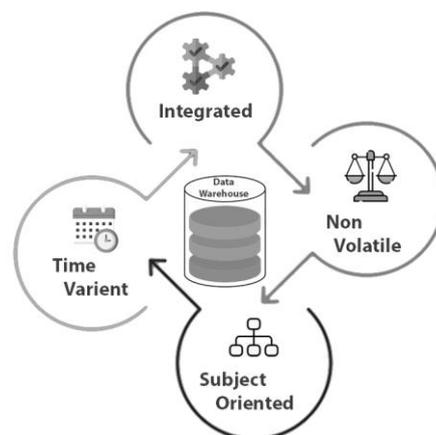


Figura 7 – Propriedades de uma DW.

como a qualidade de vários produtos da mesma família ou o desempenho de uma dada linha de produção (Shi et al., 2001).

Integrados

Shi et al. (2001) acreditam que a integração dos dados é considerada pela comunidade científica a característica mais importante e mais relevante de uma DW. Qualquer sistema de tomada de decisão tem como preferência *inputs* normalizados e integrados comparativamente a *inputs* heterogêneos, em qualquer uma das classes dos seus atributos: nome, estruturado, tipo, dimensão e codificação. Isto acontece porque dados sumarizados e normalizados permitem que exista apenas uma sintaxe global que facilita o tratamento e análise de dados, já que esta passa a ser única para todos os dados. A integração exigida pela DW garante que este requisito seja cumprido por qualquer classe de dados (Bontempo & Zagelow, 1998; Chaudhuri & Dayal, 1997).

Históricos

Todos os dados armazenados na DW possuem, acoplados a si, um valor temporal que garante à base de dados um registo histórico dos seus dados, permitindo que se analise, por exemplo, possíveis tendências ou sazonalidades. Esta característica será explorada com um maior detalhe no subcapítulo 2.3.2.1 aquando da comparação entre as bases de dados operacionais e as bases de dados analíticas, como é o caso da DW (Watson et al., 2001).

Não voláteis

Todos os dados armazenados numa DW não podem ser atualizados ou modificados, Watson et al. (2001) classificam-nos mesmo como dados de leitura. A imutabilidade destes dados garante à DW um *design* técnico bastante mais simples comparativamente a outras bases de dados em que operações de transformação são permitidas.

2.3.2.1 Base de Dados Operacional vs Base de Dados Analítica

A capacidade de compreender a diferença entre uma base de dados operacional e uma analítica é um dos maiores entraves ao desenvolvimento de uma DW. Ambas fazem parte da arquitetura do sistema que envolve uma DW e, contribuem para o sucesso do sistema de tomada de decisão de uma organização (Gardner, 1998).

Base de dados operacional

Quando se fala de uma base de dados operacional fala-se, obrigatoriamente, de *Online Transaction Processing* (OLTP), processo responsável por registar digitalmente todas as transações que traduzem valor para a empresa. Este processo é realizado diariamente em tempo real, com elevada consistência e atuando em pequenos grupos de dados de cada vez (Oliveira et al., 2022).

As aplicações do OLTP automatizam várias tarefas de processamento de dados e são suportadas a partir de uma base de dados operacional, ou seja, uma estrutura de armazenamento de dados digital que guarda toda a informação resultante das várias iterações funcionais que decorrem regularmente dentro de um negócio de uma empresa. As operações alvo do OLTP são tarefas diárias, isoladas e repetitivas, que decorrem nos vários departamentos de uma empresa e nas quais os atributos que as compõem traduzem valor para o negócio, como é exemplo o nível de temperatura interna de uma máquina de manufatura. Cada vez que uma transação é finalizada, a base de dados é atualizada incrementando a informação do *dataset* e permitindo que os colaboradores consigam monitorizar em tempo real o estado de cada processo (Chaudhuri & Dayal, 1997).

Este tipo de base de dados também é característico pela heterogeneidade dos seus dados face às restantes bases de dados intraempresa. É claro que, diferentes funções tendem a possuir não só objetivos completamente distintos, como também diversidade nos elementos com os quais interagem ao longo do seu processo. Veja-se o exemplo, no controlo da qualidade dos produtos vendidos, o código identificativo do cliente é um elemento fulcral, pois permite distinguir um pedido dos restantes da mesma família de produto. No entanto, na realização do planeamento semanal da produção, a informação associada ao cliente já é obsoleta; não garante qualquer valor à empresa armazenar o código identificativo do cliente (Oliveira et al., 2022). Em linha com a heterogeneidade das funções é importante realçar que as bases de dados operacionais são "*purpose oriented*", ou seja, cada base de dados é desenhada consoante a finalidade da operação em análise, podendo ter poucos ou muitos utilizadores, dependendo da dimensão da operação em análise (Teorey et al., 2011).

Face à elevada frequência com que os dados são gerados, é de grande importância que as empresas que adotam estas base de dados possuam um sistema de controlo de falhas que permita verificar a consistência dos dados após cada transação (Srivastava & Chen, 1999).

Base de dados analítica

Ao contrário da base de dados operacional a base de dados analítica não é "*purpose oriented*", mas sim "*subject oriented*", este tipo de bases de dados não se limita a armazenar apenas a informação

de uma função, também permite armazenar dados que representem todas as valências de uma entidade ou departamento, como é o caso da DW, que armazena num só local, centralizado, toda a informação sumarizada de uma organização que se mostre relevante. Este tipo de bases de dados é desenhado e otimizado de forma a suportar o *Online Analytical Processing* (OLAP), um processo de análise de dados que suporta a tomada de decisão, recorrendo à extração de conhecimento a partir de um grande volume dados normalizados com um considerável grau de complexidade (Bontempo & Zagelow, 1998; Cuzzocrea et al., 2013).

Neste tipo de base de dados, todos os dados necessitam de sofrer um processo de normalização e transformação, por forma a garantir a homogeneidade que respeita a estrutura desenhada por cada departamento da empresa. Num ambiente analítico, dados armazenados que se caracterizam por ser limpos, sumarizados, consolidados e capazes de possuir um valor histórico representam um maior valor para empresa do que dados individuais que representam iterações independentes de uma operação (Figura 8). De relevar que o facto de as bases de dados serem *subject oriented* leva a que diferentes utilizadores possuam diferentes requisitos. Portanto, o *design* do processamento e consequente estrutura da base de dados tem de ser ajustado à necessidade de cada departamento (Chaudhuri & Dayal, 1997). Enquanto nas bases de dados operacionais o valor histórico temporal dos dados não é relevante, acabando mesmo por se eliminar regularmente grandes quantidades de dados “antigos”, na base de dados analítica o mesmo não acontece; a sua construção é completamente dependente da existência de uma grande quantidade de dados históricos, como por exemplo séries temporais. Logo, a dimensão e volume de dados presentes numa base de dados analítica é muito superior à da base de dados que lhe fornece os dados e que geralmente acaba por ser mesmo uma base de dados operacional. Neste caso, a incrementação e registo dos dados é, normalmente, realizada em lotes a partir de atualizações, ao

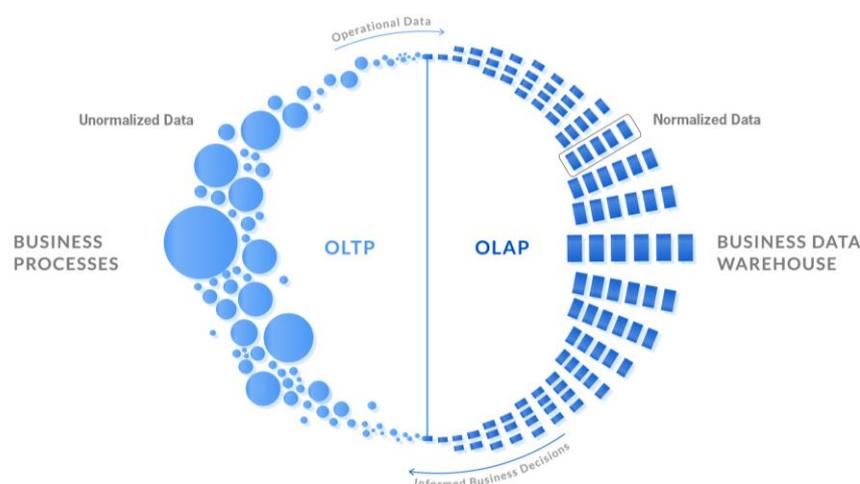


Figura 8 – Contrastes entre OLTP e OLAP. Adaptado de (Franco, 2019).

contrário das bases de dados operacionais, na qual decorre continuamente (Elmasri & Navathe, 2000; Oliveira et al., 2022; Teorey et al., 2011).

Na Tabela 5 sumarizam-se as principais distinções entre as bases de dados operacionais e as analíticas.

2.3.2.2 *Extração Transformação e Armazenamento dos Dados*

O valor resultante das diversas análises efetuadas aos dados de uma DW a partir de um sistema de tomada de decisões está totalmente dependente do cumprimento dos requisitos estruturais de uma DW. Como referido no capítulo 2.3.2.1, os dados que alimentam uma DW são normalmente gerados a partir de base de dados operacionais; bases de dados que por natureza possuem dados heterogéneos e detalhados, características estas incompatíveis com a estrutura de uma DW. É, naturalmente, necessário que este conjunto de dados seja sujeito a um conjunto de etapas de tratamento de dados de maneira que lhes seja possível garantir as características compatíveis com uma DW. É aqui que entra em ação o processo ETL, um processo de extração, limpeza e armazenagem de dados que atua como ponto de contacto entre uma DW e as respetivas fontes de informação, como está representado na Figura 9 (D. Shi et al., 2001)).

Esta sigla advém das palavras: Extração; Transformação; Armazenamento, que representam as três operações que compõem o processo ETL e que permitem que, a partir da infraestrutura do sistema que suporta o processo ETL, se consiga (Jukic et al., 2015):

1. Extrair, a partir de bases de dados operacionais, todos os dados que traduzem valor analítico para a organização;
2. Transformar todos os dados extraídos, garantindo que todas as suas propriedades ficam conformes à estrutura designada pela DW;
3. Armazenar todo o conjunto de dados transformados na DW.

Tabela 5 – Base de Dados Operacional vs Base de Dados Analítica

Base de dados operacional	Base de dados analítica
Purposed Oriented	Subject Oriented
Dados não integrados	Dados integrados
Inexistência de valor histórico temporal	Presença de valor histórico temporal
Baixo volume de dados	Grande volume de dados
Incrementação contínua	Incrementação em lotes

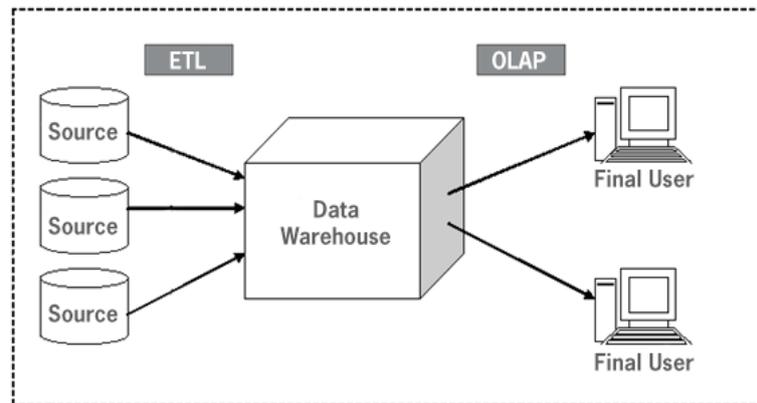


Figura 9 – Arquitetura minimalista de uma *Data Warehouse*. Adaptado de (Rostek, 2010).

De realçar que o processo ETL difere, normalmente, de DW para DW, devido à divergência dos aspetos que moldam uma base de dados de cada organização, como: o tipo de dados; a sua organização estrutural; os seus objetivos; o tipo de OLAP a realizar. É a combinação destes fatores que acaba por delinear qual é o processo de transformação pelo qual os dados necessitam passar.

2.3.2.3 *Arquitetura de uma Data Warehouse*

Uma das características que distingue uma DW da grande maioria das restantes bases de dados relacionais e analíticas é a complexidade presente na arquitetura do sistema que a rodeia. Qualquer organização que pretenda implementar uma DW com o intento de extrair todo o seu potencial, requer a presença de um sistema composto pelos seguintes elementos: fonte interna e externa de dados; ETL; DW; *Data Mart*; *Metadata*; OLAP, como representado na Figura 10.

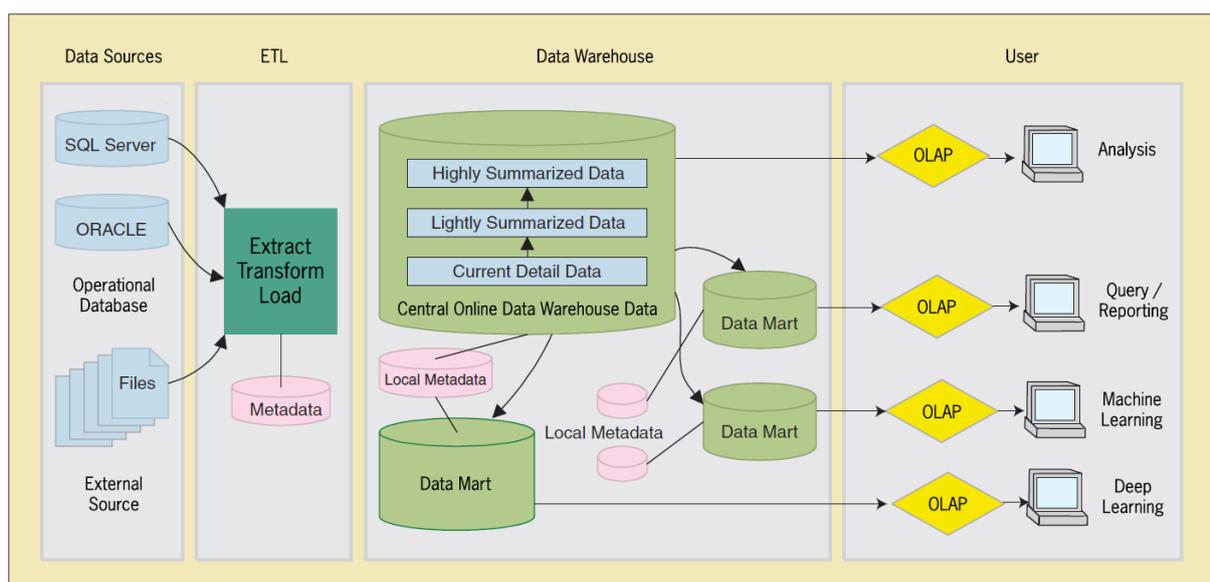


Figura 10 - Arquitetura de uma DW. Retirado de: (Shi et al., 2001).

Fonte de dados

Os dados armazenados numa DW são, geralmente, gerados a partir de dois tipos de fontes: bases de dados operacionais e fontes de informação externas. Como referido em capítulos anteriores uma base de dados operacional fornece informação sobre processos e operações industriais em curso, como por exemplo sistemas de MRP que planeiam e lançam ordens de compra e produção semanalmente, ou sistemas de *Enterprise Resource Planning* que geram a automatização de vários processos e equipamentos da organização. Já as fontes externas de dados, como o próprio nome indica, geram conhecimento externo à empresa, que reflete tendências e condições do mercado externo que podem ter impacto no negócio, tais como o *feedback* dos clientes face aos nossos produtos, ou dados específicos do mercado, como os preços de mercadorias e transporte (Jukić et al., 2015; Jukic et al., 2016; Shi et al., 2001).

Data Warehouse

A DW é o elemento fundamental e central do sistema, já que representa o local onde todo o volume de dados extraído por cada uma das fontes é armazenado, promovendo a centralização e integração da informação. Todos os dados, como supracitado, são sujeitos a um processo de limpeza e transformação prévio à armazenagem que lhes garante um grau de elevada consistência e qualidade. Os dados são recolhidos e armazenados, periodicamente, em lotes, garantindo a presença na base de dados de um historial que permite identificar mudanças e tendências no mercado. Acabam por ser sumarizados e agregados, agilizando as tarefas analíticas e, são organizados consoante o departamento a que acrescentam valor, designadamente: controlo da qualidade; planeamento da produção; marketing (Bontempo & Zagelow, 1998; Chaudhuri & Dayal, 1997; Inmon, 1996).

Data Mart

Um *Data Mart* é nada mais que um subconjunto de uma DW. Possui todas as características de uma DW, quer na tipologia estrutural quer no objetivo. A única diferença entre as mesmas assenta na singularidade da categoria dos dados que um *Data Mart* armazena. Enquanto uma DW armazena, num local centralizado, todos os atributos relevantes aos vários departamentos de uma empresa, um *Data Mart* armazena apenas informação relativa a um único departamento. Daí ser possível identificar vários *Data Marts* na Figura 10, cada elemento representa um dos diversos departamentos monitorizados (Shi et al., 2001). (Srivastava & Chen, 1999) cimenta a relação entre uma DW e um *Data Mart* afirmando que os dados que constituem um *Data Mart* são conjuntos de dados provenientes da DW, extraídos regularmente em lotes.

Metadata

Shi et al. (2001) e Gardner (1998) definem *metadata* como “sendo dados sobre dados” e “tudo o que define um objeto de uma *Data Warehouse*, como uma tabela, um atributo, uma *query*, um relatório ou um algoritmo de transformação”, respetivamente. As duas definições estão corretas, uma *metadata* é, em suma, um dicionário digital que contém toda a informação referente aos dados armazenados numa DW. Geralmente a *metadata* inclui conteúdos como (Chaudhuri & Dayal, 1997; D. Shi et al., 2001):

1. Modelo dos dados;
2. Estrutura dos dados;
3. Regras de processamento e transformação dos dados;
4. Algoritmos que potenciam a sumarização dos dados;
5. Ferramentas *back-end* e *front-end*.

Este conteúdo é a chave para a compreensão e utilização eficaz de uma DW.

OLAP e ETL

Como definido no capítulo 2.3.2.1, OLAP é um processo analítico padrão realizado em *Data Warehouses* com o objetivo de garantir aos colaboradores de uma empresa valor estatístico e analítico que se traduza em conhecimento (Shi et al., 2001). Este conceito foi introduzido por (Codd, 1998) que acreditava que nenhuma organização era capaz de extrair respostas e conhecimento a partir de uma base de dados operacional. O conceito por detrás do OLAP é simples: permitir que qualquer utilizador consiga manipular os dados alocados e sumarizados numa DW permitindo a implementação de um leque de testes e operações analíticas; como representado na Figura 10. Porém, o OLAP requer que os dados a processar estejam organizados sob requisitos específicos, relacionados com métodos de acesso e implementação concretos (Chaudhuri & Dayal, 1997; Cuzzocrea et al., 2013). É aqui que entra o ETL, garantindo que todos os requisitos são preenchidos a partir do desenvolvimento de cada uma das suas três etapas, como é demonstrado na Figura 11.

2.3.2.4 Esquema de Estrela

Apesar da tipologia integrada, sumarizada e temporal que os dados alocados a uma DW têm de respeitar, devido às características da própria DW e do seu carácter analítico, o esquema estrutural da base de dados, em si, é completamente personalizado pela organização. Contudo, esta personalização respeita, geralmente, um esquema de estrela, um esquema estrutural capaz de representar todo o modelo multidimensional de uma DW (Chaudhuri & Dayal, 1997). Este esquema de dados caracteriza-

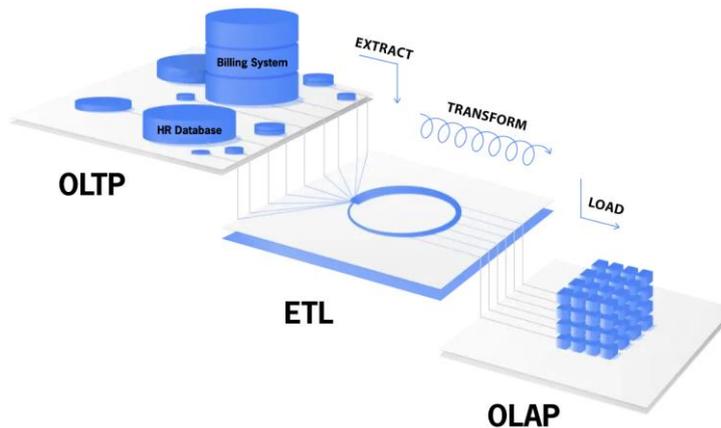


Figura 11 – Fluxo de informação. Retirado de (Borges, 2020).

se pela sua simplicidade e facilidade de paralelização, fatores que normalmente impulsionam o desenvolvimento de aplicações de tomada de decisão com um alto desempenho. É este mesmo desempenho que assegura a prioridade que este tipo de *design* tem em relação a outros (Stonebraker, 2011). A Figura 12 ilustra um exemplo de um esquema de estrela industrial.

A designação esquema de estrela advém da aparência que o diagrama do esquema possui. Uma tabela no seu centro, a tabela facta, e as restantes tabelas que se interligam a esta, as tabelas de dimensão. A tabela facta representa uma tabela singular que contém todos os apontadores que representam as chaves únicas, chaves identificativas capazes de unir cada tuplo da tabela central a uma das possíveis tabelas de dimensão. Já as tabelas de dimensão armazenam toda a informação descritiva de cada um dos campos registados na tabela facta, ou seja, cada uma destas tabelas armazena todos os atributos classificativos de uma certa dimensão que traduzem conhecimento pertinente à empresa. De um modo simples, cada tabela de dimensão pode ser interpretada por um conjunto de colunas em que cada uma corresponde a um atributo da dimensão em questão. O *design* claro e intuitivo do esquema permite uma fácil compreensão dos dados que leva a que as tabelas de facta e de dimensões possam

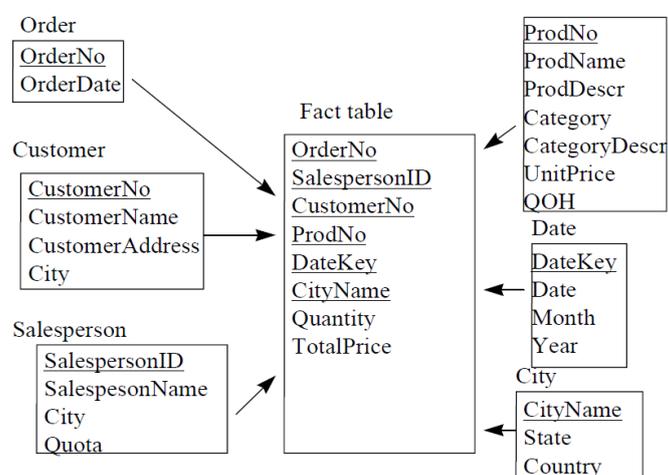


Figura 12 – Exemplo de um esquema de estrela. Retirado de (Chaudhuri & Dayal, 1997).

ser facilmente unidas para responder a questões empresariais complexas (Chaudhuri & Dayal, 1997; Myriantous, 2022).

É de referir que a fácil escalabilidade deste esquema é outro parâmetro relevante, uma vez que permite à base de dados ser facilmente expansível, sempre que necessário, algo bastante útil face à elevada taxa com que os dados são gerados no atual ambiente empresarial. São estas relações e propriedades inerentes ao esquema que permitem aos analistas de dados ou profissionais de *Business Intelligence* navegar e consultar facilmente os dados (Chaudhuri & Dayal, 1997; Stonebraker, 2011).

2.4 Monitorização do Chão de Fábrica

A digitalização dos processos produtivos e sistemas ativos em chão de fábrica permite que os gestores e engenheiros sejam capazes de extrair dados autênticos das operações em execução no chão de fábrica, preferencialmente em tempo real. Recorrendo a este conjunto de dados e a ferramentas que os permitam adquirir e processar é, agora, possível fazer uma avaliação complexa e contínua baseada em dados reais, sobre o estado atual de uma máquina e do seu processo (Mourtzis et al., 2016; Teti et al., 2022). Min et al. (2011) defendem mesmo que o uso de processos de monitorização baseados em sistemas de sensorização permite facilitar o controlo e *setup* de máquinas de grande precisão; máquinas essas, cada vez mais requisitadas face à sua eficiência produtiva e sustentabilidade. São estes fatores, que a cada dia, impulsionam a mudança do paradigma tradicional da produção, tornando-o cada vez mais digital (Teti et al., 2022).

No presente capítulo serão abordadas duas tipologias distintas de processos de monitorização: o controlo do processo e o controlo do produto, ambos suportados pela monitorização do estado, a desenvolver no presente subcapítulo. A sua divergência assenta não tanto na arquitetura e ferramentas utilizadas para implementar os mesmos, mas, maioritariamente, nos objetivos de ambos os processos e as classes dos KPI que os monitorizam.

A monitorização do estado, criado por (Prince & Shortland, 1987), é um processo de monitorização que permite avaliar o estado das máquinas produtivas instaladas em chão de fábrica e dos respetivos processos produtivos, recorrendo a dados gerados pelas máquinas ou pelo processo em que se inserem. Este tipo de operação monitoriza parâmetros capazes de refletir o condicionamento da máquina ou do processo, nomeadamente: a temperatura; a vibração; o número de produtos não conformes em um dia; o tempo de ciclo (Wohlert et al., 2020).

Este tipo de monitorização é importante na medida em que permite regular a eficiência e eficácia dos processos produtivos, já que qualquer alteração nos parâmetros medidos que vá além do espetável representa o evento de um estado indesejado que geralmente acaba por gerar falhas no produto e paragens produtivas face à manutenção necessária. Todos estes fatores prejudicam a eficiência e eficácia da organização. Logo, a implementação de sistemas deste tipo que promovem uma perceção real e contínua dos processos, designada de controlo da qualidade, acaba, normalmente, por gerar melhorias significativas no que toca ao desempenho e qualidade da empresa (Amirat et al., 2009; Lu et al., 2009).

Segundo Juran & Feo (2010) o controlo da qualidade pode adotar diferentes tipos: controlo de *setup*; controlo do processo; controlo do produto; controlo da instalação. Cada uma destas classes pode ser realizada a partir da monitorização do estado. Porém, cada uma diverge das demais face aos objetivos que pretende alcançar. Neste capítulo serão analisados mais detalhadamente dois destes quatro tipos de controlo de qualidade, os mais relevantes para o projeto a desenvolver na presente dissertação.

Controlo do Processo

O controlo do processo permite monitorizar as operações do mesmo avaliando o seu desempenho a partir da comparação dos valores dos KPI reais com os valores requeridos pela empresa. Caso seja identificado algum desvio significativo entre estes dois valores, medidas de correção serão tomadas. Este tipo de monitorização do estado requer uma implementação regular, preferencialmente contínua, de modo a alcançar-se uma monitorização em tempo real do processo. Um dos KPI de referência deste tipo de controlo é o OEE que, como referido anteriormente, permite monitorizar a eficiência de um processo (Wohlers et al., 2020).

Controlo do Produto

O controlo do produto, como a designação indica, consiste em operações de inspeção localizadas que avaliam a qualidade do produto já finalizado, comparativamente aos níveis de excelência pré-estabelecidos pela empresa. Finalizada a inspeção, a informação será registada servindo de *input* a KPI de controlo de produto, como é exemplo a taxa diária de produtos não conformes. Apesar das inspeções finais serem um dos principais métodos de controlo de qualidade, é importante perceber que em ambientes produtivos o controlo da qualidade também envolve todo o controlo e monitorização de indicadores de desempenho, durante o processo produtivo, que direta ou indiretamente influenciam a qualidade final do produto (L. Wang & Gao, 2006).

2.5 Classes de Indicadores de Desempenho

Qualquer KPI independentemente da área que representa possui a mesma finalidade: avaliar até que ponto os objetivos estratégicos previamente delineados pela empresa estão a ser alcançados (Juran & Godfrey, 1999). A figura 13 ilustra os termos-chave que devem ser devidamente considerados aquando do desenvolvimento e estudo de um KPI: Com vista a potenciar o desempenho de um processo, torna-se imperativo atuar de forma estratégica, recorrendo a indicadores que incorporem dados concretos e reais. É através da mensuração dos KPIs que se procede, em consonância com o objetivo estabelecido, à avaliação dos aspetos do processo nos quais a otimização se revela possível. Para tal, é vital a presença de um sistema que meça o desempenho da empresa, ainda que cada indústria tenha características e particularidades próprias, o que leva a que existam processos específicos a necessitarem de ser avaliados. Assim sendo, cada empresa possuirá indicadores de desempenho únicos, distintos de qualquer KPI presente nas restantes (Cristea & Cristea, 2021; Spitzer, 2007).

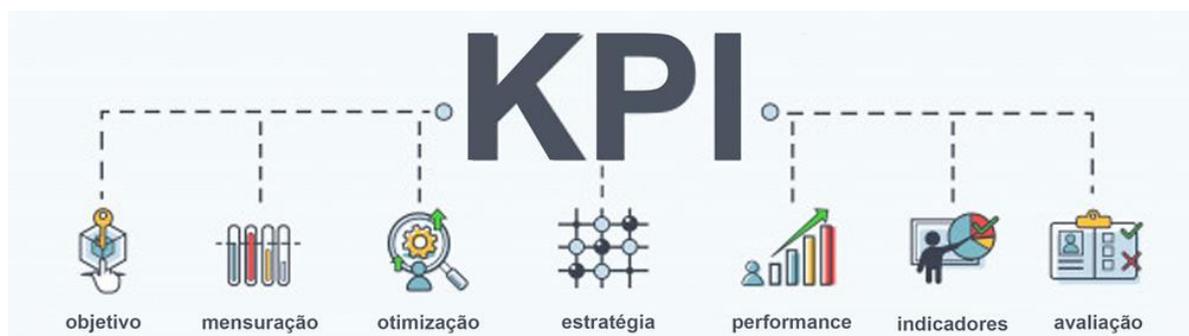


Figura 13 - Termos chave de um KPI. Retirado de: (Alonso, 2022)

Analisando a perspetiva interna da empresa relativamente à estrutura organizacional dos seus KPI, é natural que cada empresa reparta os seus indicadores por várias classes consoante a área que o indicador de desempenho avalia. Tem-se como exemplo o caso empresarial reportado por Cristea & Cristea (2021), no qual se implementou um método de hierarquização de KPI numa empresa de embalagem que subdivide todos os seus indicadores em 6 diferentes áreas: produção; qualidade; financeiro; *feedback* do cliente; *feedback* do colaborador e sustentabilidade, como demonstrado na Figura 14.

Para além da tipologia de classes ilustradas na Figura 14, os KPI podem ser subdivididos em dois tipos de sistemas que diferem no tipo de utilizador para o qual são desenhados: Sistema de medição de desempenho (PMS); sistema de monitorização de desempenho pessoal.

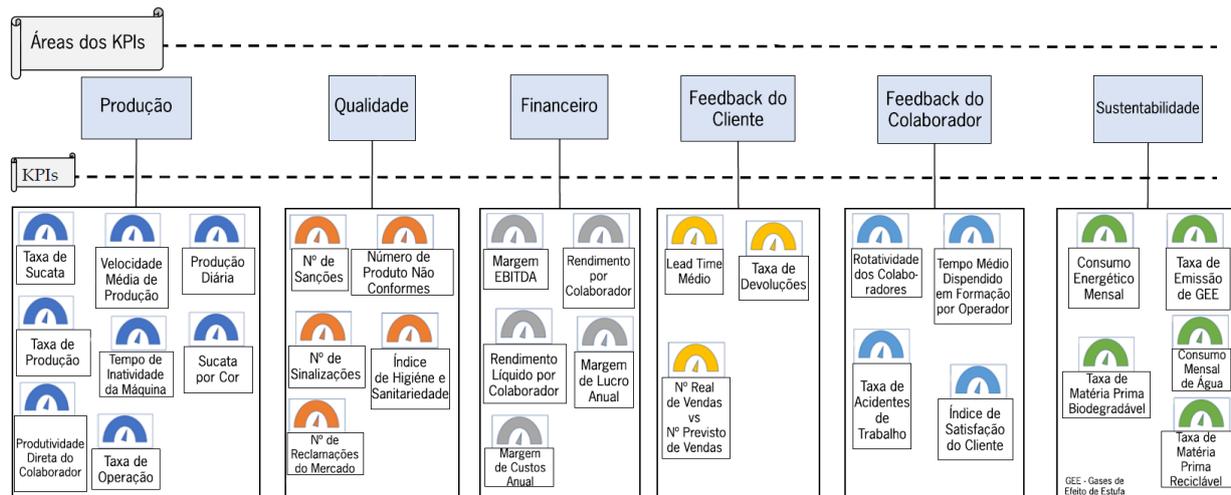


Figura 14 – Classes de KPIs. Adaptado de (Cristea & Cristea, 2021).

2.5.1.1 Sistemas de Medição de Desempenho

Em sistemas de medição de desempenho os KPI são projetados para a empresa ou para um departamento da sua hierarquia interna. Este tipo de sistema tem como objetivo realçar, visualmente, o desempenho em tempo real dos diversos indicadores, de modo a garantir que qualquer processo de tomada de decisão estratégica ou operacional possua dados recentes, reais e válidos. A Figura 15 ilustra um HMI o qual incorpora um PMS que permite aos colaboradores de uma empresa de produção de cerveja analisar, simultaneamente, a performance de diversos postos de trabalho em chão de fábrica (Wohlert et al., 2020).

Um PMS bem configurado pode revelar os pontos fortes e fracos de uma organização, ajudar a identificar áreas de melhoria, e orientar os processos internos de tomada de decisão. Geralmente o

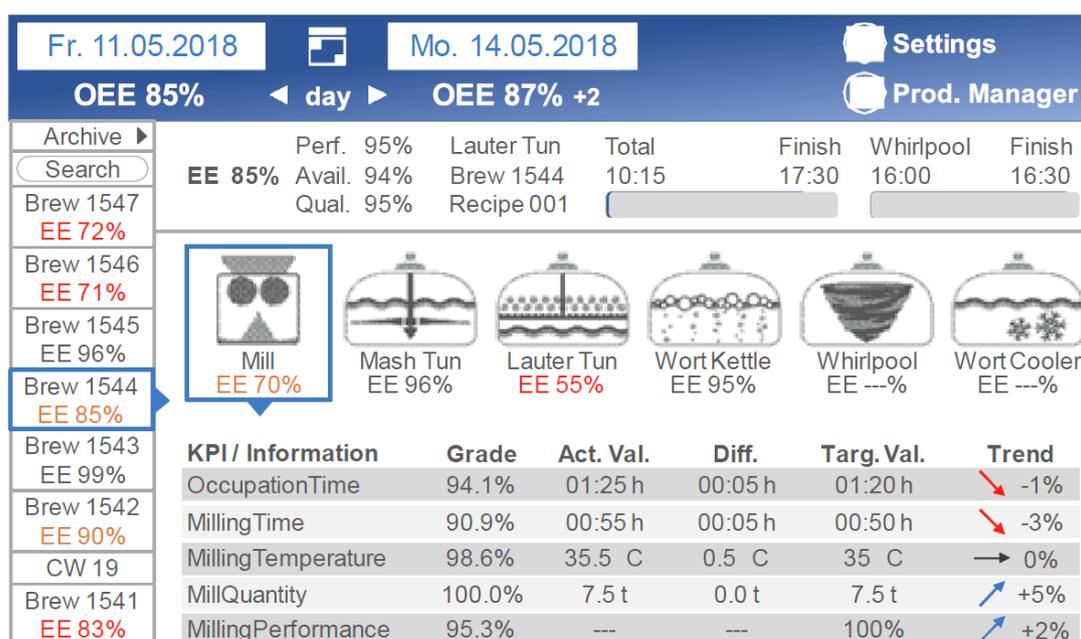


Figura 15 – Sistema de Medição de Desempenho. Retirado de (Wohlert et al., 2020).

sistema requer a utilização de KPI específicos, mensuráveis e alinhados com os objetivos organizacionais da mesma. A seleção destes KPI remete para uma das cinco etapas do processo de desenvolvimento de um PMS: identificação de metas e objetivos organizacionais; seleção dos KPI apropriados; estabelecimento de metas de desempenho; recolha e análise de dados de desempenho; comunicação dos resultados às partes interessadas. Um dos principais benefícios da implementação de um PMS é permitir às organizações medir o seu desempenho ao longo do tempo, o que pode ajudar a identificar tendências e padrões nos dados de desempenho. Esta possibilidade, por sua vez, pode ajudar as organizações a tomar decisões informadas sobre a atribuição de recursos e o desenvolvimento de estratégias. Outro aspeto importante de um PMS é proporcionar um meio de alinhar os objetivos organizacionais com os objetivos individuais de desempenho. Ao estabelecer metas de desempenho e estabelecer KPI, as organizações podem ajudar a assegurar que cada empregado está a trabalhar para os mesmos objetivos (Chenhall, 2005; Taticchi et al., 2010).

Contudo, o desenvolvimento de um PMS eficaz não está isento de desafios. Uma das principais exigências consiste em assegurar que os KPI selecionados refletem verdadeiramente as metas e os objetivos organizacionais. É importante selecionar KPI que sejam relevantes e significativos e evitar a dependência excessiva em métricas que podem não fornecer uma imagem completa e real do desempenho organizacional. Este desafio pode ser resolvido a partir da implementação do processo de hierarquização analítica, a desenvolver no Capítulo 2.6 (Chenhall, 2005; Taticchi et al., 2010).

2.5.1.2 Sistema de Monitorização de Desempenho Pessoal

Em sistemas de monitorização de desempenho pessoal os KPI são desenhados com a finalidade de alertar os operadores para a sua performance laboral. Este tipo de sistema permite que qualquer operador seja capaz de diagnosticar visualmente e em tempo real o nível da sua performance e, caso esta esteja aquém do esperado, compreender facilmente em que área está a falhar e porquê. A Figura 16 ilustra um destes sistemas implementado num posto de trabalho de uma pequena e média empresa indiana de manufatura de produtos eletrónicos (Wohlers et al., 2020).

Ambos os sistemas recorrem ao uso de um HMI. Trata-se de um sistema de monitorização de dados que conecta o operador e a respetiva máquina ou linha de produção a partir de uma interface digital interativa, como representado nas Figuras 15 e 16. Este sistema permite que os colaboradores monitorizem os KPI de uma forma visual, simples e interativa, podendo acabar por substituir centenas de botões, luzes, entre outros (Prathima et al., 2020).

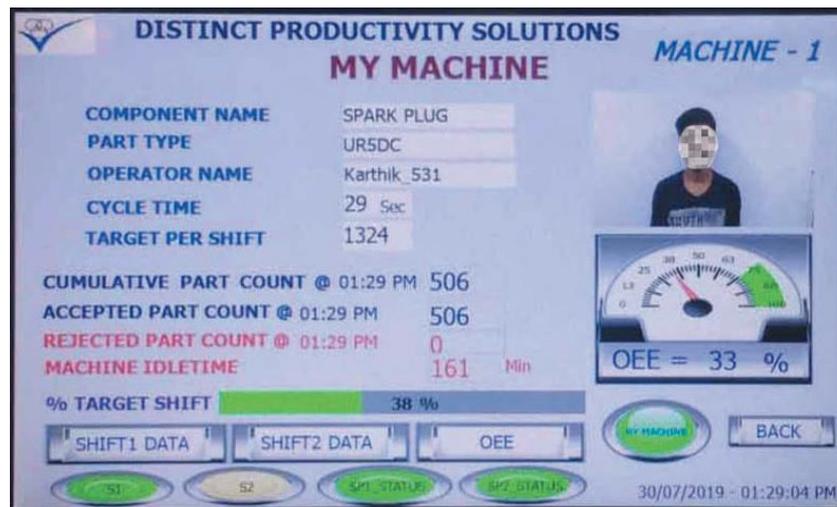


Figura 16 – Monitorização de desempenho pessoal. Retirado de (Prathima et al., 2020).

2.6 Processo de Hierarquização Analítica

2.6.1 Objetivo

A grande variedade e volume de dados gerados diariamente garante a possibilidade de se criarem diversos KPI. Apesar disso, com o acréscimo do seu volume e diversidade a gestão dos mesmos, torna-se mais um desafio. A inexistência de uma hierarquia por parte dos KPI, incapacita qualquer gestor de priorizar certos KPI relativamente a outros, sendo necessário aplicar todas as tarefas analíticas a todos os indicadores com a mesma regularidade e atenção. Este conjunto de parâmetros acaba por atrasar e dificultar a implementação de processos de tomada de decisão. Com o intuito de minimizar estes desafios, as empresas recorrem a métodos de *Multicriteria Decision-Making* (MCDM), *i.e.*, métodos estruturados que permitem modelar problemas complexos de tomada de decisão. O AHP é um dos métodos de MCDM mais utilizados devido à sua flexibilidade e simplicidade (Karasakal & Aker, 2017; Unutmaz Durmuşoğlu, 2018).

A metodologia AHP foi desenvolvida com a finalidade de identificar o peso dos diferentes indicadores de desempenho, permitindo eliminar os desafios causados e evidenciados pela inexistência de uma hierarquia relativa aos KPI.

2.6.2 Funcionamento

O peso de cada indicador é determinado a partir de duas fases (Cristea & Cristea, 2021):

1. Identificar todos os KPI adotados pela empresa – esta etapa pode ser concluída a partir da implementação de um questionário a alguns colaboradores da empresa;
2. Implementar a metodologia AHP.

A primeira fase apenas se aplica às empresas que não possuem todos os seus KPI identificados, podendo as restantes automaticamente aplicar a fase 2.

Identificar os KPI

Esta fase subdivide-se em três etapas (Cristea & Cristea, 2021):

1. Realizar uma revisão literária breve que permita ao colaborador ficar familiarizado com os KPI mais utilizados na área em questão. Desta forma, o colaborador é capaz de comparar os indicadores adotados pela empresa com os KPI mais recomendados.
2. Efetuar uma entrevista com todos os peritos que utilizam, diariamente, indicadores de desempenho, onde será discutido quais dos KPI registados na etapa anterior são analisados pela empresa e, relativamente às ausências, discutir a possibilidade de os implementar. Após cada entrevista um leque de KPI a implementar e a manter na empresa será registado.
3. Elaborar e aplicar um questionário que permita validar a implementação ou manutenção de todos os indicadores de desempenho previamente registados. Este questionário é dirigido a todos os peritos que participaram na fase 2.

Analytical Hierarchy Process

Por último, após a seleção do conjunto de KPI a monitorizar pela empresa, aplica-se o método AHP de forma a hierarquizá-los. A metodologia é subdividida em cinco etapas (Alaqael & Suryanarayanan, 2018; Cristea & Cristea, 2021; Elhuni & Ahmad, 2017; Li et al., 2018):

1. Definir qual o *business problem* da empresa – é registado o objetivo da empresa, os critérios e as variantes. Em contexto industrial, o objetivo é, geralmente, avaliar o desempenho operacional das diversas áreas de operação. Os critérios são o conjunto de KPI a hierarquizar e as variantes são áreas de desempenho, como as que se encontram representadas na Figura 14.
2. Construir a matriz de avaliação – esta matriz quadrada compara a importância de cada KPI face a cada uma das áreas de desempenho monitorizadas na empresa. A escala de importância vai de 1 a 9 como demonstrado na imagem que se encontra no anexo A da dissertação. Após a construção da matriz, é determinada a média geométrica para cada um dos KPI, a partir da equação 1. A variável n representa o número de KPI presentes na matriz e a variável x representa a importância de um dado KPI relativamente a outro.

$$\text{Média Geométrica} = (x_1 * x_2 * x_3 * \dots * x_n)^{(1/n)} \quad (1)$$

3. Formular o vetor de peso – cria-se o vetor peso a partir da equação 2, que é aplicada para cada KPI, onde λ representa o maior *eigenvalue* da matriz, x a importância média do KPI em questão e K o seu peso. Cada elemento do vetor corresponde ao peso de um KPI.

$$K = \lambda \times x \quad (2)$$

4. Verificar se existe conflito entre as valorações determinadas – todas as comparações entre pares de critérios têm de ser avaliadas face à presença de conflito. A ausência de conflito ocorre quando CR, resultante da equação 3, é inferior a 10%. Caso o CR seja superior a 10%, toda a comparação tem de ser repetida (etapa 1). A variável n representa o número de KPI presentes na matriz.

$$CR = \frac{\frac{\lambda - n}{n - 1}}{\frac{1,98 \times (n - 2)}{n}} \quad (3)$$

5. Calcular o ranking final – o rank de cada KPI é obtido a partir da multiplicação da sua média geométrica pelo correspondente valor presente no vetor de peso e, finalmente, da comparação dos KPI entre variantes.

3. CONTEXTO EMPRESARIAL DA DISSERTAÇÃO

3.1 Continental Mabor

No mundo atual, em rápida transformação, a indústria automóvel desempenha um papel crucial na configuração do panorama de transportes. No coração de cada veículo encontra-se um componente fundamental que o liga à estrada: o pneu. Um pneu bem concebido, fiável e de elevado desempenho pode fazer toda a diferença para garantir uma experiência de condução segura e confortável.

No domínio do fabrico de pneus, a Continental Mabor destaca-se como uma empresa portuguesa pioneira que se estabeleceu como líder no sector. Com uma história rica que abrange várias décadas, a Continental Mabor tem demonstrado consistentemente o seu compromisso com a inovação, qualidade e sustentabilidade, impulsionando-a para a vanguarda do mercado global de pneus. Fundada em 1989 como uma *joint venture* entre o grupo Continental AG, uma empresa alemã de renome no sector automóvel, e o Grupo Mabor, um fabricante português de pneus, a Continental Mabor emergiu como uma força significativa no panorama competitivo. Através da colaboração entre estes dois gigantes da indústria, a Continental Mabor conseguiu combinar com sucesso a experiência da engenharia alemã com a manufatura portuguesa, resultando numa marca que simboliza a excelência na produção de pneus.

As instalações de fabrico de última geração da empresa, localizadas em Lousado, Portugal, incorporam tecnologia de ponta, assegurando os mais elevados padrões de qualidade e precisão. O compromisso da Continental Mabor para com a investigação e o desenvolvimento produziu avanços notáveis no design dos pneus, permitindo à empresa satisfazer as exigências em evolução dos sectores dos veículos de passageiros e comerciais. Esta dedicação à inovação garantiu à Continental Mabor um lugar de destaque entre os fabricantes de pneus mais respeitados e procurados do mundo. Para além disso, o foco inabalável da Continental Mabor na sustentabilidade distingue-a dos seus concorrentes. Como as preocupações ambientais continuam a ganhar destaque a nível global, a empresa tomou medidas proactivas para minimizar a sua pegada ecológica. Ao implementar práticas de fabrico amigas do ambiente e ao promover ativamente a utilização de materiais sustentáveis, a Continental Mabor não só contribui para um futuro mais verde, como também melhora a reputação da sua marca enquanto cidadã empresarial responsável.

A eficiência e a otimização tornaram-se fatores cruciais para as empresas que operam no ambiente empresarial, altamente competitivo dos dias de hoje. À medida que a tecnologia continua a avançar, as organizações estão constantemente a procurar formas inovadoras de melhorar os seus processos, reduzir custos e aumentar a produtividade. A Continental Mabor, um fabricante português de

pneus de renome, reconhece a importância destes fatores e pretende racionalizar as suas tarefas de produção para manter a sua posição de líder na indústria.

3.2 Processo Produtivo – *Passenger Car & Light Truck Tires*

Como referido no subcapítulo anterior a empresa em questão subdivide o tipo de pneus que produz em duas classes:

1. *Passenger Car & Light Truck Tires* (PLT) – Pneus indicados para viaturas de passageiros e comerciais ligeiros.
2. *Commercial Specialty Tires* – Pneus indicados para veículos *off-road* como é o caso de máquinas agrícolas, máquinas industriais, camiões de portos e camiões de minas.

Esta tese irá focar-se nos dados gerados por uma operação que pertence ao processo produtivo de pneus PLT. O processo produtivo que dá origem aos pneus PLT é subdividido em 5 etapas: misturação; preparação; construção; pintura e vulcanização; inspeção final.

Misturação

O processo produtivo inicia-se nesta etapa, onde todos os compostos do pneu como a borracha, pigmentos ou a sílica são misturados, por forma a alcançar-se uma borracha mais rica que possui as propriedades químicas e físicas necessárias à etapa seguinte.

Preparação

Na preparação produzem-se os diversos tipos de elementos que compõe o esqueleto do pneu: talões, pisos e as paredes. Estes componentes são feitos recorrendo a três principais máquinas; a extrusoras, a calandras e a máquinas de corte.

Construção

Tendo todos os elementos sido transportados para a área de construção, dá-se início à terceira fase onde se constrói a carcaça do pneu, recorrendo aos talões, camada, tela e paredes como matérias-primas. De seguida constrói-se o pneu que, já construído, receberá o seu acabamento a partir da “união” entre ele e a carcaça. Este “casamento” resulta da insuflação da carcaça dentro do pneu, esta cresce até embater no conjunto do pneu.

Pintura e Vulcanização

Na quarta etapa, é finalmente dada a forma e cor pretendida ao pneu, a partir da utilização de moldes e da combinação da ação da temperatura, pressão e tempo. É importante salientar que a partir

desta fase a borracha que caracteriza o pneu passa de um estado plástico moldável para um estado elástico. Esta nota é importante, pois é após esta transição de propriedade que o pneu deixa de poder ser reintroduzido no processo produtivo.

Inspeção Final

O processo produtivo cessa nesta fase onde cada um dos pneus produzidos é analisado de modo a verificar-se cada um dos parâmetros e características do pneu. Esta análise é realizada de maneira que o departamento de qualidade da empresa consiga garantir a presença dos requisitos especificados pelo cliente.

3.3 Máquinas Extrusoras

O projeto desta dissertação centra-se na segunda fase do processo produtivo, a preparação. O ETL desenvolvido em contexto empresarial, o AHP, os relatórios iterativos e a alarmística desenvolveram-se em torno de dados gerados por duas das várias máquinas que pertencem a este processo: a extrusora 7 e 9. A principal diferença entre ambas assenta no produto semiacabado que são capazes de originar. A extrusora 7 produz apenas perfis de paredes, enquanto a extrusora 9 é capaz de produzir perfis de pisos ou paredes.

3.3.1 Paredes e Piso

As paredes e o piso são elementos vitais de um pneu, cada um com funções únicas e contribuindo para o seu desempenho global. As paredes, localizadas abaixo do piso, ligam o piso ao talão respetivo e desempenham um papel crucial no fornecimento de integridade estrutural, rigidez da parede e proteção contra danos externos. Contribuem, significativamente, para a capacidade de carga e a estabilidade do pneu. O piso, por outro lado, é a parte do pneu que entra em contacto direto com a superfície da estrada. O seu principal objetivo é garantir uma ótima aderência, tração e desempenho global. Ao canalizar eficazmente a água, a lama e os detritos para longe da área de contacto do pneu, o piso mantém o máximo contacto com a estrada, melhorando a estabilidade, a manobrabilidade e a capacidade de travagem do veículo. A figura 17 representa um corte de secção transversal de um pneu após a vulcanização ser finalizada, onde é possível analisar a localização de ambos os elementos (Király et al., 2022).

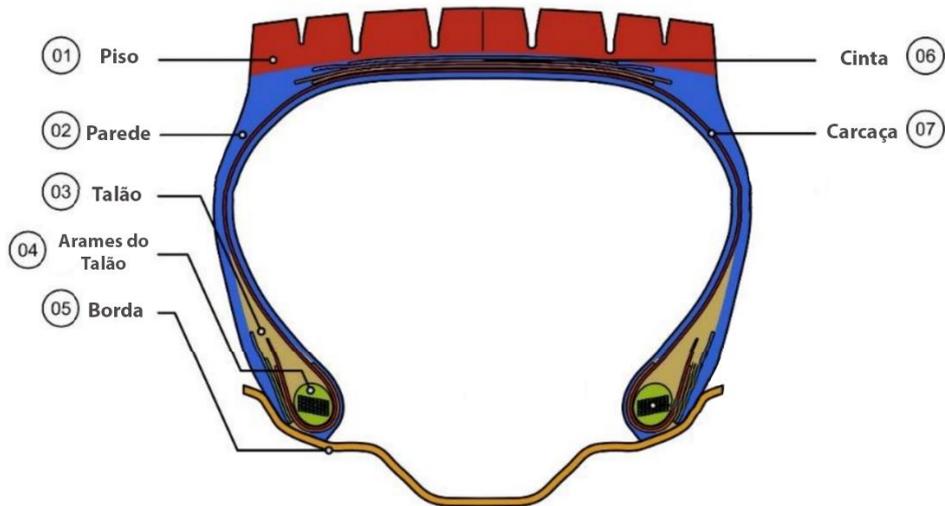


Figura 17 - Corte de secção transversal de um pneu. Adaptado de: (Király et al., 2022)

3.3.2 Extrusão

O fabrico destes componentes realiza-se a partir da extrusão, um processo que envolve o fluxo controlado de um ou mais compostos de borracha através de uma extrusora, uma máquina especialmente concebida para o efeito. A extrusão permite que os componentes da parede e do piso sejam moldados e formados com precisão, assegurando a consistência e a qualidade da sua produção. No caso concreto da Continental Mabor é utilizada uma linha de co-extrusão, denominada triplex, que possui três extrusoras. As combinações simultâneas destas extrusoras permitem à empresa ter a capacidade de extrudir três compostos de borracha diferentes em simultâneo, permitindo a criação de várias camadas ou zonas dentro do piso ou da parede lateral. Na figura 18 tem-se ilustrado um exemplo de um perfil de piso extrudido, constituído por três compostos de borrachas distintos: um no piso propriamente dito, um outro na zona do ombro e um terceiro na base do piso, a capa. Estas camadas

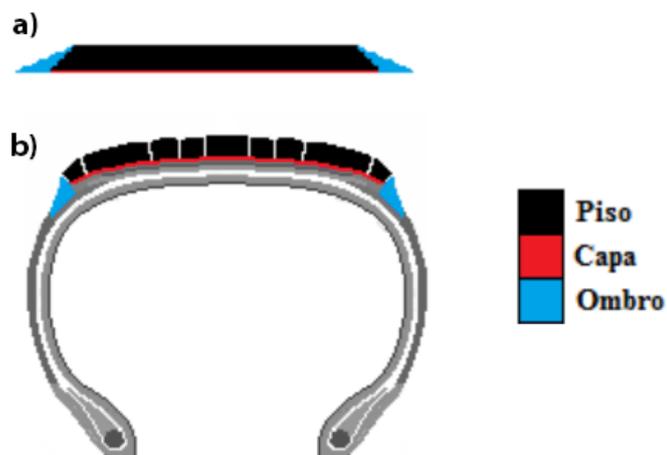
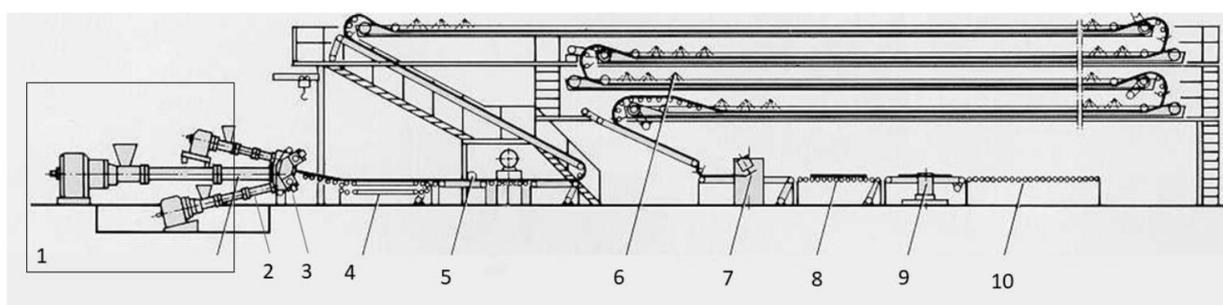


Figura 18 - (a) Corte de secção transversal do piso extrudido; (b) Corte de secção transversal de um pneu. Adaptado de: (Caetano, 2019)

podem ter propriedades e funções diferentes, resultando em características personalizadas que garantem um desempenho superior ao produto final (Caetano, 2019; Gent & Walter, 2006).

Na figura 19 tem-se esquematizado um exemplo de uma linha de co-extrusão triplex semelhante às linhas presentes na Continental Mabor. A partir da figura é possível inferir que uma das etapas deste subprocesso tem como objetivo monitorizar as medidas do perfil. É a partir desta etapa que as extrusoras 7 e 9 permitem obter informação relativamente às dimensões finais do perfil, sendo este um piso ou uma parede. Estas dimensões podem variar de ordem para ordem produção, existem diferentes tipos de perfis por onde optar, cada um com características distintas que se identificam em propriedades como a espessura, largura e área do perfil.



- | | |
|---------------------------------------|--|
| 1- Zona de alimentação | 6- Tanques de arrefecimento |
| 2- Extrusoras | 7- Corte ao comprimento |
| 3- Cabeça da co-extrusora | 8- Passadeira de sopradores de secagem |
| 4- Passadeira de relaxação de tensões | 9- Câmara de Verificação das dimensões finais. |
| 5- Marcação de linhas | 10- Armazenamento automático |

Figura 19 - Linha de co-extrusão triplex. Adaptado de: (Costa, 2022)

O *setup* que permite alterar estas dimensões de ordem para ordem é, principalmente, alcançado a partir da alteração de um elemento na linha de co-extrusão, nomeado de *fieira*. A *fieira*, utilizada pelas extrusoras na etapa 3 do subprocesso representado na figura 19, é uma espécie de molde que permite que o composto de borracha seja extrudido com as dimensões pretendidas. Na figura 20 ilustra-se um exemplo de uma *fieira* utilizada para originar um piso.



Figura 20 - Fieira utilizada para a produção de perfis de pisos. Retirado de: (Costa, 2022)

3.3.3 Perfilômetro

Como referido no capítulo anterior, ao longo da linha de co-extrusão é realizada a monitorização das medidas dos perfis extrudidos. Esta monitorização é vital para para a empresa, porque permite ao departamento de qualidade ser proativo ao ponto de detetar possíveis falhas no produto semi-acabado, permitindo diminuir a produção de *scrap*. Na extrusora 7 e 9 esta monitorização é realizada a partir de um perfilômetro. Como demonstra a figura 21, o perfil, já com a fisionomia desejada, é transportado ao longo de uma passadeira para que possa ser armazenado de modo a mais tarde ser utilizado como matéria-prima no processo a jusante. Antes deste armazenamento ocorrer o perfil é transportado pelo interior de um perfilômetro que, a partir de uma arquitetura de aparelhos e *software* tecnológicos, mede e calcula diferentes métricas do perfil em análise.

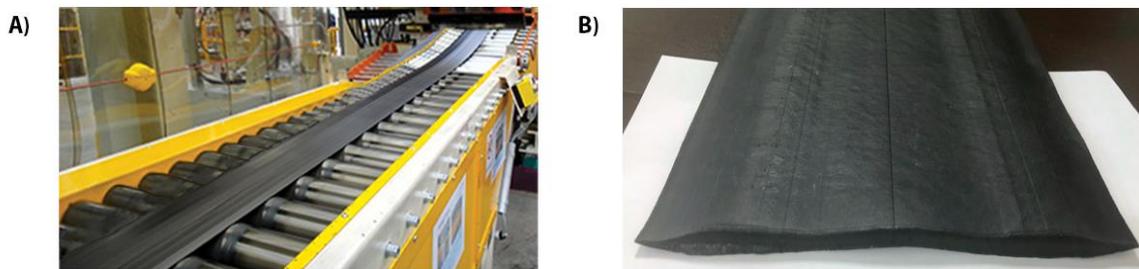


Figura 21 - (A) Perfil já finalizado a ser transportado por uma passadeira; (B) Perspetiva transversal de um perfil de paredes. Adaptado de: (NEXEN TIRE, 2017)

A figura 22 esboça um nível simplista a arquitetura deste perfilômetro. O elemento A representa a passadeira, o B representa o perfil, o C um conjunto de “lâmpadas” e o D as diversas câmaras de verificação das dimensões. As câmaras geralmente em intervalos de dez segundos realizam a emissão daquele feixe de luz vermelho, que sinaliza que nesse instante foi realizada a medição dos parâmetros desejados do perfil. A medição, como a figura demonstra, é realizada sobre o eixo transversal do perfil, naquele instante. O eixo medido encontra-se representado na figura, pela linha vermelha.

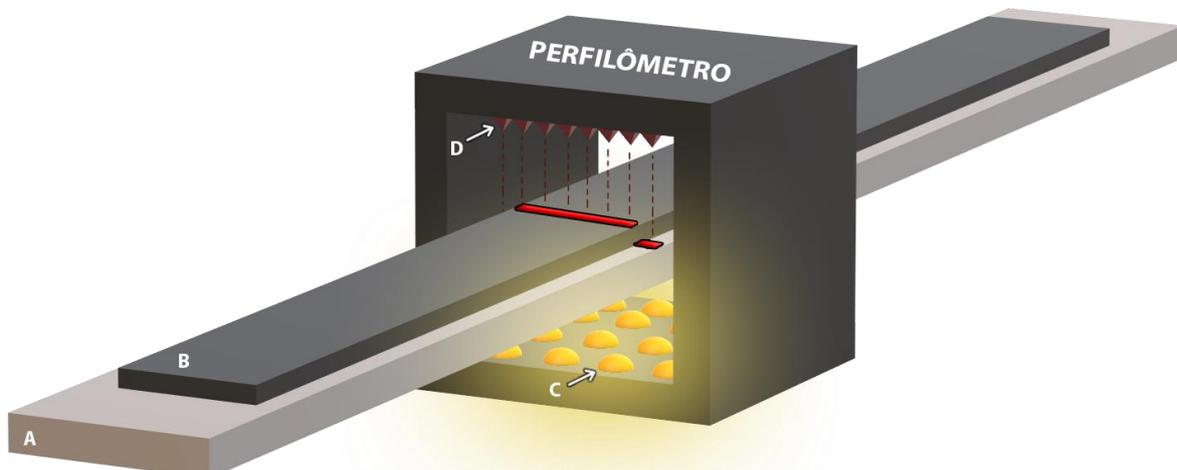


Figura 22 - Esboço figurativo de um perfilômetro utilizado em uma linha de co-extrusão.

4. DIAGNÓSTICO DE PROBLEMAS E PROPOSTA DE SOLUÇÕES

4.1 Estado Inicial

O processo de fabrico de pneus da Continental Mabor envolve uma multiplicidade de etapas interligadas, que vão desde a aquisição de matérias-primas até à montagem do produto final. Cada etapa gera uma grande quantidade de dados, incluindo estatísticas de produção, medições de controlo de qualidade e métricas de desempenho do equipamento. No entanto, o desafio reside em aproveitar eficazmente estes dados e utilizá-los para tomar decisões informadas, identificar estrangulamentos e promover a melhoria contínua. As extrusoras 7 e 9 não são uma exceção, ambas geram continuamente dados ao longo de cada dia, dados estes captados e processados pelo perfilômetro anteriormente referido.

Por cada ordem de produção é gerada uma pasta com diversos ficheiros do tipo TSX e apenas um ficheiro do tipo NOX. Como referido anteriormente, de dez em dez segundos (parcialmente) as câmaras do perfilômetro analisam as dimensões do perfil associado à ordem de produção em vigor, nesse instante. Cada uma destas iterações gera um ficheiro TSX. Este ficheiro detém todas as métricas medidas e calculadas referentes ao eixo transversal do perfil que foi medido naquele instante. Portanto, o ficheiro TSX é o ficheiro que detém a informação real sobre o perfil produzido. Todos os ficheiros TSX são nomeados a partir do instante em que foi realizada a medição, sendo o formato do tempo: HH-MM-SS. A análise de todos os ficheiros TSX, presentes na pasta gerada pela máquina, permitem ao departamento de qualidade ter uma visão real das métricas do perfil ao longo da sua corrida. De notar que a corrida é o conjunto de todas as medições realizadas, desde a primeira à última, para a mesma ordem de produção. Em média, por corrida, são gerados 97 ficheiros TSX.

O ficheiro NOX, por outro lado, representa as dimensões que o perfil deve possuir ao longo do seu comprimento. Como referido anteriormente, é apenas gerado um ficheiro deste tipo por ordem de produção, o que faz sentido, visto que para cada um dos ficheiros TSX o desejado é que as métricas lá presentes sejam iguais às indicadas no ficheiro NOX. Este ficheiro é internamente nomeado ficheiro nominal porque representa o que se pretende alcançar ao longo da totalidade do perfil.

A figura 23 é ilustrativa da arquitetura com que é gerada e armazenada a pasta que detém os ficheiros. A pasta “mãe”, armazenada localmente num dos servidores da empresa, é nomeada a partir

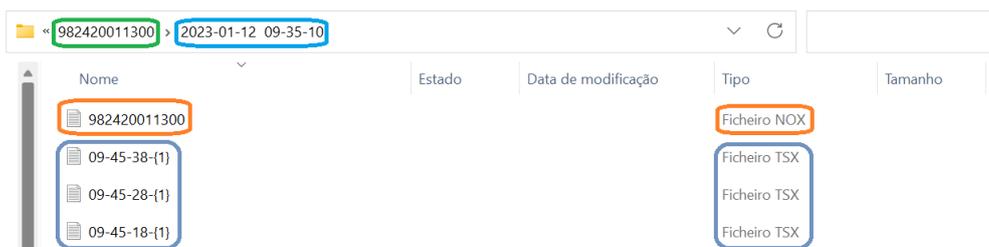


Figura 23 - Arquitetura dos ficheiros gerados pela extrusora.

do código ID escolhido para representar o tipo de perfil produzido. A pasta interna é nomeada com o dia e instante em que se iniciou a ordem de produção. Finalmente, os ficheiros TSX são nomeados a partir do instante em que decorreu a iteração de medição em causa, enquanto o ficheiro NOX é nomeado com o código ID do tipo de perfil.

4.2 Desafios Associados

4.2.1 Dados Não Estruturados

Tanto os ficheiros TSX como os NOX são ficheiros não estruturados. Os dados não se encontram estruturados num formato convencional, podem mesmo ser interpretados como texto bruto.

Os ficheiros TSX são compostos, em média, por mais de 3000 linhas. De todas as linhas, 15 são preenchidas por *strings* (texto), que servem como um separador de um livro. Permitem dividir o ficheiro em quinze partes, onde os dados de cada parte fornecem informação sobre um determinado tema. Os quinze títulos presentes no ficheiro TSX são nomeados da seguinte maneira:

- | | |
|-------------------------|------------------------------|
| 1- Header | 9- More Measures |
| 2- Contour | 10- Perfil Error |
| 3- Dimensions | 11- Perfil Type |
| 4- Date | 12- Piece Size Markers |
| 5- Central Thickness EX | 13- Section |
| 6- Conicity Surfaces | 14- Shoulder Measurements EX |
| 7- Conti Measures AECS | 15- Stripes in Lines |
| 8- Encoder | |

Analisando as restantes linhas, 99% destas são preenchidas por valores numéricos. Para dificultar a situação não existe qualquer tipo de *metadata* que permita decifrar qual o fio condutor usado pelo perfilômetro para estruturar estes dados. Na figura 24 apresenta-se um segmento pixelizado retirado de um ficheiro TSX real. Ao longo da dissertação várias figuras encontram-se pixelizadas, este método foi aplicado para que as regras de confidencialidade da Continental Mabor sejam cumpridas. Analisando a figura podemos inferir que a estrutura dos dados entre a secção *CONTI_AECS_MESSWERTE* (*measured*

```

CONTI_AECS_MESSWERTE
0.0000 0.0000 0.0000
0.0000 0.0000 0.0000
0.0000 0.0000 0.0000
0.0000 0.0000 0.0000
0.0000 0.0000 0.0000
SCHULTERMASSE_EX
-1.000000, 00
-1.000000, 00
-1.000000, 00
-1.000000, 00
0.0000

```

Figura 24 - Segmento pixelizado retirado de um ficheiro TSX real.

values em inglês) e a secção *SCHULTERMASSE_EX* (*shoulder dimension*) é completamente diferente. O número de atributos máximo por linha é distinto, 3 e 1 respetivamente.

Os ficheiros NOX por sua vez possuem em média, 226 linhas, onde 62 são preenchidas por *strings*. No entanto, ao contrário das 15 *strings* presentes no ficheiro TSX que representam cada uma um tema distinto, das 62 *strings* do ficheiro NOX existem *strings* que representam temas, outras subtemas, outras propriedades do perfil e outras em que se é mesmo incapaz de detetar o que representam, como por exemplo a string “YJR”. A complexidade e ausência de estrutura é mais acentuada neste tipo de ficheiro. De todas as 62 *strings*, apenas 34 foram assinaladas como delimitadores de secções.

A não estruturação dos ficheiros impossibilita o uso de métodos de tratamento e carregamento de dados convencionais, levando a que os dados permaneçam não só invisíveis como armazenados localmente, fora da DW da empresa.

4.2.2 Ineficácia e Limitações do Software de BI

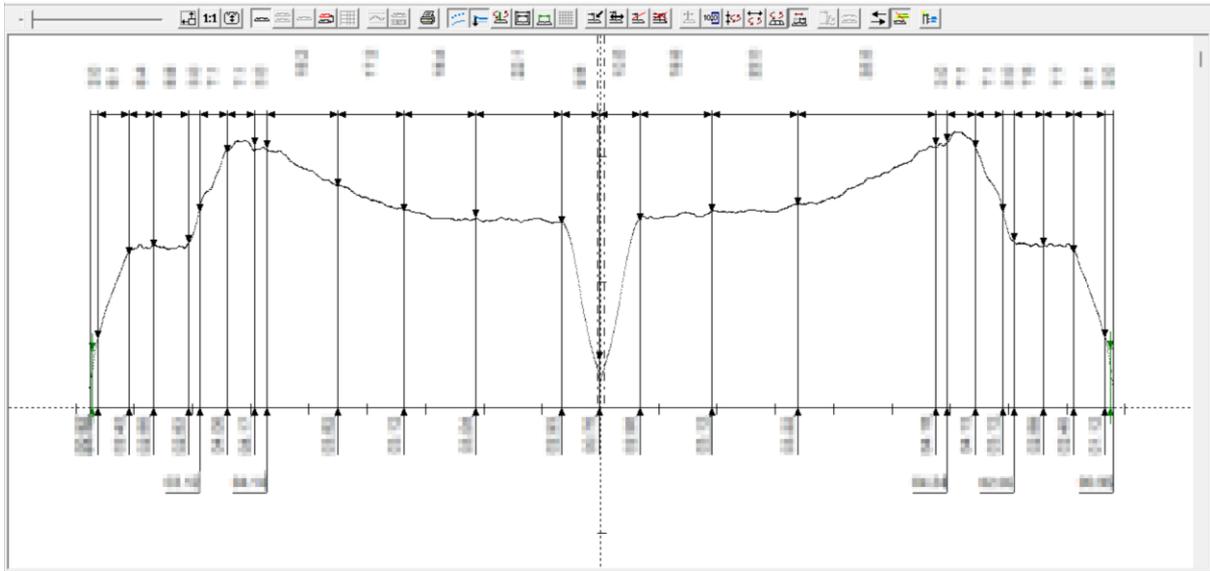
Apesar da inexistência de uma estruturação convencional por parte dos ficheiros, a empresa que desenvolveu o perfilômetro, desenvolveu, também, um *software* capaz de processar os ficheiros TSX e NOX. O *software* é capaz de:

1. Expor graficamente a largura e espessura ao longo do eixo transversal do perfil (real ou nominal);
2. Determinar um leque específico de KPI;
3. Comparar o perfil real com o respetivo nominal.

As diversas funcionalidades do *software* encontram-se evidenciadas na figura 25, 26 e 27.

A figura 25, secção A, apresenta a 1ª funcionalidade: o eixo transversal de um perfil de paredes real encontra-se representado graficamente. Este eixo corresponde a um dado ficheiro TSX que fora

A)



B)

EL-TRISCAN offline : Product parameters

Product-no.

General data

Denotation

Date of production

Preformer no.

Revision no.

Operator

Measured at 19.04.2023 17:48

Modified at 19.04.2023 17:48

Production machine ?

Contour Stretching (Porosity) 1.800

Mixtures

Compounds	Material	Meter weight [g/m]	Spec. Weight [kg/dm ³]	Cross-section [mm ²]
CAP	<input type="text"/>	0.0	0.000	0.0
WINGS	<input type="text"/>	0.0	0.000	0.0
BASE	<input type="text"/>	0.0	0.000	0.0
UP	<input type="text"/>	0.0	0.000	0.0
Mix5	<input type="text"/>	0.0	0.000	0.0
Formula		Calculated	0.0	0.000

Cross-sect. Widths

	NOMINAL	REAL	
Total width	<input type="text" value="280.0"/>	<input type="text" value="263.4"/>	[mm]
Width left	<input type="text" value="0.0"/>	<input type="text" value="131.5"/>	[mm]
Width right	<input type="text" value="0.0"/>	<input type="text" value="131.9"/>	[mm]
Total cross-section	<input type="text" value="773.0"/>	<input type="text" value="696.7"/>	[mm ²]
Cross-sec. left	<input type="text" value="386.5"/>	<input type="text" value="348.3"/>	[mm ²]
Cross-sec. right	<input type="text" value="386.5"/>	<input type="text" value="348.4"/>	[mm ²]
Cross section diff. (left - right)	<input type="text" value="0.0"/>	<input type="text" value="-1.1"/>	[mm ²]
Meter weight	<input type="text" value="0.0"/>	<input type="text" value="0.0"/>	[g/m]
Wing's width	<input type="text" value="0.0"/>	<input type="text" value="0.0"/>	[mm]

Symmetry

Cross-sect. diff. (outer) Max. [mm²]

Cross sect. diff. (inner) Max. [mm²]

Conicity mm

Comment

Sidewall profile

Change directory

Figura 25 - Exemplos concretos da 1ª e 2ª funcionalidade do *software*.

importado pelo *software*, o qual representa uma iteração captada pelo perfilômetro. Apesar da pixelização de parte da figura, é possível inferir que o gráfico encontra-se legendando com as métricas do perfil, no que toca à sua espessura e largura. A secção B, apresenta a 2ª funcionalidade: resumo de um leque específico de KPI, como a largura e a área do perfil.

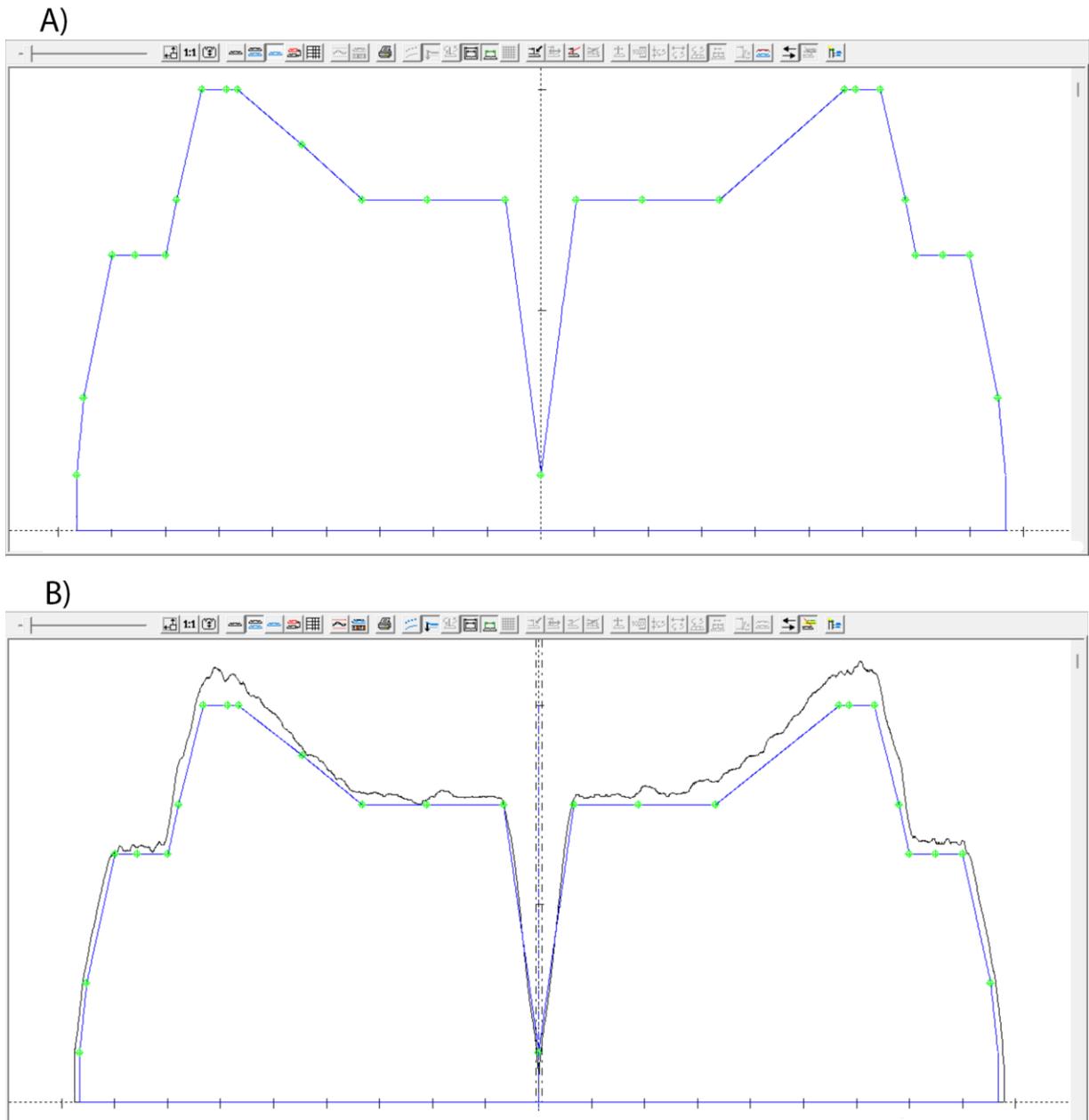


Figura 26 - Exemplos concreto da 1ª e 3ª funcionalidade do *software*, para um perfil de paredes.

A figura 26, secção A, apresenta graficamente um perfil de paredes nominal, de um modelo produzido regularmente em chão de fábrica. Este output do *software* recorre à importação de um ficheiro NOX, ficheiro que detém a informação do perfil nominal. Como referido anteriormente, é a partir deste perfil que o colaborador entende o que é pretendido alcançar: a largura do perfil, a espessura ao longo do eixo e uma simetria perfeita entre o lado esquerdo e direito do perfil.

A secção B da figura representa a 3ª funcionalidade do *software*, a comparação do perfil real para uma dada iteração e o seu perfil nominal correspondente. Utiliza-se o mesmo tipo de perfil da secção A. A partir do gráfico é possível inferir rapidamente que ao longo do eixo, a espessura do perfil é superior ao que é pretendido pelo departamento de produção e qualidade.

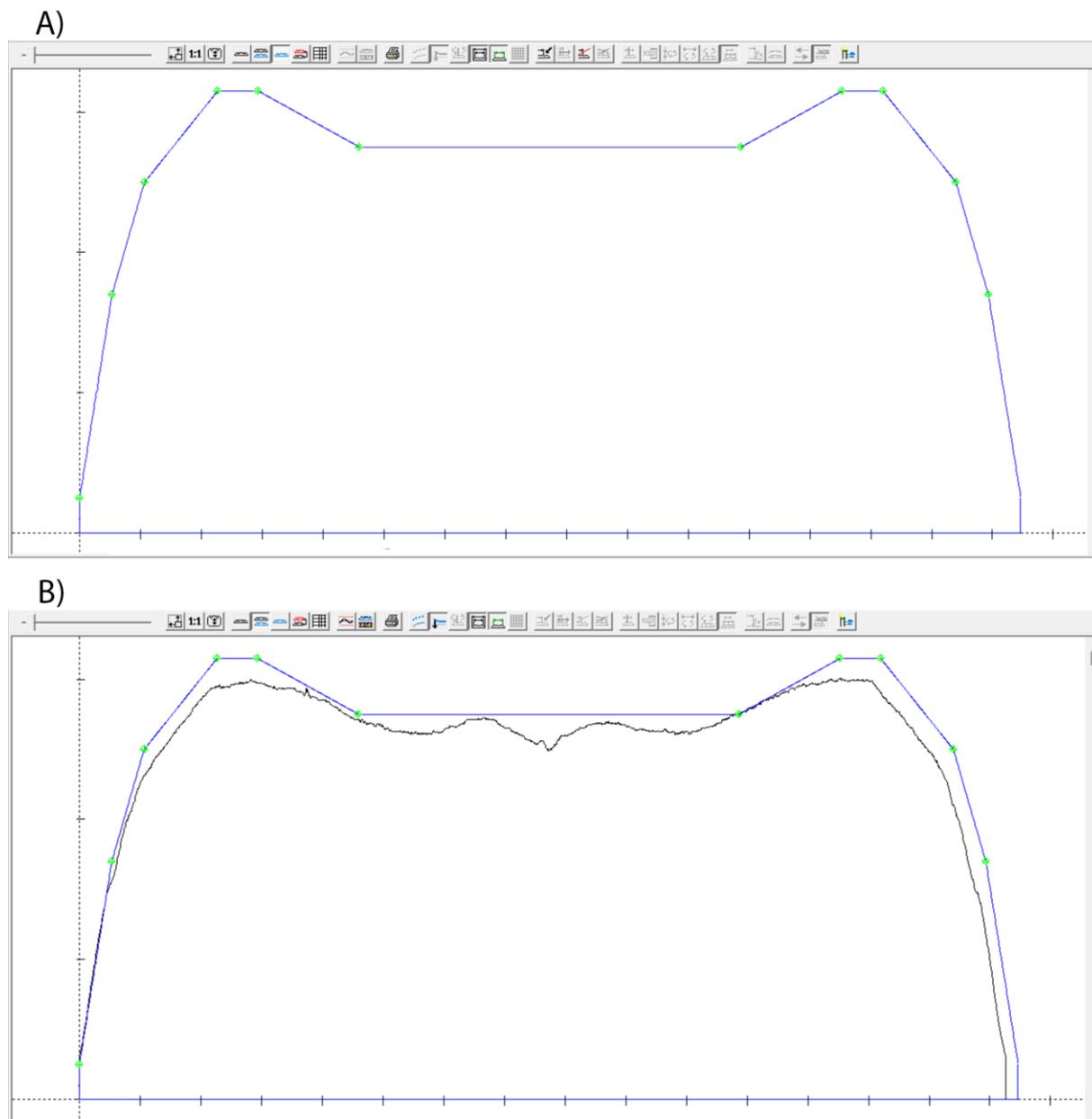


Figura 27 - Exemplos concreto da 1ª e 3ª funcionalidade do *software*, para um perfil de pisos.

A figura 27 é em tudo semelhante à figura 26, exceto na classe e tipo de perfil que representa. Enquanto na figura 25 é ilustrado um perfil de paredes, aqui é ilustrado um perfil de pisos. A diferença mais notável entre as classes de perfis apresenta-se na presença de uma espécie de “vale” no meio do eixo do perfil de paredes.

Apesar destas funcionalidades serem úteis, existem várias limitações que impedem este *software* de se tornar capaz de satisfazer os requisitos da empresa:

1. Não permite importar de uma só vez todos os ficheiros TSX que compõe a corrida, levando a que seja impossível ter uma visão geral do perfil na sua totalidade. Isto é, todos os KPI e o gráfico remetem apenas para os dados de uma única interação;

2. Não permite configurar novos KPI a partir do enorme conjunto de dados presentes nos ficheiros;
3. Possui KPI fixos, imutáveis e simplistas;
4. Incapaz de exportar os dados em formato convencional, após o processamento dos mesmos, de modo que seja possível importá-los para a DW;
5. A importação é manual, morosa e arcaica.

O departamento de qualidade, principalmente devido à 1ª limitação, olha a este *software* como sendo obsoleto.

4.2.3 Inexistência de Monitorização e Alarmísticas

Existe uma elevada correlação entre o *scrap* produzido e a qualidade dos perfis de paredes e pisos. No entanto, face à invisibilidade dos dados, é impossível para a empresa desenvolver uma monitorização em tempo real das métricas dos perfis produzidos. A inexistência desta monitorização torna impraticável a implementação de um sistema de alarmes. A ausência de alarmística incapacita o departamento de qualidade de ser proativo ao ponto de detetar uma falha no perfil, estando este ainda ativo no processo produtivo, a tempo de conseguir emitir alguma ordem corretiva que possa reaproveitar o perfil ou até mesmo corrigi-lo. Em suma, é do interesse da empresa assegurar que todos estes dados passem a ser visíveis, interpretáveis e monitorizados “em tempo real”, o mais possível.

4.3 Soluções Propostas

O presente projeto, de modo a ultrapassar os desafios citados no subcapítulo anterior, contemplou estratégias diversas, evidenciadas na tabela 6.

Tabela 6 - Soluções implementadas para cada problema identificado no diagnóstico inicial.

Desafios					Soluções Implementadas
Dados Não Estruturados	Invisibilidade dos Dados	Ineficácia e Limitação por parte do Software de BI	Inexistência de Monitorização	Inexistência de Alarmísticas	
X					Realização de engenharia inversa a partir do software do perfilômetro, por forma a identificar os atributos presentes em cada uma das 15 secções do ficheiro.
X	X				Desenvolvimento de um ETL utilizando python que recebe como input os ficheiros TSX, NOX e a informação resultante da engenharia inversa. O ETL permite armazenar os dados em formato tabelar na DW da empresa.
	X	X	X		Concepção de <i>reports</i> em Power Bi que recebem como input os dados diretamente da DW da empresa. Estes, permitem monitorizar e analisar interativamente as métricas dos perfis.
				X	Produção de alarmísticas capazes de monitorizar estados críticos à eficácia e eficiência da empresa. Recorreu-se a um software desenvolvido internamente pelo departamento "Aplicações e Processos".

5. DESENVOLVIMENTO DO ETL

Com o intuito de fundamentar o presente capítulo devido à sua complexidade, foi realizado um fluxograma, ilustrado na figura 28, que representa todas as etapas presentes no ETL desenvolvido no projeto e, em vigor na Continental Mabor.

A etapa inicial de qualquer processo de desenvolvimento de um ETL passa pela decisão da(s) ferramenta(s) a usar para a construção do mesmo. Face à complexa natureza do ficheiro, ao seu formato e à estrutura interna pouco convencional dos dados, é necessário optar por uma ferramenta capaz de transformar e manusear dados complexos em estado bruto. O python, uma das linguagens de programação mais usadas a nível mundial, a partir das diversas bibliotecas às quais tem acesso, permite não só satisfazer todos estes requisitos como, também, realizar cada uma das três fases do ETL. Por tal

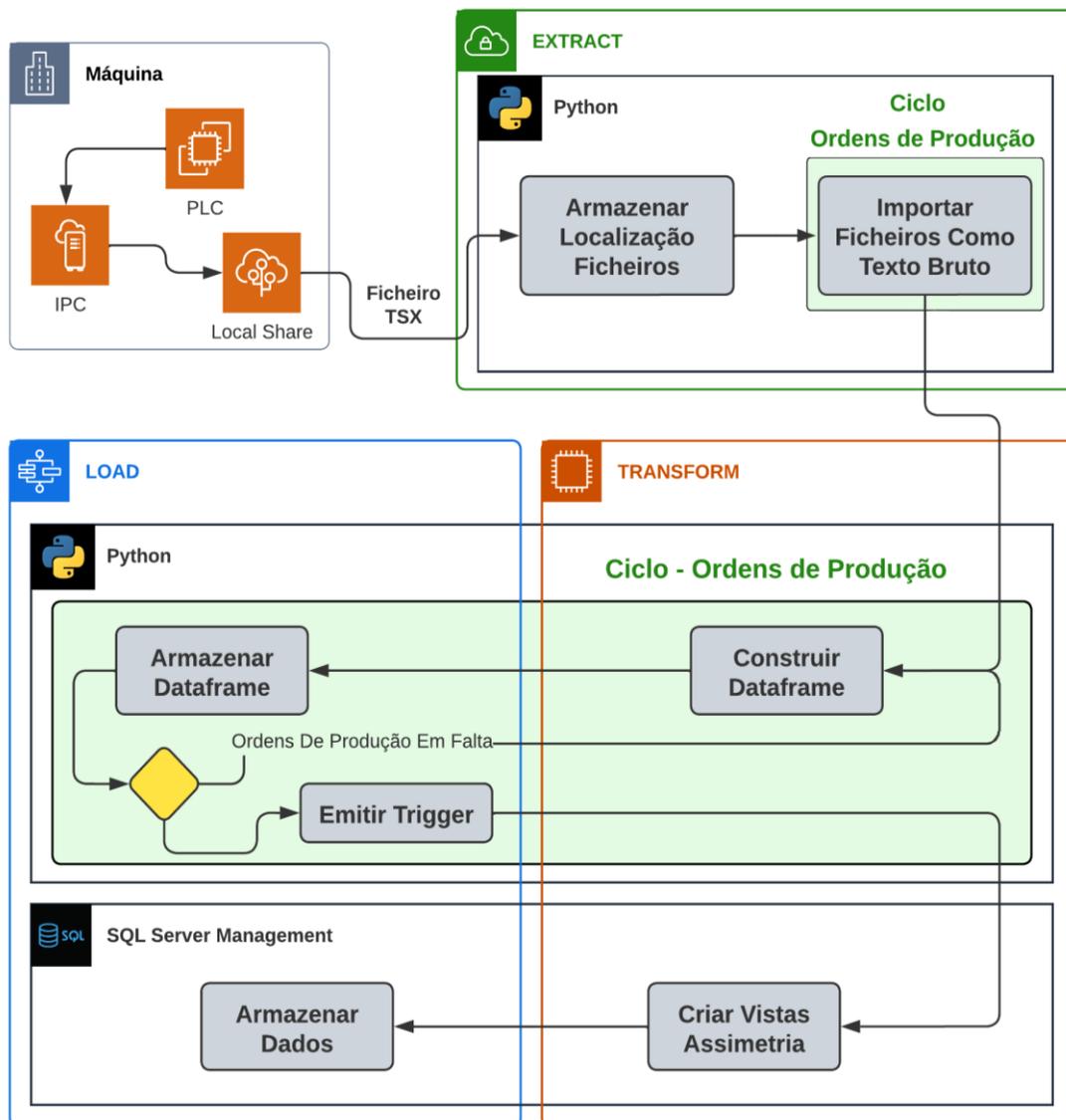


Figura 28 - Fluxograma do ETL.

facto, decidiu-se inicialmente usar o python como meio de centralização, permitindo realizar todas as três fases do ETL a partir da mesma ferramenta.

O ETL consome os ficheiros TSX e NOX a partir de uma pasta local nomeada *Generated Files*, que se encontra armazenada, localmente, no servidor onde o ETL é executado. Esta particularidade irá ser explanada com maior detalhe no subcapítulo 5.6.

5.1 Extração

O primeiro passo do ETL consiste em percorrer todas as diretorias presentes na pasta *Generated Files* e armazená-las em uma de duas listas. Caso seja um ficheiro TSX é armazenado numa, caso seja um ficheiro NOX é armazenado noutra. O objetivo deste passo é simples: conseguir armazenar cronologicamente e por ordem de produção a diretoria de todos os ficheiros de um dado tipo, num só local. Posteriormente, ambas as listas sofrerão tratamentos distintos, como ilustrado na figura 29.

1. A lista que detém os ficheiros TSX será subdividida tendo em conta a ordem de produção do ficheiro, i.e. existirá no final da divisão uma lista por cada ordem de produção. A ordem de produção é obtida a partir da junção dos nomes das primeira e segunda pastas às quais o ficheiro TSX pertence, como exemplificado na figura 23;
2. Já a lista que detém os ficheiros NOX não sofrerá qualquer tipo de subdivisão. Ordens de produção distintas que pretendam produzir o mesmo tipo de perfil, possuem ficheiros NOX iguais. Como não se pretende ter ficheiros duplicados na DW, em vez de se subdividir a lista, serão eliminadas todas as diretorias que armazenem ficheiros NOX que representem duplicados. Para tal, a lista NOX é analisada quanto ao tipo de perfil, obtido a partir do nome da primeira pasta, como se verifica na figura 23.

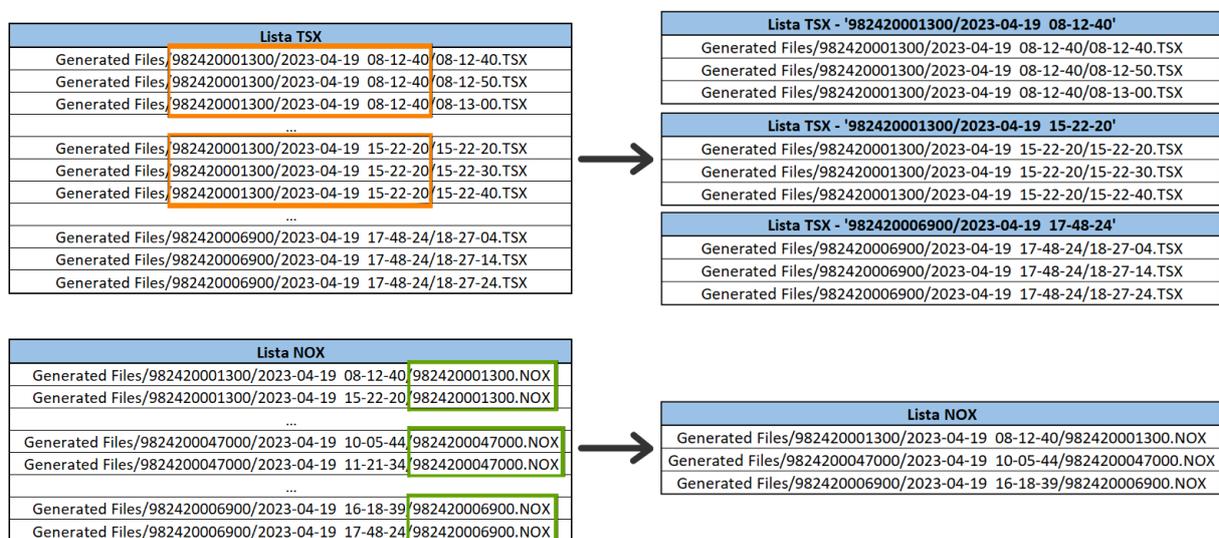


Figura 29 - Processamento associado a cada tipo de lista.

1. Por forma a tornar o texto mais legível foram removidos todos os caracteres especiais e desnecessários presentes no texto, como o 'x00';
2. Para se separar o texto em diferentes linhas como representado na figura 28, foi necessário realizar um estudo prévio do ficheiro que permitisse descobrir qual era o carácter que servia como delimitador entre linhas. Conclui-se que o carácter usado era o parágrafo, representado em texto bruto por '\r'. Indicando-se o '\r' como delimitador subdividiu-se, então, a *string* total em várias *strings* mais pequenas, cada uma a representar uma linha do ficheiro. Estas *strings* foram todas armazenadas numa lista.

Como anteriormente citado, os ficheiros TSX subdividem-se em quinze secções e os NOX em trinta e quatro, no que respeita ao tema que as suas métricas representam. Assim, tomou-se a decisão de separar a lista de *strings* do ficheiro TSX em quinze listas mais pequenas e a do ficheiro NOX em trinta e quatro, separando o conteúdo do ficheiro pela sua raiz. Para tal, decidiu-se utilizar um dicionário em python. Cada uma destas secções é identificada por uma *string* que representa o título do campo, como por exemplo a *string* *KONTUR*, contorno em alemão, que indica qual a secção do ficheiro que está relacionada com o contorno do perfil, no ficheiro TSX. Como se sabe de antemão o nome de cada um destes títulos, a estratégia implementada foi criar um dicionário em python, em que para cada título possível corresponde um elemento do dicionário. No dicionário, esse elemento, armazena o índice (posição) em que o título respetivo aparece na lista. Utilizando essas quinze e trinta e quatro posições como delimitadores para o ficheiro TSX e NOX, respetivamente, é possível dividir a lista do ficheiro TSX em quinze listas mais curtas e, a do ficheiro NOX em trinta e quatro. Cada uma das listas possui informação relativa a um dado campo do perfil. Na figura 31 apresenta-se como exemplo uma das quinze listas criadas para um dado ficheiro TSX, nomeadamente a lista da secção *Date*.

5.2 Transformação – 1ª parte

5.2.1 Etapa 1 – Estruturação dos Dados

5.2.1.1 Engenharia Inversa

Antes de se iniciar a terceira etapa, “Construir Dataframe”, é necessário preencher-se um requisito: identificar a estrutura que os dados presentes nas listas TSX e NOX seguem, e quais os atributos que estes representam. Com vista a preencher estes requisitos foi necessário realizar-se o

```
>>> date
['DATUMDERMESSUNG', '2023', '1', '4', '12', '9', '36', '0', '781']
```

Figura 31 - Lista tratada da secção *date*.

processo anteriormente referido como engenharia inversa. Este processo foi realizado apenas uma vez, previamente à construção do ETL e subdividiu-se em três etapas:

1. Analisar a estrutura interna dos ficheiros, etapa referida nos subcapítulos anteriores, onde se concluiu que os ficheiros não são estruturados, não possuem metadata, podem ser subdivididos em diversas secções;
2. Importar um ficheiro TSX e NOX a partir do *software TRISCAN*, *software* referido no subcapítulo 4.2.2, e realizar-se uma comparação meticulosa entre os valores que são expressos no *software*, estes, já com metadata associada como representado pela figura 25, com os valores armazenados no ficheiro. A partir da metadata do *software* foi possível entender em qual das secções do ficheiro a comparação teria de ser feita, para cada conjunto de dados do *software*. A partir desta comparação vários atributos e estruturas de valores presentes no ficheiro foram capazes de ser identificados;
3. Verificar se estas conclusões se mantinham para ficheiros TSX de outras iterações e ficheiros NOX de outro tipo de perfil. Este estudo presente na etapa três foi realizado para mais de cem ficheiros, nos quais as conclusões da etapa dois permaneceram válidas.

Após a conclusão das três etapas, deu-se por completa a engenharia inversa, onde se alcançou sucesso parcial. O termo sucesso parcial deve-se a dois fatores: apenas se conseguiu identificar os atributos e estrutura dos dados para cinco de quinze secções do ficheiro TSX e uma das trinta e quatro secções do ficheiro NOX, o que representa uma taxa de sucesso de 33% e 3%, respetivamente. Ainda assim, este processo foi encarado como sucesso, pois as seis secções onde a engenharia inversa foi completada com sucesso, abrangem toda a informação necessária para determinar os indicadores de performance que o departamento de qualidade da empresa pretende monitorizar e estudar. Isto é, todos os indicadores que precisam de ser analisados e monitorizados, determinados pelo AHP realizado e a explicar no subcapítulo 5.2.2, são encontrados numa destas seis secções. Na realidade, existem dados invisíveis ainda por identificar, mas estes não são, para já, essenciais à empresa. As seis secções identificadas com sucesso foram a *Header*, a *Contour*, a *Date*, a *Section* e a *Dimensions* para o ficheiro TSX e a *Nominal* para o ficheiro NOX.

Header

A secção *Header* pode ser encarada como uma tabela com dezasseis atributos e uma linha. Os onze atributos, representados na tabela 20, no apêndice A, representam métricas gerais e simples que caracterizam rapidamente o perfil. Temos o exemplo do *Perfil_Dimension_ID* e do

Nominal_Total_Area_mm2, que representam o tipo de perfil e a área que o perfil deveria possuir para ir ao encontro dos níveis de qualidade pretendidos, respetivamente. Os atributos sinalizados a laranja são atributos que pertencem à tabela *header*, mas que resultam do processamento que os dados irão sofrer, logo, representam informação que apenas será adicionada num ponto mais tardio do ETL. O processo em que estes atributos são calculados e o que representam será explicado no subcapítulo 5.2.3.

Contour

A secção *Contour* pode ser encarada como uma tabela com um atributo, representado na tabela 21, no apêndice A, e N linhas. O único atributo desta tabela, *Contour_Thickness_mm*, representa a espessura que o perfil possui ao longo do seu eixo transversal. O eixo representa a largura do perfil e é composto por N pontos. O valor de N depende da largura do perfil; quanto maior a largura maior o número de pontos registados pelo perfilômetro. Em média existem dois mil duzentos e cinquenta pontos por cada iteração captada pelo sensor, cada um destes pontos será representado por uma linha na tabela *contour*. Ao serem traçados estes pontos, sequencialmente, consegue-se alcançar um gráfico semelhante ao presente no exemplo representado na figura 25.

Date

A secção *Date* pode ser encarada como uma tabela com sete atributos e uma linha. Os sete atributos, representados na tabela 22, no apêndice A, representam métricas relacionadas com a data e instante em que o perfilômetro mediu os dados deste ficheiro. O colaborador, a partir desta tabela consegue analisar do nível mais simplista (ano) ao nível mais complexo (milésimas de segundo) em que instante atuou o perfilômetro.

Dimensions

A secção *Dimensions* pode ser encarada como uma tabela com três atributos e n linhas. Ao analisar-se os atributos, presentes na tabela 23, no apêndice A, pode concluir-se que existe uma similaridade de objetivos entre os dados desta secção e da secção *Contour*.

A tabela *dimensions* tem um objetivo semelhante à tabela *contour*, porém apesar de ambas pretenderem registar e ilustrar a variação da espessura ao longo do perfil, a tabela *dimensions* apenas apresenta a espessura para um valor muito reduzido de pontos. Ou seja, enquanto a tabela *contour* possui N linhas, uma para cada ponto monitorizado pelo perfilômetro, sendo este N em média dois mil duzentos e cinquenta, o n é dado pelo atributo *Number_Marks*, que em média é vinte. Esta tabela possui

uma pequena percentagem dos pontos reais do perfil, permitindo representar resumidamente a espessura do mesmo.

Nominal

A secção *Nominal* pode ser encarada como uma tabela com dois atributos e k linhas. Ao serem analisados os atributos, presentes na tabela 24, no apêndice A, pode verificar-se que existe uma correspondência de objetivos entre os dados desta secção, os da *Contour* e os da *Dimensions*. Apesar de todas as tabelas registarem a espessura do perfil ao longo da sua largura, a tabela nominal apresenta a espessura que o perfil deveria ter em alguns dos pontos assinalados pelo perfilômetro. O termo “alguns” é aplicado porque tal como a tabela *dimensions*, também a tabela nominal possui um número reduzido de linhas, em média vinte.

Section

A secção *Section* pode ser encarada como uma tabela com dois atributos e uma linha. O atributo *Left_Width_mm* indica a largura do lado esquerdo do perfil e o *Right_Width_mm* indica a largura do lado direito. Ambos se encontram ilustrados na tabela 25, no apêndice A.

Casos de Insucesso

Em todas as secções restantes, presentes em ambos os ficheiros, não foi possível concluir com sucesso a engenharia inversa. Para o ficheiro TSX, concretamente, todas as secções restantes possuem uma nomenclatura semelhante a uma matriz; todas as linhas da mesma secção possuem o mesmo número de colunas. De modo a não perder a informação, decidiu-se converter cada uma das secções para uma tabela, utilizando-se a seguinte lógica:

1. Todas as secções em que o número de colunas é diferente de um, são convertidas para tabela sem alterar em nada a estrutura, i.e. se a matriz possui três colunas e nove linhas, a tabela final possuirá três atributos e nove linhas. A nomenclatura usada para nomear as colunas é a seguinte: a1, a2, ... , a N. Sendo N o número máximo de colunas.
2. Todas as secções em que o número de colunas é igual a um, são convertidas transpondo a tabela, i.e. se a matriz possui uma coluna e nove linhas, a tabela final possuirá uma linha e nove atributos. A nomenclatura usada para nomear as colunas é a seguinte: a1, a2, ... , a N. Sendo N o número máximo de colunas.

Na figura 32 ilustram-se essas possibilidades, a partir de um exemplo puramente abstrato.

Relativamente ao ficheiro NOX, face à complexidade das secções e à irregularidade estrutural entres estas, optou-se por não converter as mesmas em *dataframe*, as secções nas quais a engenharia

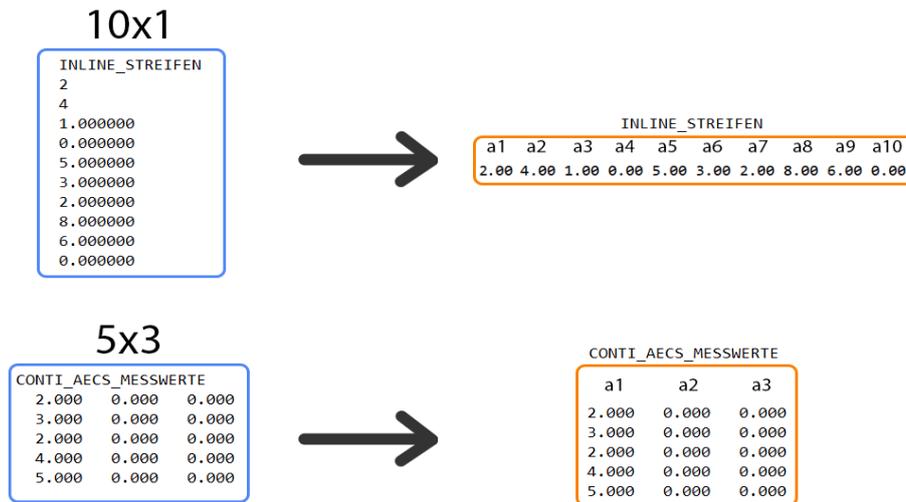


Figura 32 - Secções em que a engenharia inversa não teve sucesso.

inversa não funcionou. Esta decisão também foi sustentada pelo facto de que a única secção do ficheiro NOX onde a engenharia inversa funcionou armazena toda a informação essencial e requerida para calcular os KPI seleccionados no subcapítulo 5.2.2.

5.2.1.2 Construção e Configuração dos Dataframes

Estando as estruturas e os atributos determinados para cada secção é iniciada a terceira fase do fluxograma, “Construir Dataframe”. Esta etapa recorre à livreria *pandas*, onde é utilizada a função *DataFrame* que permite transformar cada lista num *dataframe*, uma estrutura de dados bidimensional, semelhante a uma tabela, que organiza os dados em linhas e colunas.

Após a construção de todos os *dataframes* é necessário definir qual o tipo de dados associados a cada um dos seus atributos. Até ao momento, todos os valores são considerados uma *string*, independentemente da *string* possuir apenas valores numéricos como caracteres. Tendo como base os esquemas de tabelas, presentes em anexo, todos os *dataframes* que representam secções em que a engenharia inversa foi implementada com sucesso são configurados de modo que todos os seus tipos de dados passem a ir ao encontro dos tipos de dados presentes no esquema que lhe corresponde. É importante salientar que o facto dos dados já possuírem o tipo de dados correto irá facilitar o processo de armazenamento dos dados na DW.

5.2.2 Implementação do Processo de Hierarquização Analítica

Finalizada a etapa inicial da fase transformação do ETL, é necessário, de seguida, processar os dados a fim de, facilmente, se determinarem os KPI pretendidos pela empresa. No entanto, como referido na revisão bibliográfica, uma empresa não deve monitorizar todos os KPI passíveis de serem calculados

a partir dos dados gerados. Ao invés disso, as empresas devem otimizar o seu tempo limitado num leque restrito de KPI que englobe os indicadores que mais influenciam a eficiência e eficácia das mesmas.

A seleção deste leque de KPI foi realizada a partir da implementação do AHP. Diversos colaboradores da Continental Mabor participaram neste processo; colaboradores de diversos departamentos: PRODUÇÃO, QUALIDADE E PROCESSOS. De um leque inicial de dez indicadores de performance, representados na tabela 7, apenas quatro validaram os requisitos delineados pelo grupo de colaboradores. Na etapa inicial do AHP foi determinado que apenas se iria monitorizar o conjunto de KPI com maior valoração que representasse $\frac{3}{4}$ do valor de todos os KPI passíveis de determinar a partir dos dados gerados pelas extrusoras. Como ilustrado pela tabela 8, foram selecionados os KPI: diferença entre o peso real do perfil e o peso esperado; assimetria; diferença entre a área real do perfil e a área esperada; diferença entre a espessura real do perfil e a espessura esperada.

5.2.3 Etapa 2 – Processamento dos Dados

Após a seleção de quais os KPI a monitorizar é necessário processar os dados de modo a facilitar o cálculo dos indicadores de performance a realizar a jusante do espetro. Independentemente da ferramenta selecionada que permita criar os relatórios interativos, a ferramenta possuirá mecanismos internos que lhe irão permitir transformar e processar os dados de modo que se calculem diversos indicadores. No entanto, o tipo de transformações que estas plataformas permitem aplicar são mais limitadas comparativamente às que as linguagens de programação como o python ou SQL oferecem. Isto é, o colaborador terá maior facilidade em alcançar transformações complexas a partir de uma linguagem de programação do que a partir, por exemplo, do Power Bi.

Tabela 7 - Leque inicial de KPI em estudo no AHP.

Leque Inicial	
Velocidade da Máquina	Ratio Ficheiros com Denotação ".out"
Diferença Área Real face à Área Nominal	Assimetria
Intervalo entre Medições	Diferença Largura Real face à Largura Nominal
Diferença Espessura Real face à Espessura Nominal	Ratio Ficheiros Sem Indicação do Centro do Perfil
Diferença Volume Real face ao Volume Nominal	Diferença Peso Real face à Peso Nominal

Tabela 8 - Leque de KPI selecionados após o AHP.

KPI Selecionados	Peso	Peso Cumulativo	Rank
Diferença Peso Real face ao Peso Nominal	27%	27%	1
Assimetria	21%	48%	2
Diferença Área Real face à Área Nominal	15%	63%	3
Diferença Espessura Real face à Espessura Nominal	15%	78%	4
Ratio Ficheiros Sem Indicação do Centro do Perfil	9%	87%	5
Ratio Ficheiros com Denotação ".out"	5%	92%	6
Diferença Largura Real face à Largura Nominal	4%	96%	7
Diferença Volume Real face ao Volume Nominal	2%	98%	8
Velocidade da Máquina	1%	99%	9
Intervalo entre Medições	1%	100%	10

Focando no exemplo do indicador “assimetria”, existem várias métricas que podem classificar-se como KPI de assimetria, sendo uma delas o *P-value*, valor numérico resultante do teste *T-test*, a explicar em maior detalhe no decorrer deste capítulo. Caso se optasse por realizar o cálculo desta métrica em Power Bi, seria necessário implementar de raiz todas as equações matemáticas que compõem o teste e, conseqüentemente, teriam de se armazenar todos os resultados provenientes de cálculos intermédios. Este processo seria moroso e passível de originar erros. Utilizando a linguagem python em conjunto com a função *ttest_rel* da livreria *scipy* é possível calcular o *P-value* em menos de um segundo, sem necessidade de realizar qualquer configuração na função. Dado que os KPI selecionados pela empresa apresentam um nível razoável de complexidade, optou-se por realizar o processamento dos dados recorrendo a linguagens de programação.

A linguagem selecionada para desenvolver todo o processamento foi o python, devido ao acesso que este possui no que toca a livrerias que facilitam toda a transformação de dados inerente a este processo. Ainda assim, foi necessário recorrer adicionalmente à linguagem SQL, acabando esta por realizar cerca de 20% do processamento dos dados. A justificação para a ausência da centralização do processamento em apenas uma linguagem, será apresentada à posteriori no subcapítulo 5.2.3.8.

5.2.3.1 Chaves dos Dataframes

Analisando os atributos iniciais dos esquemas das tabelas *header*, *contour*, *dimensions*, *nominal*, *section* e *date*, presentes em anexo, é possível concluir que não existe uma chave comum que permita criar relações de união entre elas. A capacidade de criar uniões entre tabelas é um requisito para o processamento a realizar no ETL, por conseguinte, foram criadas para cada uma das tabelas sete novos atributos.

ID

Este atributo numérico indica em que posição foi inserida a linha na tabela, i.e. se uma dada linha vai ser armazenada numa tabela da DW, que anteriormente se encontrava vazia, o ID dessa linha tomará o valor de um. Este atributo representa uma autoincrementação, que atua sempre que uma linha de dados é armazenada na base de dados. Este processamento é alcançado a partir da funcionalidade de autoincrementação fornecida pelo Microsoft SQL Server, o RDBMS selecionado e usado pela Continental Mabor.

Upload_Time

O presente atributo indica o instante em que a linha de dados em questão foi armazenada na base de dados. À semelhança do atributo anterior, também é adicionado apenas aquando do

armazenando dos dados, na base de dados. É determinado a partir da função *getDate* do Microsoft SQL Server.

Perfil_Dimensions_ID

O *Perfil_Dimensions_ID* é o código que indica o tipo de perfil ao qual os dados dizem respeito. É fornecido a partir da tabela *header*, como demonstrado na tabela 20, no apêndice A.

Date_Time

Este atributo que representa o instante em que o perfil foi analisado pelo perfilômetro, é determinado a partir da concatenação dos atributos: *Year; Month; Day; Hour; Minute; Second*, presentes na tabela *Date*, ilustrada na tabela 22, no apêndice A. A nomenclatura do atributo é a seguinte: 'YYYY-MM-DD HH:MM:SS'.

Production_ID

Como supracitado, no conjunto de dados a processar pelo ETL podem existir diferentes ordens de produção com o mesmo *Perfils_Dimensions_ID*. Por forma a distinguir estas ordens de produção foi criado este atributo, que resulta da concatenação dos atributos: *Perfils_Dimensions_ID*, *Date_Time*, a nomenclatura da coluna é dada por: *Perfils_Dimensions_ID-Date_Time*.

Work_Center

O ETL irá interagir com diferentes parâmetros que irão estar em constante mudança, tendo, portanto, de se ajustar. Focando no nome da base de dados, concretamente; a designação que se dá hoje à base de dados da DW, onde os dados são armazenados, pode não ser a mesma daqui a cinco anos. Até mesmo o nome das tabelas da DW, onde os dados são carregados, pode alterar ao longo do tempo. Esta informação é considerada e é relevante para o ETL, tendo de existir um mecanismo que permita ao ETL ajustar-se estas mudanças de parâmetros de modo a que o mesmo não dê erro.

Para solucionar esta questão foi criado um ficheiro Excel denominado *Parameters*, que armazena a informação atualizada sobre cada um destes diversos parâmetros. O ETL cada vez que é executado consome a informação diretamente deste ficheiro, assegurando que os parâmetros estão em consonância com a realidade. A estrutura do excel encontra-se representada na tabela 9.

O atributo *Work_Center* indica o ID do posto de trabalho onde está localizada a máquina que gerou os dados em questão. A informação é fornecida a partir do Excel *Parameters*.

Tabela 9 - Parâmetros configuráveis do ETL.

Parameter	Name
Server	LOAS030\LO_I2
Database	PLT_PREPARATION_EXE09
User_ID	user123
Password	pw123
Table 1	[PERF.Header]
Table 2	[PERF.Date]
Table 3	[PERF.Dimensions]
Table 4	[PERF.Contour]
Table 5	[PERF.More_Measures]
Table 6	[PERF.Conicity_Surfaces]
Table 7	[PERF.Perfil_Error]
Table 8	[PERF.Perfil_Type]
Table 9	[PERF.Stripes_In_Lines]
Table 10	[PERF.Piece_Size_Markers]
Table 11	[PERF.Conti_Measures_AECS]
Table 12	[PERF.Shoulder_Measurements_EX]
Table 13	[PERF.Central_Thickness_EX]
Table 14	[PERF.Section]
Table 15	[PERF.Encoder]
Table 16	[PERF.Nominal]
Table 17	[PERF.Overlay]
Table 18	[PERF.Trigger]
Work Center	EXE09
Log Table	[PERF.Applications_Logs]

Todos os dados, independentemente da tabela que representam, irão ser importados para a DW da empresa. As chaves que permitem unir todas as tabelas da DW são o *Date_Time* e o *Work_Center*.

5.2.3.2 KPI - Espessura

De modo a facilitar o cálculo das métricas relacionadas com a espessura foram adicionados quatro atributos à tabela *contour*.

Nominal_Thickness_mm

Estudar a espessura do perfil por si só, não traz qualquer benefício, é necessário comparar os valores reais do perfil com os valores que se pretendem alcançar, representados pelo ficheiro NOX. As espessuras do perfil real encontram-se armazenadas na tabela *contour* originada pelo ficheiro TSX e as espessuras do perfil nominal encontram-se armazenadas na tabela *nominal* proveniente do ficheiro NOX. No entanto, o número de pontos representados pela tabela *contour* em média é de dois mil duzentos e cinquenta, enquanto o número de pontos representados pela tabela *nominal* em média é de vinte. A diferença entre o número de pontos impossibilita a comparação direta entre tabelas. Por forma a contornar a situação implementaram-se os passos que a seguir se discriminam.

Primeiramente, analisou-se a partir do atributo *Nominal_Total_Width_mm* a largura total do perfil nominal. Sabe-se que a distância entre os pontos medidos pelo perfilômetro é de 1mm, portanto sabe-se que o número de pontos máximo é igual à largura do perfil em milímetros mais 1, como demonstrado na figura 33. O número de pontos na tabela *contour* é sempre máximo porque o perfilômetro mede todos os pontos de 1 em 1 milímetro, daí que para comparar o perfil real com o perfil

nominal é necessário converter os vinte pontos indicados pelo ficheiro NOX no número de pontos máximo para o tipo de perfil em questão. Tendo calculado o número máximo de pontos para o perfil nominal, a

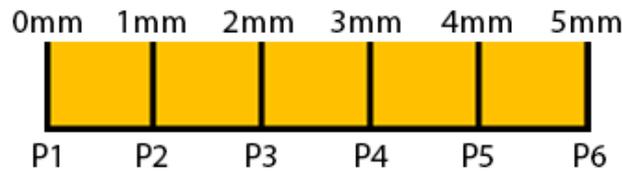


Figura 33 - Número de pontos por 5 milímetros.

partir da espessura dos vinte pontos fornecidos é realizada interpolação linear para determinar a espessura para todos os pontos.

Num cenário perfeito, poder-se-ia criar o atributo *Nominal_Thickness_mm* e a adicionar à tabela *contour* para que se calculasse a diferença entre os atributos: *Contour_Thickness_mm* e *Nominal_Thickness_mm*. Porém, a largura do perfil real quer no seu global, quer do lado esquerdo ou direito é geralmente distinta do indicado pelo perfil nominal, levando a que o perfil real possa ter menos ou mais pontos que o perfil nominal. A solução para este problema é simples, acrescentar linhas com o número 0 para o atributo com menor número de pontos: *Contour_Thickness_mm* ou *Nominal_Thickness_mm*. No entanto é necessário primeiro perceber qual o perfil mais pequeno, e em quais dos lados; esquerdo, direito ou ambos.

Face à complexidade deste estudo os passos irão ser explicados um a um e sustentados, graficamente, pela figura 34.

1. A partir dos atributos *Left_Width* e *Right_Width* presentes na tabela *section*, conseguiu-se calcular o número de pontos totais do lado esquerdo e direito.

$$Number\ Side\ Dots = Side\ Width + 1 \quad (4)$$

2. Sabendo que o perfil é suposto ser assimétrico pode-se inferir que o número de pontos do lado esquerdo é igual ao do lado direito do perfil da nominal. Recorreu-se então à equação:

$$Number\ Side\ Dots\ Nominal = \frac{Number\ Total\ Dots}{2} \quad (5)$$

3. Após determinado o número de pontos para cada lado do perfil real e nominal, passa-se à comparação, como demonstrado na secção A da figura 34.
4. Identificado para cada perfil o lado ou lados com menor número de pontos, é então necessário corrigir esse lado(s), adicionando 0 no respetivo atributo para que o número de pontos se iguale.

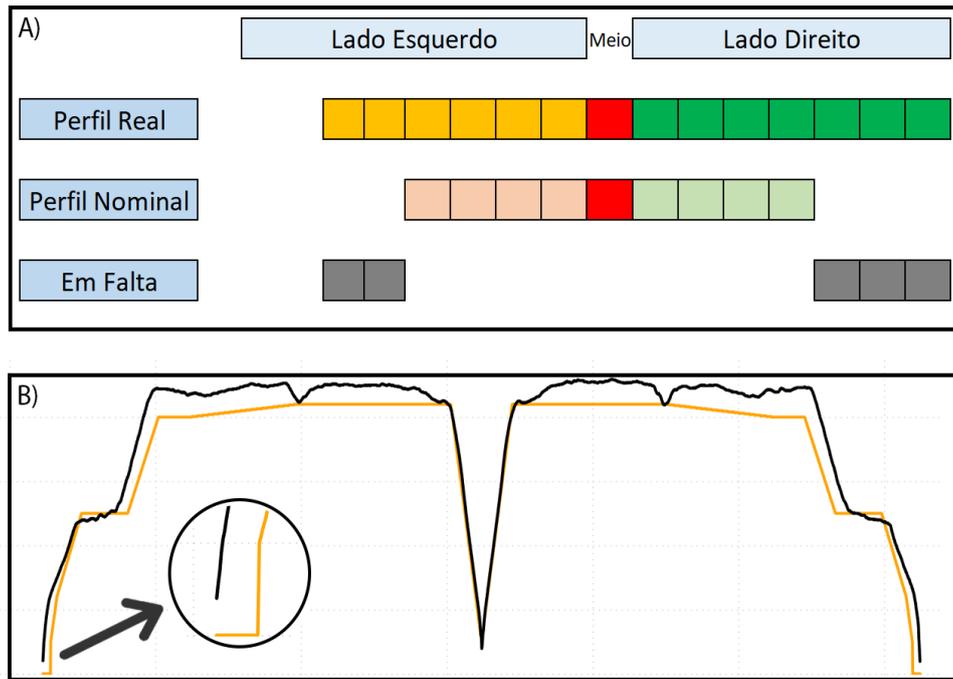


Figura 34 - Discrepância entre a largura do perfil real e nominal.

5. Ao adicionar as linhas é indispensável ter cuidado com um detalhe: as linhas estão ordenadas por tempo e o perfilômetro mede a espessura do lado esquerdo para o lado direito, i.e. as primeiras linhas da tabela, para um dado perfil, começam por representar o lado esquerdo do perfil e, conseqüentemente, as últimas linhas representam o lado direito do mesmo. Em suma, caso o perfil real seja apenas maior que o perfil nominal no seu lado esquerdo, as linhas com o valor zero irão ser inseridas no atributo *Nominal_Thickness_mm* no topo das linhas que dizem respeito a essa iteração. Esta particularidade mais complexa é ilustrada pela figura 35.

Na secção B da figura 34 é apresentado um caso real de uma iteração em que foi necessário adicionar zero no início e no fim das linhas do atributo *Nominal_Thickness_mm*, dado que o perfil nominal era mais pequeno que o perfil real, em ambos os lados.

Ordem de Produção	Iteração	Contour_Thickness	Nominal_Thickness
A	1	Amarelo	Púrpura
A	1	Amarelo	Púrpura
A	1	MEIO	MEIO
A	1	Verde	Verde
A	1	Verde	Púrpura
A	2	Amarelo	Púrpura
A	2	Amarelo	Púrpura
A	2	Amarelo	Amarelo
A	2	MEIO	MEIO
A	2	Verde	Verde
A	2	Verde	Verde

Figura 35 - Adição de 0 para igualar número de pontos.

Após a implementação destes cinco passos, o atributo *Nominal_Thickness_mm* é adicionado à tabela *contour*, facilitando o cálculo da diferença de espessura entre o perfil real e o pretendido.

Is_Middle

A partir do número de pontos do lado esquerdo e direito do perfil real, determinados na etapa anterior, é estipulado para cada iteração presente na tabela *contour* qual a linha que representa o meio do perfil. No caso especial em que o número de pontos total é par, o ponto médio é o último ponto do lado esquerdo. A coluna *Is_Middle* é adicionada à tabela *contour*, sinalizando com zero todas as linhas que não representam o meio e com 1 as linhas que representam o meio para a respectiva iteração. Este atributo é útil porque permite determinar o atributo seguinte: *What_Side*.

What_Side

O atributo *What_Side* é adicionado à tabela e é composto por duas possíveis *strings*: *LEFT* ou *RIGHT*, que, logicamente, indicam se uma linha pertence ao lado esquerdo ou direito do perfil real. A utilidade deste atributo face à espessura está presente na capacidade que irá dar aos colaboradores do departamento de qualidade de conseguirem estudar as métricas relacionados com a espessura em três possíveis ambientes: face ao perfil total; face ao lado esquerdo do perfil; face ao lado direito do perfil.

Cumulative_Width_mm

Em linha com a utilidade que é oferecida ao relatório interativo, foi desenvolvido o atributo *Cumulative_Width_mm*. Esta coluna indica a largura a que se encontram os pontos do perfil ao longo do eixo, para cada uma das iterações das ordens de produção. A cada ponto o atributo é acrescido em 1mm, lembrando que a distância entre pontos é 1 milímetro. Esta coluna permite que seja possível num gráfico do relatório demonstrar a espessura do perfil ao longo eixo, colocando como eixo do Y a espessura e como eixo do X a largura do perfil a que se encontra cada ponto.

5.2.3.3 KPI - Área

Com vista a facilitar o cálculo das métricas relacionadas com a área foram adicionados quatro atributos à tabela *contour*.

py_Total_Area_mm2

As métricas relacionadas com a área são da máxima importância e face à grande precisão do perfilômetro, no que concerne aos pontos medidos, 1mm entre estes, é forçoso garantir que qualquer métrica relacionada com a área é medida com a maior precisão possível. Para tal, foi utilizada a função *trapz* proveniente da livreria *numpy*. Esta função recebe como *input* os atributos *Cumulative_Width_mm*

e *Contour_Thick_mm*, que a partir da equação da área do trapézio, calcula, primeiramente, a área criada por cada par de pontos e, por fim, soma todas as áreas determinadas.

Na figura 36 é ilustrado o conceito utilizado para o cálculo da área do perfil. Decidiu-se usar um perfil nominal para demonstrar o conceito, pois caso se optasse por usar um perfil real, seriam cerca de 2250 pontos, logo 2249 intervalos, o que levava a que a ilustração se tornasse confusa. Por conseguinte, é importante notar que apesar do exemplo apresentar um perfil nominal, no ETL apenas se calcula a área para o perfil real de cada iteração.

O atributo *py_Total_Area_mm2* é calculado para a tabela *contour* e para a tabela *dimensions*, como se pode comprovar pela tabela 21 e 23, no apêndice A, ambas possuem os *inputs* necessários à função.

py_Left_Area_mm2

O atributo *py_Left_Area_mm2* representa a área do perfil no lado esquerdo e o processamento dos dados que deu origem a este atributo recorre ao mesmo algoritmo explicado na secção *py_Total_Area_mm2*. A única distinção é que se aplica o algoritmo tendo apenas como *input* os pontos que representam o lado esquerdo do perfil.

A seleção dos pontos do lado esquerdo na tabela *contour* utiliza como filtro o atributo *What_Side*. Já a tabela *dimensions*, necessita de analisar em que *Cumulative_Thickness_mm* o perfil transita do lado esquerdo para o direito, sinalizar essa medida e, utilizá-la como filtro para dividir os seus pontos.

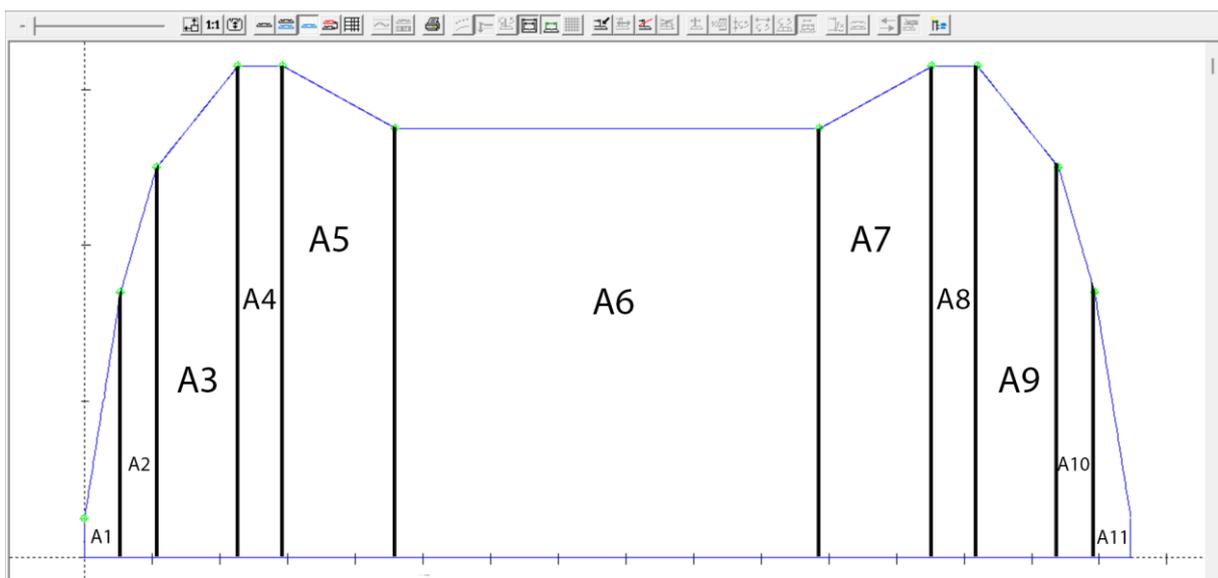


Figura 36 - Ilustração dos diversos trapézios.

py_Right_Area_mm2

O atributo *py_Right_Area_mm2* representa a área do perfil no lado direito e o processamento dos dados que deu origem a este atributo recorre ao mesmo algoritmo explicado na secção *py_Left_Area_mm2*.

py_Area_Difference_mm2

O atributo *py_Area_Difference_mm2* representa a diferença entre a área total do perfil real e a área do perfil nominal. O processamento dos dados que deu origem a este atributo resulta da diferença entre os seguintes atributos: *py_Total_Area_mm2* e *Nominal_Total_Area_mm2*.

Apesar dos atributos *py_Left_Width_mm* e *py_Right_Width_mm*, representados na tabela 20, no apêndice A, não conseguirem trazer nenhum contributo adicional ao cálculo de métricas relacionadas com a área, o departamento de qualidade requisitou, ainda assim, estes atributos na tabela *header*. Pelo facto de os atributos *Nominal_Left_Width_mm* e *Nominal_Right_Width_mm* pertencerem a esta. Ao coexistirem estes dois atributos, com o *Nominal_Left_Width_mm* e *Nominal_Right_Width_mm* na mesma tabela, assegura-se que qualquer comparação futura entre os respetivos pares que possa decorrer é facilitada.

5.2.3.4 KPI – Assimetria

Is_Contour

O atributo *Is_Contour* presente na tabela *contour*, tal como o *Is_Middle*, é um booleano que indica se uma dada linha da tabela *contour* é real ou não, i.e. a solução em vigor para resolver o problema da discrepância entre o número de pontos do perfil real e o perfil nominal consiste em adicionar novas linhas (pontos) na tabela *contour* em que o valor dos atributos *Contour_Thickness_mm* ou *Nominal_Thickness_mm* é igual a zero. Estas linhas ao serem usadas para calcular a área ou assimetria vão dar origem a erros, na medida em que vão levar a que pontos que não existam sejam tidos em conta.

Ao criar este atributo sempre que se pretender realizar estudos relacionados com a assimetria ou área, basta filtrar os pontos a partir do atributo *Is_Contour* considerando as linhas da tabela em que o atributo é igual a 1, i.e. é um ponto real.

assimetry_differences_Pvalue_Paired_T_Test

Através do teste *P-value Paired T*, é possível examinar com precisão a existência de assimetria entre o lado esquerdo e direito do perfil. Esse teste estatístico baseia-se na comparação das diferenças entre pares de pontos relacionados, obtidos em cada lado do perfil. O conceito “par de pontos” encontra-se ilustrado na figura 37.

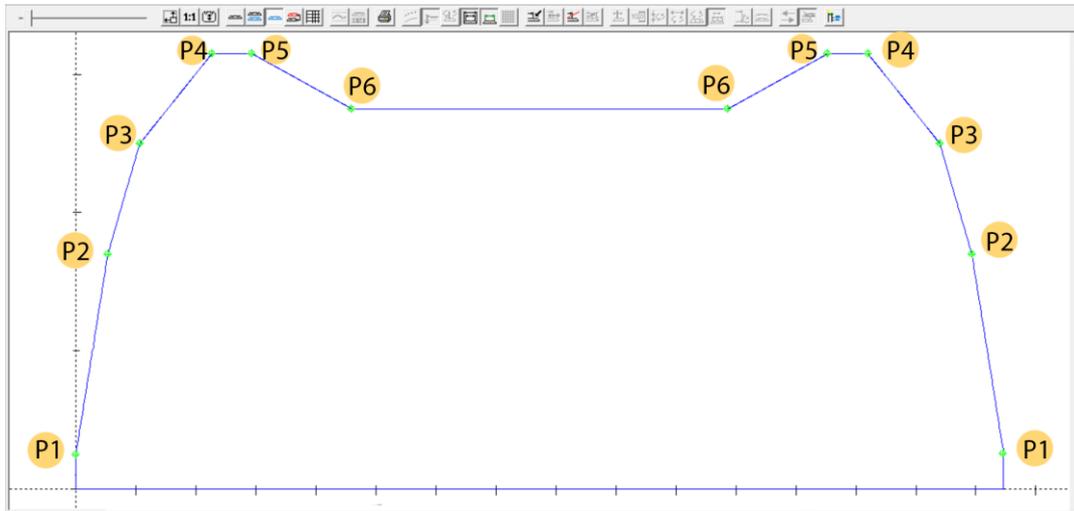


Figura 37 - Exemplos de par de pontos para um perfil nominal de um piso.

Ao aplicar o teste, recolhem-se as medidas para todos os pares de pontos longo do perfil. De seguida, calculam-se as diferenças entre os pares e, por fim, analisa-se se a média das diferenças é estatisticamente diferente de zero. Este estudo estatístico resulta no valor p , uma medida estatística que indica a evidência contra a hipótese nula presente num teste de hipóteses como o atual: se p for menor que o nível de significância pré-determinado, é possível inferir que há evidências estatísticas que rejeitam a hipótese nula de que não há assimetria.

Ao utilizar este teste, é possível identificar e quantificar qualquer assimetria entre os lados esquerdo e direito do perfil do pneu, fornecendo informações essenciais a ajustes e melhorias no processo de fabricação, que acabarão por garantir um produto com maior qualidade e desempenho.

Como referido no capítulo 5.2.3 o teste foi realizado a partir da função `ttest_rel` proveniente da livraria `scipy`. Esta função recebe como input o atributo `Contour_Thickness_mm`.

assymetry_differences_Pvalue_Paired_Wilcoxon

O teste *Pvalue Paired Wilcoxon*, tal como o *t-test*, estuda a possível presença de assimetria face à espessura do perfil ao longo do eixo transversal. A principal diferença entre eles é que o *wilcoxon* é não paramétrico, i.e. não faz suposições sobre a distribuição dos dados.

Tal como o *t-test* também o teste *wilcoxon* tem como output o valor p . Assim, caso p seja inferior ao nível de significância pré-determinado, é possível deduzir que há evidências estatísticas que sustentam a presença de assimetria.

O teste foi realizado a parti da função *wilcoxon* presente na livraria `scipy`. Esta função também recebe como input o atributo `Contour_Thickness_mm`.

5.2.3.5 KPI – Peso

Has_Out

Como ilustrado pela figura 23, o instante temporal em que o perfilômetro mediu os dados presentes no ficheiro TSX é a variável usada para nomear os ficheiros TSX. No entanto, a legenda *out* é uma *string* adicional que por vezes aparece no nome dos ficheiros, como por exemplo: “09-35-29-out”. Esta legenda é escrita no nome do ficheiro para sinalizar que aquele “setor” do perfil que está a ser medido pela máquina foi criado enquanto a máquina estava a sofrer algum tipo de *setup*. O *setup* leva a que as métricas deste ficheiro fiquem muito aquém do requerido, apresentando grandes desvios face a qualquer um dos KPI, como a espessura, a área e a assimetria. De realçar que a percentagem de ficheiros TSX com a legenda *out* é inferior a 1% e encontra-se presente nas iterações iniciais da corrida.

Este atributo representa um booleano que permite identificar que iterações representam estes momentos de *setup*, em conformidade com os atributos anteriores: quando o valor é 1 existe *setup* relacionado com o instante; quando o valor é zero não existe *setup* relacionado com o instante.

Todos os atributos identificados até à data analisam apenas o perfil iteração a iteração, não analisam a corrida como um todo ou fases desta. Todavia, irão ser apresentados KPI ao longo da dissertação que analisam a corrida do perfil como um todo ou fases desta e, nesses casos, é necessário não considerar as iterações em que o *setup* influenciou as métricas do perfil.

dif_perc_area

O perfilômetro é incapaz de medir o peso do perfil. Porém, em conjunto com o departamento de qualidade foi definido um processo que permite estimar a diferença percentual entre o peso do perfil real e o peso da nominal.

1. Antes de tudo, foi delineado qual o KPI a usar para estimar a diferença percentual entre o peso do perfil real e da nominal. Tendo em conta os dados fornecidos pelo perfilômetro, decidiu-se que a métrica mais realista seria a resultante da equação 6:

$$\text{Diferença Percentual} = \frac{\text{Média das Diferenças entre a Área do Perfil Real e Nominal}}{\text{Área do Perfil Nominal para Uma Iteração}} \quad (6)$$

O fundamento assenta no seguinte raciocínio: Se em média, para cada iteração, a área real do perfil é, por exemplo, inferior 4% à área do perfil nominal, pode estimar-se que o volume total do perfil é 4% mais pequeno, já que o perfil deve ser igual para cada iteração de uma corrida. Se a cada intervalo a área é inferior em 4%, o seu volume total será, também, inferior em cerca de 4%. Como o peso se encontra igualmente distribuído pela

borracha do perfil, é possível concluir que com uma redução de 4% face ao volume total pretendido, virá acoplada uma redução em 4% do seu peso.

2. Concluída a 1ª etapa, foram definidas quais as fases da corrida a analisar. Foi determinado que se iriam analisar as seguintes fases:
 - a. 1º quarto da corrida;
 - b. 2º quarto da corrida;
 - c. 3º quarto da corrida;
 - d. 4º quarto da corrida;
 - e. Totalidade da corrida.
3. De seguida, foi decidido que para cada fase da corrida seria analisado o perfil a partir de três perspetivas: perfil total; lado esquerdo do perfil; lado direito do perfil. A combinação das cinco fases com as três perspetivas, resulta num total de quinze combinações.
4. Por fim, procedeu-se ao cálculo dos novos atributos:
 - a. Para aplicar a equação 7 a cada uma das 15 combinações utilizaram-se como input os atributos: $py_Total_Area_mm^2$, $py_Left_Area_mm^2$, $py_Right_Area_mm^2$, $Nominal_Total_Area_mm^2$.
 - b. Registaram-se os resultados na tabela *header*, a partir da adição de quinze novos atributos, como é ilustrado pela tabela 20, no apêndice A. De relevar que como este atributo não representa uma iteração, mas sim um conjunto de várias, o valor é repetido ao longo das várias linhas da ordem de produção em questão.

Este é um caso concreto de uma métrica que engloba várias iterações e pretende classificar a corrida no seu todo ou as fases desta, donde se conclui que deve filtrar os dados a partir do atributo *Has_Out*, considerando apenas as iterações em que o *setup* não influenciou o perfil.

5.2.3.6 *Volume dos Dados*

A precisão dos dados face ao elevado número de pontos marcado pelo perfilômetro, até agora tem sido uma grande mais-valia para a empresa, permitindo calcular KPI com o mínimo de erro possível. Por outro lado, com a aproximação da 3ª etapa do ETL, Armazenamento, esta característica torna-se uma desvantagem, posto que torna o processo de armazenagem demorado.

O ETL desenvolvido pelo presente projeto atua de quatro em quatro horas, i.e. irá extrair, processar e armazenar todos os ficheiros gerados pela extrusora nas últimas quatro horas. A rotina do ETL irá ser desenvolvida com maior detalhe no subcapítulo 5.6.

Em média a extrusora realiza dezasseis ordens de produção a cada intervalo de quatro horas e cada ordem gera em média noventa e sete ficheiros TSX. Olhando à tabela com maior número de linhas temos a tabela *contour* com uma média de duas mil duzentas e cinquenta linhas por iteração. Multiplicando estes três valores para calcular o número de linhas a importar para a DW, apenas no que se refere à tabela *contour*, tem-se um total de 3.420.000 linhas. Um número elevadíssimo de linhas a importar de quatro em quatro horas.

O *Power Bi, software* selecionado para realizar os relatórios interativos, não possui uma capacidade de armazenamento ilimitada, i.e. o histórico de dados importados para o mesmo não pode ser ilimitado. Face à quantidade de dados que são armazenados a cada iteração, rapidamente se atinge a limitação máxima do Power Bi com a importação de um espetro relativamente curto. O relatório em Power Bi desenvolvido no projeto que considera todas as iterações da corrida, a explicar mais pormenorizadamente no capítulo 6, pode importar um histórico máximo de quinze dias.

Is_Beginning_Mid_End

Estudar os dados num espetro máximo de quinze dias é ineficaz. O departamento de qualidade pretende, no mínimo, conseguir analisar o último trimestre. Para resolver este desafio foi adicionado à tabela *header* e *contour* um novo atributo nomeado *Is_Beginning_Mid_End*. Este atributo possui um de quatro valores: *Beginning*; *Middle*; *End*; *None*. As iterações que possuem o valor *Beginning* representam a primeira iteração da corrida gravada pelo perfilômetro, as que possuem o valor *Middle* representam a iteração do meio, as que possuem *End* representam a iteração final e, naturalmente, todas as outras iterações da corrida possuem o valor *None*.

Esta atributo permite analisar as métricas do perfil para todas as corridas em três instantes distintos: início, meio e fim da corrida. De notar que qualquer iteração com *has_out* igual a 1 não pode ser considerada aquando da identificação destas três classes de instantes.

A partir deste atributo foi possível desenvolver um outro relatório que importa todas as ordens de produção presentes na DW (histórico total), em que são apenas importadas três iterações por ordem de produção. De realçar, mais uma vez, que toda a informação detalhada relacionada com os relatórios interativos desenvolvidos, encontra-se expressa no capítulo 6 da dissertação.

5.2.3.7 *Sobreposição das Espessuras*

Em síntese, existem dois relatórios interativos por extrusora: um relatório que importa as ordens de produção dos últimos quinze dias, com todas as iterações; outro relatório que consegue importar todo o historial de ordens de produção, mas apenas três iterações por ordem.

Face ao relatório que importa todo o historial, o departamento de qualidade requisitou que se fosse capaz de expressar graficamente a sobreposição das espessuras ao longo do eixo, para cada uma das três iterações do perfil, de modo a facilitar a identificação de diferenças entre estas.

Para tal, foi desenvolvida a tabela *overlay*, uma tabela com treze atributos, como ilustrado na tabela 26, no apêndice A. Os atributos *Beg_Thickness_mm*, *Mid_Thickness_mm*, *End_Thickness_mm*, indicam a espessura do perfil ao longo do seu eixo transversal para a iteração inicial, para a do meio e para a final, respetivamente. A partir destes três atributos é possível satisfazer o requisito do departamento, já que para cada ordem de produção a primeira linha irá armazenar a espessura para o primeiro ponto do eixo transversal do perfil em cada uma das três iterações, a segunda irá armazenar a espessura para o segundo, e assim sucessivamente.

O atributo *Nominal_Thickness_mm* está presente para facilitar a comparação com o que é esperado obter em cada iteração. Já os atributos *Beg_What_Side*, *Mid_What_Side* e *End_What_Side* permitem filtrar os dados quanto ao lado do perfil a que correspondem.

5.2.3.8 Descentralização do ETL

Ainda existe um requisito por preencher que irá facilitar o estudo de métricas complexas relacionadas com a assimetria. Após o processo de engenharia inversa ser realizado, assim como o AHP, o departamento de qualidade da empresa emitiu os seguintes requisitos adicionais:

1. Criar uma tabela que seja capaz de registar a diferença entre espessuras para cada par de pontos do perfil real, i.e. é necessário armazenar todas as diferenças de assimetria que são usadas, por exemplo, aquando da realização dos testes de assimetria: *Pvalue Paired T-Test*, *Pvalue Paired Wilcoxon*.
2. Criar uma tabela semelhante à *overlay* que permita expressar graficamente a sobreposição das diferenças de assimetria, para cada uma das três iterações do perfil: início, meio e fim.

As tabelas a criar foram nomeadas *Assimetry_Differences* e *Assimetry_Differences_BME*, respetivamente, onde a sigla BME, significa *Beggining Middle End*.

Apesar do processamento dos dados ser facilitado a partir do python relativamente ao SQL, a transformação dos dados que resultará nos *dataframes* a serem armazenados nestas tabelas é realizada a partir do Microsoft SQL Server. Esta decisão leva a que as fases Transformação e Armazenamento sejam subdivididas em duas partes: 1ª parte em que se recorre ao python; 2ª parte em que se recorre o SQL Server.

A razão para se optar pelo SQL Server, está diretamente relacionada com o volume de dados a importar de quatro em quatro horas para a DW e com a simplicidade da transformação necessária à segunda parte do ETL. Se por intervalo de quatro horas, como referido anteriormente, a tabela *contour* possuir cerca de três milhões de linhas, a tabela da assimetria como retrata a diferença da espessura entre par de pontos, irá possuir cerca de 1,5 milhões de linhas. Logo, caso se optasse por realizar a criação desta tabela em *python* iria haver um acréscimo de 1,5 milhões de linhas a importar para a base dados. Este aumento substancial iria trazer as desvantagens citadas anteriormente: aumento substancial da duração do ETL e acréscimo da probabilidade de se perderem dados.

Ao recorrer-se ao Microsoft SQL Server, esta transformação ocorrerá após os dados estarem armazenados na DW, impossibilitando a perda de dados. Para além da segurança dos dados, a duração da fase Armazenamento é inferior utilizando o SQL Server comparativamente ao cenário em que se utiliza o *python* como na primeira parte do ETL. Ao executar as *queries* diretamente no SQL Server, o seu tempo de transformação é semelhante caso se realizasse o mesmo em *python*. Contudo, o tempo de armazenamento é mais significativamente reduzido. Sendo a duração mais curta e a transformação dos dados relativamente acessível, não existe qualquer razão para realizar a transformação e armazenamento dos dados destas duas tabelas, em *python*.

5.3 Armazenamento – 1ª parte

5.3.1.1 Armazenamento dos Dados

Com a criação da tabela *overlay* dá-se por terminada a terceira etapa do ETL, iniciando a etapa seguinte “Armazenar Dataframe”, apresentada na figura 29, que dispara a terceira fase do ETL, o Armazenamento. Neste subcapítulo, irá ser detalhada a fase Armazenamento do ETL que recorre ao *python*.

Primeiramente, foram criadas duas bases de dados no servidor do RDBMS, uma para cada extrusora. A necessidade de criar duas bases de dados distintas é justificada caso se executasse o ETL para cada uma das respetivas máquinas, e se a terceira fase do ETL (Armazenamento) de ambas a tabela, por alguma razão, decorresse em simultâneo, iria haver funções de ETL distintos a escrever na mesma tabela, ao mesmo tempo, arriscando a perda de informação. Por tal facto, para permitir que o ETL decorresse ao mesmo tempo sem haver possibilidade de erro, foram criadas duas bases de dados, uma para cada extrusora.

Em cada uma das bases de dados foi necessário criar cada uma das dezassete tabelas a importar nesta primeira parte na DW: *header*; *date*; *dimensions*; *contour*; *more_measures*;

conicity_surfaces; perfil_error; perfil_type; stripes_in_lines; piece_size_markers; conti_measures_aecs; shoulder_measurements_ex; central_thickness_ex; section; encoder; nominal; overlay;

Seguidamente, foi necessário criar conexão entre o *python* e a DW da empresa. Para tal, foi utilizada a função *connect* da livreria *pyodbc*. A nomenclatura da função encontra-se representada na figura 38. A função necessita que lhe seja fornecida informação, mais propriamente: nome do servidor onde se encontra a DW; nome da base de dados; ID do utilizador; password do utilizador; toda esta informação necessária à terceira fase do ETL é importada a partir do excel *Parameters*.

Após estar criada a conexão, recorreu-se à função *execute*, também proveniente da mesma livreria, a qual recebe como *input* uma *query* em SQL e que é executada na base de dados à qual o *python* se encontra conectada. Como anteriormente citado, o RDBMS utilizado para interagir com a DW é o Microsoft SQL Server. Portanto, o *input* da função necessita de ser representado por uma *query* em linguagem SQL. A *query* usada para armazenar os dados na base (de dados), recorre à função *insert on*, embebida no SQL Server. Esta *query* será executada N vezes, sendo N o número total de linhas a armazenar na base de dados. A função *execute* necessita de receber como *input* os nomes das tabelas onde os dados serão importados. Esta informação, mais uma vez, é importada a partir do Excel *Parameters*.

```
connection_string = "Driver={SQL Server};Server=" + server + ";Database=" + database + ";ID=" + user + ";Password=" + password
cnxn = pyodbc.connect(connection_string)
```

Figura 38 - Nomenclatura da função *connect* proveniente da livreria *pyodbc*.

5.3.1.2 Transformação e Armazenamento com ou sem Processamento

Como foi referido no subcapítulo 5.1, é necessário ter precaução com a duração do ETL, para que nem o processo se torne demasiado lento, nem haja perda de informação. Assim sendo, para se poder verificar se a adição do processamento antes da armazenagem na DW resulta numa grande discrepância na duração do ETL, foi estudada a duração das etapas Transformação e Armazenamento com e sem processamento.

Foram selecionados vinte conjuntos distintos de dezasseis ordens de produção, foram selecionadas dezasseis ordens de produção por conjunto, dado que é o valor médio de ordens de produção realizadas em chão de fábrica a cada quatro horas. A primeira fase do ETL foi implementada para cada um destes vinte conjuntos, duas vezes. Uma vez com o processamento e uma segunda vez sem o processamento. A comparação entre a duração do ETL encontra-se apresentada na tabela 10. De referenciar, que este teste foi realizado em duas extrusoras, nomeadas extrusora sete e extrusora nove. Ao analisar a tabela 10, é possível perceber que para a extrusora sete e a extrusora nove houve um

Tabela 10 - Comparação entre ETL com e sem processamento.

Tipo de Máquina	Tempo de Extração		Tempo de Transformação e Armazenamento	
	Isento de Processamento	Com Processamento	Isento de Processamento	Com Processamento
Extrusora 7	-	-	1h 32m	1h 36m
Extrusora 9	-	-	1h 41m	1h 44m

aumento residual na duração de quatro e três minutos, respetivamente. Pode, então, concluir-se que o processamento pode manter-se no ETL.

5.4 Transformação e Armazenamento – 2ª parte

Como ilustrado na figura 29, a segunda parte do ETL apenas dispara quando a primeira parte do ETL tiver atuado para todas as ordens de produção presentes na pasta *Generated Files*. Quando a última ordem de produção é extraída, processada e armazenada na DW, é executada uma script em *python*. Esta script em *python* irá inserir uma linha na tabela *PERF.Trigger*, tabela presente na base de dados da extrusora em questão. Existe um *trigger* que indica que sempre que uma linha é inserida nesta tabela, é ativado o procedimento interno *create_assimetry* armazenado na RDBMS. É este procedimento que inicia a segunda parte do ETL. A emissão do sinal corresponde à quinta do ETL, figura 27.

É importante assegurar que o processamento realizado na segunda fase, pelas *scripts* em SQL, apenas se aplica aos dados que foram armazenados nesta iteração do ETL e não aos que já foram armazenados à priori. Para tal, criou-se uma tabela denominada *PERF.Contour_Last*, uma réplica da tabela *PERF.Contour*, que possui uma duplicação de todos os dados armazenados na primeira parte do ETL.

Com a execução do procedimento dá-se início à sexta etapa do ETL, “Criar Vistas Assimetria”. Nesta etapa, tal como o nome indica, serão criadas tabelas temporárias que irão armazenar os dados, temporariamente, até estes serem armazenados nas respetivas tabelas finais: *Assimetry_Differences* e *Assimetry_Differences_BME*.

Assimetry_Differences

O procedimento *create_assimetry* recebe como input todos os valores armazenadas na tabela *Contour_Last*, processa-os e cria a tabela temporária que apresenta os dados a armazenar na tabela *Assimetry_Differences*. O processo do procedimento pode ser subdividido nos seguintes passos:

1. Inicia-se um ciclo que irá percorrer todos as iterações presentes na tabela *Contour_Last*;
2. Em cada iteração do ciclo começa-se por dividir o atributo *Contour_Thickness_mm* em duas colunas distintas (armazenadas numa tabela temporária), o filtro utilizado para

realizar a divisão é o atributo *What_Side*, i.e. uma coluna irá armazenar as espessuras do lado esquerdo do perfil e a outra as espessuras do lado direito;

3. A ordem das linhas da coluna que contém as espessuras do lado esquerdo é invertida, de modo que o primeiro valor dessa coluna seja o par do primeiro valor da coluna que mostra os valores do lado direito do perfil (par de pontos);
4. Caso o comprimento das colunas seja distinto, são adicionados zeros na coluna com menor número de linhas até que o comprimento seja igualado;
5. Realizar a diferença entre os valores de ambas as colunas. Armazenar esses valores numa nova tabela temporária, e nomear esse atributo *Assimetry_Differences_mm*;
6. Adicionar os restantes atributos ilustrados na tabela 27, no apêndice A, já que estes são constantes para todos os pontos dentro de cada iteração;

Após a conclusão da transformação dos dados, é iniciado o segundo e último procedimento interno: *create_assimetry_BME*.

Assimetry_Differences_BME

O procedimento *create_assimetry_BME* recebe como *input* todos os valores armazenados na tabela temporária, resultante do procedimento anterior. Acende-se a esta tabela temporária e não à tabela *PERF.Assimetry_Difference* por se pretender atuar apenas sobre as linhas importadas nesta última iteração, e não sobre todas as linhas armazenadas ao longo do histórico.

O procedimento em questão irá processar esses dados e criar a tabela temporária que tem a informação a ser armazenada na tabela *Assimetry_Differences_BME*. O processo do procedimento pode ser subdividido nos seguintes passos:

1. Inicia-se um ciclo que irá percorrer todos as ordens de produção presentes na tabela temporária, resultante do procedimento *create_assimetry*;
2. A cada iteração do ciclo começa-se por dividir o atributo *Assimetry_Differences_mm* em três colunas distintas (armazenadas numa tabela temporária), o filtro utilizado para realizar a divisão é o atributo *is_Beginning_Mid_End*, i.e. uma coluna armazena a diferença de espessuras do par de pontos da iteração que representa o início da corrida, a outra coluna armazena a diferença de espessuras do par de pontos da iteração que representa o meio da corrida e, por fim, a última coluna armazena a diferença de espessuras do par de pontos da iteração que representa o final da corrida.

3. É adicionado ao atributo *Cumulative_Width_mm*, um valor numérico que se inicia em zero na primeira linha, de cada ordem de produção, e que a cada linha incrementa em 0,1.
4. Adicionar os restantes atributos, ilustrados na tabela 28, no apêndice A, dado que estes são constantes para todos as iterações dentro da mesma ordem de produção.

Após a conclusão da transformação dos dados, é iniciada a sétima e última etapa do ETL: “Armazenar Dados”. Todos os dados presentes na tabela temporária originada pelo procedimento *create_assimetry* são inseridos na tabela *PERF.Assimetry_Differences* e todos os dados presentes na tabela temporária resultantes do procedimento *create_assimetry_BME* são inseridos na tabela *PERF.Assimetry_Differences_BME*. Após a conclusão da inserção dos dados, dá-se por completo o ETL. Na figura 39 é ilustrado o esquema estrela da base de dados.

5.5 Conversão do ETL num Executável

O *trigger* e os procedimentos que compõem a segunda parte do ETL encontram-se armazenados no servidor da DW. No entanto, toda a 1ª parte do ETL é gerada a partir de vários ficheiros *python* interligados. Foi tomada a decisão de se converter estes ficheiros num só, executável. A escolha de converter em executável o ficheiro, baseou-se nas seguintes premissas:

1. O facto de optar-se por realizar um executável permite que não seja necessário instalar o *python* no servidor em que irá ser corrido o ETL.
2. Ao utilizar-se *python* recorre-se a bastantes livrarias. O problema de utilizar livrarias é o facto de estas possuírem versões. O ETL foi realizado com um conjunto de determinadas versões, uma para cada tipo de livraria. Nada garante que ao correr a mesma *script* com versões mais recentes não possa ocorrer algum erro. Ao converter-se os ficheiros *python* interligados num só executável, assegura-se que as livrarias mantêm sempre a versão.

5.6 Rotina do ETL

Como referido anteriormente, a pasta local onde se encontra armazenado o ETL necessita de preencher os seguintes requisitos:

1. Na pasta local necessita de estar armazenado o ficheiro de configuração dos parâmetros do ETL, *Parameters*, preenchido de acordo com os parâmetros atuais da base de dados da respetiva extrusora;

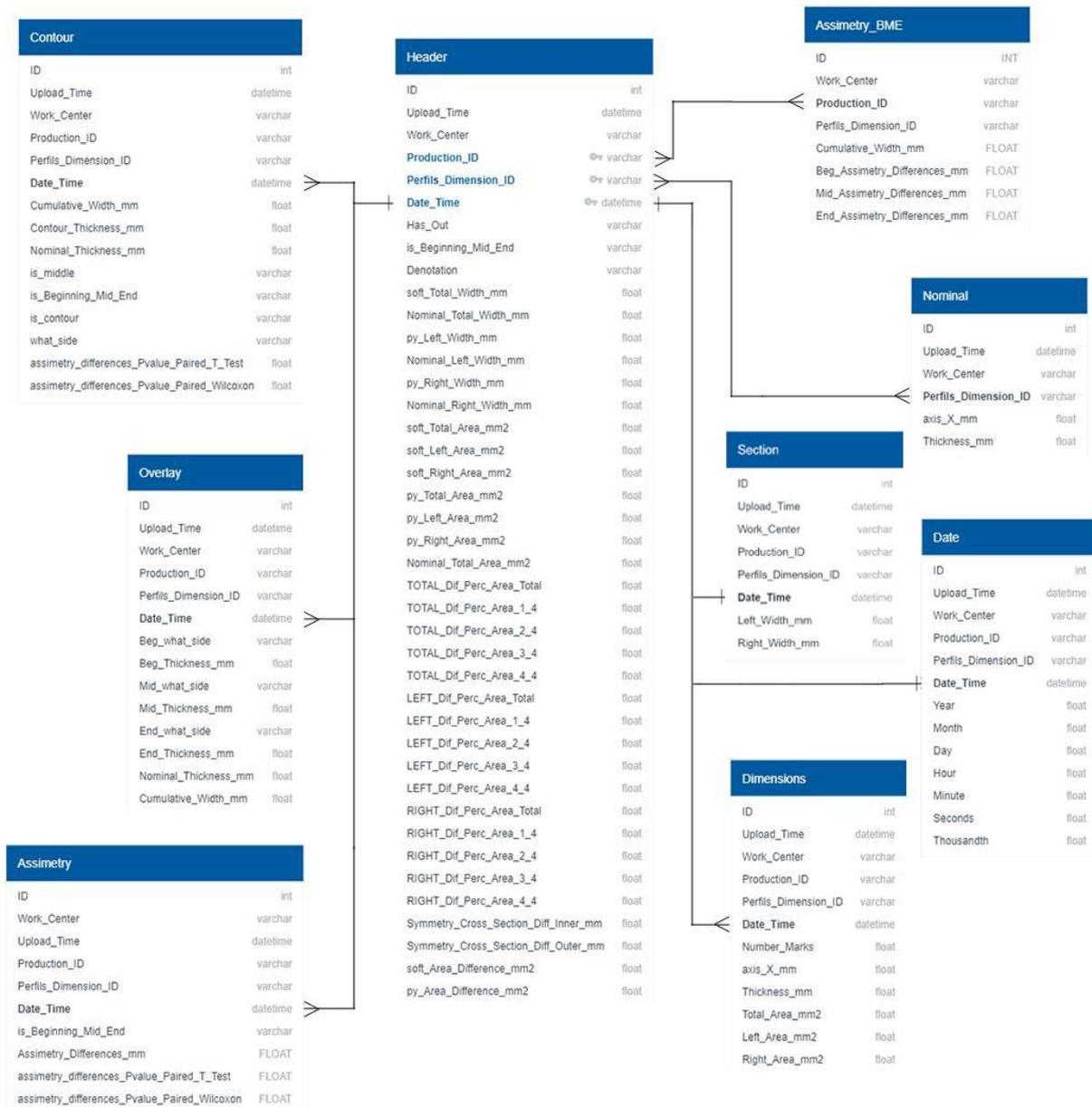


Figura 39 - Esquema da base de dados.

- Na pasta local necessita de estar armazenada a pasta *Generated Files*, preenchida com todos os ficheiros gerados pela respetiva extrusora durante as últimas quatro horas.

A arquitetura da pasta encontra-se ilustrada pela figura 40.

De modo a compreender-se a rotina do ETL é necessário entender como e quando são armazenados os ficheiros gerados pela extrusora. Após cada ordem de produção ser finalizada, os

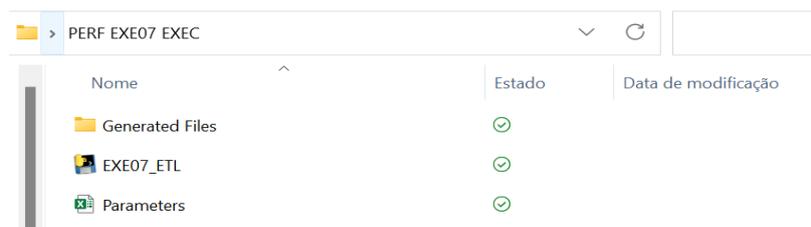


Figura 40 - Arquitetura da pasta local onde se encontra o executável.

ficheiros produzidos pela máquina são armazenados, localmente, num dos vários servidores da empresa, numa pasta nomeada *today*. No final do dia é criada uma pasta nomeada a partir da data do dia em questão e o conteúdo da pasta *today* é movido para lá.

A rotina do ETL é simples, porém possui vários passos:

1. De quatro em quatro horas, todas as pastas presentes na pasta *today* que tenham uma data de modificação que esteja compreendida entre o intervalo de tempo das últimas quatro horas irão ser copiadas para a pasta *Generated Files* presente na pasta onde o ETL se encontra;
2. Depois de todos os ficheiros serem copiados, o executável é iniciado;
3. Como ilustrado na figura 28, a primeira parte do ETL atua, sequencialmente, por todas ordens de produção. Logo que cada ordem de produção seja armazenada na DW, a pasta a que corresponde a ordem de produção é eliminada.
4. Caso ocorra algum erro para uma ordem de produção ao longo do ETL, a *string* que detalha o erro é inserida na tabela *PERF.Applications_Logs*. Esta tabela *log*, presente na base de dados da extrusora em questão, permite armazenar a documentação de todos os erros que possam ocorrer durante a execução do ETL.

As principais características do ETL encontram-se ilustradas, em síntese, pela tabela 11.

Tabela 11 - Características principais do ETL.

Tipo de Máquina	Setor	Formato do Ficheiro	Ferramenta de Processamento	RDBMS	Intervalo entre ETL	Duração do ETL		Número de Tabelas com que Interage
						1ª Parte	2ª Parte	
Extrusora 7	PLT	TSX; NOX	Python; SQL	MS SQL	4h	1h 36m	12m	21
Extrusora 9	PLT	TSX; NOX	Python; SQL	MS SQL	4h	1h 44m	13m	21

6. DESENVOLVIMENTO DO RELATÓRIO INTERATIVO

6.1 Seleção do *Software*

O processo de seleção do *software* a utilizar para a criar os relatórios foi simples e rápido. O departamento de qualidade da empresa, no início do projeto, já possuía, à priori, inúmeros relatórios ativos que permitem analisar diferentes tipos de máquinas e classes de KPI. A grande maioria destes relatórios foram criados a partir do Power Bi, com exceção de alguma minoria realizada a partir do Grafana. Para manter esta centralização, foi decidido que se utilizaria o Power Bi para analisar os dados provenientes das extrusoras.

6.2 Relatório – Last 15 Days

O primeiro relatório desenvolvido, denominado *Profilometer – Last 15 Days*, permite analisar todas as ordens de produção realizadas nos quinze dias últimos. Para cada ordem de produção é possível analisar cada iteração.

6.2.1 Dados Importados

Para o presente *dashboard* interativo foram importados os dados das tabelas:

1. *PERF.Contour*
2. *PERF.Header*
3. *PERF.Assimetry_Differences*

De relevar, que os dados presentes nas tabelas foram filtrados a partir do atributo *Date_Time* para que apenas as ordens de produção dos últimos 15 dias fossem carregadas para o *dashboard*. Este filtro foi aplicado a partir de três vistas realizadas no SQL Server, uma vista por cada uma das tabelas.

6.2.2 Arquitetura do Relatório

O relatório é constituído por quatro páginas: Introdução; Espessura; Área; Assimetria. Cada página é composta por métricas que se focam num dos KPI selecionados a partir do AHP, excetuando a página *Area* com enfoque nos KPI relacionados com a área e o peso.

Uma nota importantíssima a ter em conta ao analisar a arquitetura do relatório é que qualquer KPI que represente a diferença entre dois atributos segue a seguinte nomenclatura:

1. Caso o indicador represente a diferença entre um atributo do perfil real e um do nominal, é o valor do perfil real menos o do perfil nominal.

2. Caso seja diferença entre atributo do lado esquerdo do perfil e o mesmo atributo do lado direito do perfil, é o valor do lado esquerdo menos o do lado direito.

6.2.2.1 Introdução

A primeira página permite introduzir qual é o foco do relatório, quais os colaboradores da empresa que se encontram responsáveis pela sua manutenção, quando foi a última vez que o *dashboard* sofreu uma alteração e, ainda, qual a sua arquitetura. A primeira página encontra-se representada pela figura 47, do apêndice B.

6.2.2.2 Espessura

A segunda página do *dashboard* pode ser subdivida em duas secções, a secção das métricas e a secção dos filtros. Esta página possui todas as métricas relacionadas com a espessura do perfil e encontra-se ilustrada pelas figuras 48, 49 e 50, presentes no apêndice B. A figura 48 mostra a vista principal da página; a figura 49 permite analisar a barra lateral da página que detém alguns filtros; a figura 50 permite ilustrar a interação dos KPI com o filtro *Perfil's Side Selected*. A documentação de cada um dos KPI e filtros presentes na segunda página, encontra-se detalhada na tabela 12.

Os filtros podem ser subdivididos em duas classes:

1. Filtros de 1ª camada- que permitem filtrar os dados consoante a data, a ordem de produção, o tipo de perfil ou mesmo as iterações da corrida;

Tabela 12 - KPI e filtros associados à página Espessura.

KPI		
Nome	Figura	Conteúdo
Perfil Thickness - Graph	42	O gráfico demonstra a variação da espessura do perfil ao longo da sua largura.
Thickness Difference - Graph	42	O gráfico demonstra a variação da diferença entre a espessura do perfil real e do nominal ao longo da largura do perfil.
Average Difference	42	Este KPI indica a diferença média entre as espessuras do perfil real e nominal.
Standard Deviation	42	Este KPI indica o desvio padrão associado à diferença das espessuras entre o perfil real e nominal.
Maximum Difference	42	Este KPI indica o valor máximo das diferenças das espessuras entre o perfil real e nominal.
Dots Above Tolerance Level	42	Este KPI indica o número de pontos do perfil real que possuem, face ao perfil nominal, uma diferença de espessura superior ao que é tolerado pelo departamento de qualidade. O departamento de qualidade definiu que a tolerância é 0.5 mm, logo qualquer ponto que possua um desvio superior a este valor está fora dos limites aceitáveis de qualidade.
Filtros		
Nome	Figura	Conteúdo
Date Range	43	Este parâmetro filtra os dados quanto ao intervalo de tempo que se pretende analisar.
Perfil Dimensions ID	43	Este parâmetro filtra os dados quanto ao tipo de perfil que se pretende analisar.
Production ID	43	Este parâmetro filtra os dados quanto à ordem de produção que se pretende analisar.
Date Time	43	Este parâmetro filtra os dados quanto a que iteração ou iterações da corrida se pretende analisar.
Perfil's Type Selected	42	Este parâmetro filtra o primeiro gráfico, quanto ao tipo de perfil ou perfis que se pretende representar: real; nominal; real e nominal.
Perfil's Side Selected	44	Este parâmetro filtra o primeiro gráfico, quanto ao lado do perfil ou perfis que se pretende representar: esquerdo; direito; total.
Y Axis Graph 1	42	Este parâmetro define o valor do eixo Y do 1º gráfico.
Y Axis Graph 2	42	Este parâmetro define o valor do eixo Y do 2º gráfico.

2. Filtros de 2ª camada – que permitem filtrar os dados mais uma vez, mas a um nível mais baixo, permitindo analisar as métricas do relatório face a uma configuração mais específica dos parâmetros, como o lado do perfil.

Os filtros de 1ª camada encontram-se na aba lateral de cada uma das páginas do relatório, respetivamente. Os filtros de 2ª camada encontra-se no painel principal de cada uma das páginas.

6.2.2.3 Área

A terceira página do *dashboard* também pode ser subdivida em duas secções, a secção das métricas e a secção dos filtros. Esta página possui todas as métricas relacionadas com a área do perfil e, encontra-se ilustrada pela figura 51, presente no apêndice B. A documentação de cada um dos KPI e filtros presentes nesta página, encontra-se detalhada na tabela 13.

Tabela 13 - KPI e filtros associados à página Área.

KPI		
Nome	Figura	Conteúdo
Perfil Area - Graph	45	O gráfico demonstra a variação da espessura do perfil ao longo da sua largura, mas desta vez com um grafismo que realça a área do perfil.
Left Area	45	Este KPI indica a soma das áreas dos perfis reais, considerando apenas o lado esquerdo do perfil.
Right Area	45	Este KPI indica a soma das áreas dos perfis reais, considerando apenas o lado direito do perfil.
Nominal Left Area	45	Este KPI indica a soma das áreas dos perfis nominais, considerando apenas o lado esquerdo do perfil.
Nominal Right Area	45	Este KPI indica a soma das áreas dos perfis nominais, considerando apenas o lado direito do perfil.
Left Real Width	45	Este KPI indica a soma das larguras dos perfis reais, considerando apenas o lado esquerdo do perfil.
Rigth Real Width	45	Este KPI indica a soma das larguras dos perfis reais, considerando apenas o lado direito do perfil.
Left Nominal Width	45	Este KPI indica a soma das larguras dos perfis nominais, considerando apenas o lado esquerdo do perfil.
Rigth Nominal Width	45	Este KPI indica a soma das larguras dos perfis nominais, considerando apenas o lado direito do perfil.
Area Difference Real vs Nominal	45	Este KPI indica a diferença entre a soma das áreas dos perfis reais e nominais.
Area Difference Real vs Nominal %	45	Este KPI indica a diferença percentual entre a soma das áreas dos perfis reais e nominais.
Width Difference Real vs Nominal	45	Este KPI indica a diferença entre a soma das larguras dos perfis reais e nominais.
Width Difference Real vs Nominal %	45	Este KPI indica a diferença percentual entre a soma das larguras dos perfis reais e nominais.
Filtros		
Nome	Figura	Conteúdo
Date Range	-	Este parâmetro filtra os dados quanto ao intervalo de tempo que se pretende analisar.
Perfil Dimensions ID	-	Este parâmetro filtra os dados quanto ao tipo de perfil que se pretende analisar.
Production ID	-	Este parâmetro filtra os dados quanto à ordem de produção que se pretende analisar.
Date Time	-	Este parâmetro filtra os dados quanto a que iteração da corrida que se pretende analisar.
Perfil's Type Selected	45	Este parâmetro filtra o primeiro gráfico, quanto ao tipo de perfil ou perfis que se pretende representar: real; nominal; real e nominal.
Perfil's Side Selected	45	Este parâmetro filtra o primeiro gráfico e os KPI restantes, quanto ao lado do perfil ou perfis que se pretende representar: esquerdo; direito; total.
Y Axis Zoom	45	Este parâmetro define o valor do eixo Y do gráfico.

6.2.2.4 Assimetria

À semelhança das duas páginas que a antecedem também a quarta página do relatório poder ser subdivida em duas secções; a secção das métricas e a secção dos filtros. Esta página possui todas as métricas relacionadas com a assimetria do perfil e encontra-se ilustrada pela figura 52, presente no apêndice B. A documentação de cada um dos KPI e filtros presentes nesta página encontra-se detalhada na tabela 14.

Tabela 14 - KPI e filtros associados à página Assimetria.

KPI		
Nome	Figura	Conteúdo
Assimetry Differences - Graph	46	Este KPI demonstra a variação da diferença da espessura todos os pares de pontos.
Positive Dots %	46	Este KPI indica a percentagem de pontos do gráfico em que a diferença da espessura entre os pares de pontos é positiva.
Negative Dots %	46	Este KPI indica a percentagem de pontos do gráfico em que a diferença da espessura entre os pares de pontos é negativa.
Assimetry Differences Mean	46	Este KPI indica a média das diferenças de espessura entre os pares de pontos.
Assimetry Differences Std. Deviation	46	Este KPI indica o desvio padrão associado à média das diferenças de espessura entre os pares de pontos.
Maximum Assimetry Differences	46	Este KPI indica o valor máximo das diferenças de espessura entre os pares de pontos.
Assimetry P-Value T-Test	46	Este KPI indica o P-Value para o "T-Test" realizado na 1ª parte da fase <i>Transform</i> do ETL.
Assimetry P-Value Wilcoxon-Test	46	Este KPI indica o P-Value para o "Wilcoxon-Test" realizado na 1ª parte da fase <i>Transform</i> do ETL.
Filtros		
Nome	Figura	Conteúdo
Date Range	-	Este parâmetro filtra os dados quanto ao intervalo de tempo que se pretende analisar.
Perfil Dimensions ID	-	Este parâmetro filtra os dados quanto ao tipo de perfil que se pretende analisar.
Production ID	-	Este parâmetro filtra os dados quanto à ordem de produção que se pretende analisar.
Date Time	-	Este parâmetro filtra os dados quanto a que iteração da corrida que se pretende analisar.
Perfil's Type Selected	46	Este parâmetro filtra o primeiro gráfico, quanto ao tipo de perfil ou perfis que se pretende representar: real; nominal; real e nominal.
Perfil's Side Selected	46	Este parâmetro filtra o primeiro gráfico, quanto ao lado do perfil ou perfis que se pretende representar: esquerdo; direito; total.
Y Axis Zoom	46	Este parâmetro define o valor do eixo Y do gráfico.

6.3 Relatório – *Beginning Middle End*

O segundo relatório desenvolvido, nomeado *Profilometer – BME*, permite analisar todas as ordens de produção armazenadas na DW. Porém para cada ordem de produção é apenas possível analisar três iterações: a iteração inicial; a iteração do meio; a iteração final.

Face à capacidade deste *dashboard* em permitir que se analise todo o histórico produtivo da empresa relativamente à extrusora em questão, este relatório é considerado o mais importante.

6.3.1 Dados Importados

Para o presente *dashboard* interativo foram importados os dados das tabelas:

1. *PERF.Contour*
2. *PERF.Header*
3. *PERF.Overlay*
4. *PERF.Assimetry_Differences*
5. *PERF.Assimetry_Differences_BME*

De realçar que os dados presentes nas tabelas foram filtrados a partir do atributo *Is_Beginning_Mid_End* de forma que apenas as iterações que representam o início, meio ou fim de uma corrida fossem importadas para o *dashboard*. Este filtro foi aplicado a partir de três vistas realizadas no Microsoft SQL Server; uma vista por cada uma das seguintes tabelas: *PERF.Contour*; *PERF.Header*; *PERF.Overlay*.

6.3.2 Arquitetura do Relatório

Tal como acontece com o relatório *Last 15 Days* também este é constituído pelas páginas: *Introdução; Espessura; Área; Assimetria.*

6.3.2.1 Introdução

A primeira página permite introduzir qual é o foco do relatório, quais os colaboradores da empresa que se encontram responsáveis pela sua manutenção, quando foi a última vez que o *dashboard* sofreu uma alteração e, também, qual a sua arquitetura. A primeira página encontra-se representada pela figura 53, do apêndice B.

6.3.2.2 Espessura

A segunda página do *dashboard* pode ser subdivida em duas secções, a secção das métricas e a secção dos filtros. Esta página possui todas as métricas relacionadas com a espessura do perfil e encontra-se ilustrada pelas figuras 54 e 55, presentes no apêndice B. A figura 54 mostra a vista principal da página; a figura 55 permite analisar a barra lateral da página que detém alguns filtros. A documentação de cada um dos KPI e filtros presentes na segunda página encontra-se detalhada na tabela 15. Tal como o relatório *Last 15 Days*, os filtros podem ser subdivididos em duas classes: filtros de 1ª camada; filtros de 2ª camada.

6.3.2.3 Área

A terceira página do *dashboard* também poder ser subdivida em duas secções, a secção das métricas e a secção dos filtros. Esta página possui todas as métricas relacionadas com a área do perfil

Tabela 15 - KPI e filtros associados à página Espessura.

KPI		
Nome	Figura	Conteúdo
Perfil Thickness - Graph	48	O gráfico demonstra a variação da espessura do perfil ao longo da sua largura.
Thickness Difference - Graph	48	O gráfico demonstra a variação da diferença entre a espessura do perfil real e do nominal ao longo da largura do perfil.
Average Difference	48	Este KPI indica a diferença média entre as espessuras do perfil real e nominal.
Standard Deviation	48	Este KPI indica o desvio padrão associado à diferença das espessuras entre o perfil real e nominal.
Maximum Difference	48	Este KPI indica o valor máximo das diferenças das espessuras entre o perfil real e nominal.
Dots Above Tolerance Level	48	Este KPI indica o número de pontos do perfil real que possuem, face ao perfil nominal, uma diferença de espessura superior ao que é tolerado pelo departamento de qualidade. O departamento de qualidade definiu que a tolerância é 0.5 mm, logo qualquer ponto que possua um desvio superior a este valor está fora dos limites aceitáveis de qualidade.
Filtros		
Nome	Figura	Conteúdo
Date Range	49	Este parâmetro filtra os dados quanto ao intervalo de tempo que se pretende analisar.
Perfil Dimensions ID	49	Este parâmetro filtra os dados quanto ao tipo de perfil que se pretende analisar.
Production ID	49	Este parâmetro filtra os dados quanto à ordem de produção que se pretende analisar.
Selected Stages Graph 1	48	Este parâmetro filtra os dados quanto à iteração ou iterações representadas no 1º gráfico: <i>beginning; middle; end; nominal.</i>
Selected Stages Graph 2	48	Este parâmetro filtra os dados quanto à iteração ou iterações representadas no 2º gráfico: <i>beginning; middle; end; nominal.</i>
Y Axis Graph 1	48	Este parâmetro define o valor do eixo Y do 1º gráfico.
Y Axis Graph 2	48	Este parâmetro define o valor do eixo Y do 2º gráfico.
KPI Filter - Selected Stage	48	Este parâmetro filtra os KPI quanto à iteração ou iterações analisadas: <i>beginning; middle; end.</i>
KPI Filter - Selected Side	48	Este parâmetro filtra os KPI quanto ao lado do perfil analisado: esquerda; direita; total.

e encontra-se ilustrada pela figura 56, presente no apêndice B. A documentação de cada um dos KPI e filtros presentes nesta página encontra-se detalhada na tabela 16.

Tabela 16 - KPI e filtros associados à página Área.

KPI		
Nome	Figura	Conteúdo
Perfil Area - Graph	45	O gráfico demonstra a variação da espessura do perfil ao longo da sua largura, mas desta vez com um grafismo que realça a área do perfil.
Left Area	45	Este KPI indica a soma das áreas dos perfis reais, considerando apenas o lado esquerdo do perfil.
Right Area	45	Este KPI indica a soma das áreas dos perfis reais, considerando apenas o lado direito do perfil.
Nominal Left Area	45	Este KPI indica a soma das áreas dos perfis nominais, considerando apenas o lado esquerdo do perfil.
Nominal Right Area	45	Este KPI indica a soma das áreas dos perfis nominais, considerando apenas o lado direito do perfil.
Left Real Width	45	Este KPI indica a soma das larguras dos perfis reais, considerando apenas o lado esquerdo do perfil.
Rigth Real Width	45	Este KPI indica a soma das larguras dos perfis reais, considerando apenas o lado direito do perfil.
Left Nominal Width	45	Este KPI indica a soma das larguras dos perfis nominais, considerando apenas o lado esquerdo do perfil.
Rigth Nominal Width	45	Este KPI indica a soma das larguras dos perfis nominais, considerando apenas o lado direito do perfil.
Area Difference Real vs Nominal	45	Este KPI indica a diferença entre a soma das áreas dos perfis reais e nominais.
Area Difference Real vs Nominal %	45	Este KPI indica a diferença percentual entre a soma das áreas dos perfis reais e nominais.
Width Difference Real vs Nominal	45	Este KPI indica a diferença entre a soma das larguras dos perfis reais e nominais.
Width Difference Real vs Nominal %	45	Este KPI indica a diferença percentual entre a soma das larguras dos perfis reais e nominais.
Filtros		
Nome	Figura	Conteúdo
Date Range	-	Este parâmetro filtra os dados quanto ao intervalo de tempo que se pretende analisar.
Perfil Dimensions ID	-	Este parâmetro filtra os dados quanto ao tipo de perfil que se pretende analisar.
Production ID	-	Este parâmetro filtra os dados quanto à ordem de produção que se pretende analisar.
Date Time	-	Este parâmetro filtra os dados quanto a que iteração da corrida que se pretende analisar.
Perfil's Type Selected	45	Este parâmetro filtra o primeiro gráfico, quanto ao tipo de perfil ou perfis que se pretende representar: real; nominal; real e nominal.
Perfil's Side Selected	45	Este parâmetro filtra o primeiro gráfico e os KPI restantes, quanto ao lado do perfil ou perfis que se pretende representar: esquerdo; direito, total.
Y Axis Zoom	45	Este parâmetro define o valor do eixo Y do gráfico.

6.3.2.4 Assimetria

Da mesma forma que as duas páginas que a antecedem, esta pode ser subdivida em duas secções, a secção das métricas e a secção dos filtros. Esta página possui todas as métricas relacionadas com a assimetria do perfil e encontra-se ilustrada pela figura 57, presente no apêndice B. A documentação de cada um dos KPI e filtros presentes nesta página, encontra-se detalhada na tabela 17.

Como referido, anteriormente, a capacidade por relatório em Power Bi é limitada. Consequentemente, de modo a conseguir importar-se o máximo de dados por Power Bi foi decidido que seriam desenvolvidos por cada extrusora um relatório *BME* e um *Last 15 Days*.

Tabela 17 - KPI e filtros associados à página Assimetria.

KPI		
Nome	Figura	Conteúdo
Assimetry Differences - Graph	51	Este KPI demonstra a variação da diferença da espessura todos os pares de pontos.
Positive Dots %	51	Este KPI indica a percentagem de pontos do gráfico em que a diferença da espessura entre os pares de pontos é positiva.
Negative Dots %	51	Este KPI indica a percentagem de pontos do gráfico em que a diferença da espessura entre os pares de pontos é negativa.
Assimetry Differences Mean	51	Este KPI indica a média das diferenças de espessura entre os pares de pontos.
Assimetry Differences Std. Deviation	51	Este KPI indica o desvio padrão associado à média das diferenças de espessura entre os pares de pontos.
Maximum Assimetry Differences	51	Este KPI indica o valor máximo das diferenças de espessura entre os pares de pontos.
Assimetry P-Value T-Test	51	Este KPI indica o P-Value para o "T-Test" realizado na 1ª parte da fase Transform do ETL.
Assimetry P-Value Wilcoxon-Test	51	Este KPI indica o P-Value para o "Wilcoxon-Test" realizado na 1ª parte da fase Transform do ETL.
Filtros		
Nome	Figura	Conteúdo
Date Range	-	Este parâmetro filtra os dados quanto ao intervalo de tempo que se pretende analisar.
Perfil Dimensions ID	-	Este parâmetro filtra os dados quanto ao tipo de perfil que se pretende analisar.
Production ID	-	Este parâmetro filtra os dados quanto à ordem de produção que se pretende analisar.
Y Axis Zoom	51	Este parâmetro define o valor do eixo Y do gráfico.

7. DESENVOLVIMENTO DA ALARMÍSTICA

Apesar dos relatórios permitirem analisar e monitorizar os dados, estes são incapazes de monitorizar em tempo real os dados, i.e. é necessário o colaborador abrir o *dashboard*, atualizar, filtrar os dados e analisar os KPI. Como referenciado no subcapítulo 3.2, após a vulcanização o pneu deixa de poder ser reintroduzido no processo produtivo, caso possua algum defeito. Logo, é importantíssimo desenvolver um sistema que permita sinalizar, aos colaboradores, se algum estado crítico é atingido pelo perfil dos pneus, pisos ou paredes, antes que o pneu ultrapasse a vulcanização.

Com o intuito de assegurar monitorização em tempo real, foi desenvolvido um sistema de alarmísticas que atua regularmente, de vinte em vinte em minutos e, permite alertar os colaboradores do departamento de qualidade que um estado crítico foi atingido.

7.1 Seleção dos KPI a Monitorizar

O primeiro passo para criar o sistema consistiu em selecionar quais os indicadores de performance a monitorizar. Face à hierarquia dos KPI resultante do AHP, foram selecionados os seguintes indicadores:

1. Diferença percentual entre a área do perfil real e a área do perfil nominal, perfil total;
2. Diferença percentual entre a área do perfil real e a área do perfil nominal, lado esquerdo;
3. Diferença percentual entre a área do perfil real e a área do perfil nominal, lado direito.

Os níveis de tolerância para cada KPI, i.e. a percentagem mínima que é selecionada para que o alerta seja emitido foi definida. Ainda assim, uma vez mais, por razões de confidencialidade não se pode revelar o valor.

Para além destes três indicadores também foi decidido que se iria monitorizar a tabela `PERF.Application_Logs`. Relembrando, sempre que ocorre um erro durante a execução do ETL é inserida a documentação desse erro nesta tabela. Por tal facto, os colaboradores pretendem ser alertados, sempre que um erro ocorre durante o ETL.

7.2 Comparação e Seleção do *Software*

O passo seguinte compreendeu a comparação de diversos *softwares* que se encontram ativos na empresa, que permitem alcançar os objetivos estipulados, e a seleção de um destes para desenvolver o sistema de alarmísticas. Dos *softwares* internos utilizados na empresa, dois preenchem os requisitos:

Grafana

O Grafana é um *software* poderoso para visualização de dados e monitorização em tempo real, capaz de extrair informações valiosas de uma variedade de fontes de dados. Com uma interface intuitiva e flexível, esta plataforma permite criar painéis personalizados e interativos que exibem informações em tempo real de forma clara e visualmente apelativa. Com suporte para várias bases de dados, serviços de *cloud* e protocolos de comunicação, o Grafana oferece uma solução completa para monitorizar e analisar dados em ambientes complexos. Para além da vertente analítica, o Grafana permite internamente desenvolver alarmes para os indicadores que se encontram presentes nos seus diversos *dashboard* personalizados. Na figura 41 é ilustrado um caso real de um *dashboard* realizado a partir do Grafana, em que é possível analisar a diferença percentual entre a área do perfil real e a área do perfil nominal, para a corrida total e para o 1º quarto da corrida.

Relativamente à secção de alertas, foi desenvolvido um sistema de alarmísticas a partir do Grafana para se conseguir estudar com maior detalhe as vantagens e desvantagens do mesmo, no âmbito do projeto. A plataforma permite sinalizar quando cada um dos quatro cenários pretendidos, tanto os três cenários relativos à diferença percentual da área, como o cenário relacionado com o erro. O sistema é capaz de analisar, de vinte em vinte minutos, os dados presentes na DW como pretendido e, caso alguma das quatro condições predefinidas para cada alarme seja validada, o sistema irá enviar um *email* para um leque de colaboradores, a documentar o estado crítico atingido. O sistema, no conteúdo



Figura 41 - Dashboard desenvolvido pelo Grafana.

No alerta relacionado com o erro do ETL, o *email* e o SMS informam: que tipo de erro ocorreu; a que ordem de produção e extrusora diz respeito; a que horas estes dados foram armazenados na DW; qual a documentação do erro. Na figura 46 é possível analisar um caso real que ocorreu na empresa.

From: italerts@conti.de

IT Alerts: PERF_Area_LEFT_EXE09

Data Base	Production_ID	AVERAGE	Datetime	Upload_Time
PLT_PREPARATION_EXE09	982420104100_2023-03-03 13-57-05	⚠️ [REDACTED]	03/03/2023 14:00:15	16/05/2023 17:00:00
PLT_PREPARATION_EXE09	982420103900_2023-03-03 22-37-10	⚠️ [REDACTED]	03/03/2023 22:42:32	16/05/2023 16:58:00
PLT_PREPARATION_EXE09	982420103200_2023-03-03 14-32-20	⚠️ [REDACTED]	03/03/2023 14:56:50	16/05/2023 16:48:00

Datetime: 16/05/2023 17:07:33

Figura 44 – Alerta IT Soft., *PERF_Area_LEFT_EXE09*

Apesar do visual apelativo, escalável e complexo do Grafana, a informação de que o mesmo é capaz de dispor é bastante reduzida comparativamente ao *software* desenvolvido pelo departamento de IT. Para além disso, enquanto o Grafana apenas permite passar informação aos colaboradores por *email*, o *software* do departamento de IT permite sinalizar por *email* e SMS. Tendo em conta esta superioridade face ao Grafana, decidiu manter-se implementado o *software Continental Mabor's IT Alert Software*.

From: italerts@conti.de

IT Alerts: PERF_Area_RIGHT_EXE09

Data Base	Production_ID	AVERAGE	Datetime	Upload_Time
PLT_PREPARATION_EXE09	982420103900_2023-03-03 22-37-10	⚠️ [REDACTED]	03/03/2023 22:42:32	16/05/2023 16:58:00
PLT_PREPARATION_EXE09	982420103200_2023-03-03 14-32-20	⚠️ [REDACTED]	03/03/2023 14:56:50	16/05/2023 16:48:00

Datetime: 16/05/2023 17:07:33

Figura 45 – Alerta IT Soft., *PERF_Area_RIGHT_EXE09*

Com a seleção do sistema de alarmística a manter, dá-se por completa a implementação de cada uma das soluções ilustradas pela tabela 6.

From: italerts@conti.de

IT Alerts: PERF_Logs_Alert_EXE09

ID	Upload_Time	Work_Center	Production_ID	Application_ETL	Log_Type	Message
4	16/05/2023 17:01:03	EXE09	9894201029933/2023-04-19 05-44-56 (2)	log_BI_PLT_ETL_PERF.txt	Error	Error trying to build the dataframes within the order ID: "9894201029933/2023-04-19 05-44-56 (2)".
5	16/05/2023 17:01:04	EXE09	9924201029912/2023-04-19 05-44-56 (2)	log_BI_PLT_ETL_PERF.txt	Error	Error trying to build the dataframes within the order ID: "9924201029912/2023-04-19 05-44-56 (2)".

Datetime: 16/05/2023 17:07:33

Figura 46 – Alerta IT Soft., *PERF_Logs_Alert_EXE09*

8. ANÁLISE DE RESULTADOS

Após a conclusão de todas as etapas propostas é, agora, necessário analisar os resultados obtidos para cada uma delas. Deste modo, poder-se-á avaliar para cada solução a presença ou ausência de sucesso.

8.1 Hierarquização dos KPI

Ao longo da dissertação realçou-se que não é a partir da quantidade de KPI, que se encontram a ser monitorizados por uma empresa, que é definida a qualidade do seu departamento de *Business Intelligence*. O valor que o departamento de BI traz à empresa é diretamente proporcional à qualidade do suporte que os dados dão à mesma, cada vez que uma decisão estratégica necessita de ser tomada.

A implementação do AHP foi considerada um sucesso, pois permitiu que se reunissem vários colaboradores da empresa num só local, com um fim e objetivo comum. A partir da discussão, do conhecimento partilhado e experiência do grupo, selecionou-se um leque reduzido de KPI, que na visão dos colaboradores representam um suporte essencial em qualquer decisão que necessite de ser tomada e que de forma mais ou menos relevante é influenciada pela qualidade dos perfis. Obviamente, a redução de KPI a monitorizar impulsionou a redução do número de relatórios e, bem como do tempo alocado à análise das métricas que os compõem.

Outra vantagem, menos quantificável, que o AHP trouxe à empresa, consistiu no facto de vários colaboradores com nível hierárquico de destaque, na empresa, terem estado a par deste processo e presenciaram, em primeira mão, os benefícios do mesmo. Este contacto direto permite que os colaboradores conseguiram levar o ideal do AHP e as suas mais-valias para os restantes processos empresariais.

8.2 Processo de ETL

O ETL desenvolvido e em vigor na empresa, permitiu reduzir a quantidade de dados não visíveis para duas das máquinas mais relevantes no processo produtivo de pneus. A visibilidade anteriormente nula passou a ser parcialmente total. O termo parcialmente total vem em concordância com o que foi referido no capítulo 5. Apesar de apenas 33% e 3% da informação dos ficheiros TSX e NOX, respetivamente, ter transitado de invisível para visível, a verdade é que esta percentagem de dados permite satisfazer os requisitos essenciais indicados pelo departamento de qualidade. Isto é, os

diferentes *clusters* de informação que se tornaram visíveis permitiram construir todos os KPI mais importantes, selecionados pelo AHP.

Para além disso, a não estruturação dos dados limitava os colaboradores aos KPI que eram fornecidos pelo *software TRISCAN*, limitados e desajustados aos objetivos e prioridades da Continental. Atualmente, a partir do ETL é possível armazenar as métricas num formato tabelar, normalizado, possibilitando configurar todo um novo leque de KPI, a partir destes.

A visibilidade e estrutura que o ETL dá aos dados, resulta num conjunto de oportunidades que permitem aos colaboradores estudar tendências e correlações entre os dados e os resultados que, eventualmente, vão permitir tomar decisões que aumentarão a eficácia do processo produtivo. Face às inúmeras vantagens que o ETL põe ao dispor da empresa, sugere-se que este processo seja implementado nas restantes extrusoras instaladas em chão de fábrica.

Apesar dos indicadores de performance determinados pelo AHP estarem ativos, poderá existir informação presente nos restante 73% e 96% dos dados do ficheiro TSX e NOX, respetivamente, que traga valor adicional à empresa face aos KPI que necessita de monitorizar. Daí que, por forma a incrementar estas percentagens é aconselhável que se entre em contacto com os engenheiros da empresa *Ehardt Leimer*, empresa que produziu as extrusoras e, conseqüentemente, o *software TRISCAN*, para que a partir da cooperação entre empresas seja possível identificar a estrutura e os atributos dos restantes dados presentes nos dois tipos de ficheiros.

8.3 Relatório Interativos

Apesar do processamento presente no ETL facilitar o cálculo dos KPI pretendidos, na realidade, os atributos presentes na DW, por si só, não representam os indicadores de performance, ficando por realizar a fase final do cálculo dos KPI.

A implementação dos diversos relatórios em Power Bi, não só permite que estes cálculos sejam realizados, armazenando os indicadores finais em formato de tabela, como também, garante toda uma arquitetura interativa que os permite analisar num ambiente favorável ao sucesso do processo de BI. A tabela 18 resume as principais vantagens que foram alcançadas com o desenvolvimento dos relatórios.

A par da sugestão presente no subcapítulo 8.2, de expandir o ETL para as restantes extrusoras, é necessário que sejam duplicados os relatórios *BME* e *Last 15 Days* para cada uma das novas máquinas, caso esta expansão ocorra.

Tabela 18 - Resultados obtidos a partir da implementação de relatórios usando o Power Bi.

Vantagens Associadas aos Reports em Power Bi	
Antes do Projeto	Após a Projeto
O software ELTRISCAN não permitia analisar os KPI olhando à corrida como um todo.	Os <i>reports</i> permitem analisar os KPI olhando à corrida como um todo e, caso pretendido, a fases da mesma.
O software ELTRISCAN não analisa os KPI que representam maior valor para a empresa.	Os <i>reports</i> analisam os KPI que representam maior valor para a empresa. Permitindo, também filtrar estes consoante parâmetros previamente estabelecidos, como o lado do perfil.
O software ELTRISCAN importa os dados a partir de ficheiros locais, recorrendo a um processo manual e moroso.	Os <i>reports</i> importam os dados diretamente da DW, recorrendo a um processo automático e rápido.
O software ELTRISCAN é estático.	Os <i>reports</i> são interativos.
O software ELTRISCAN não é intuitivo.	Os reports são intuitivos.
O software ELTRISCAN não possui suporte técnico acessível.	Os reports possuem suporte técnico acessível interno à empresa.

8.4 Sistema de Alarmísticas

Os colaboradores do departamento de qualidade possuem diversas tarefas a desempenhar, diariamente: reuniões, planeamento e análise de relatórios, entre outras. É humanamente impossível para eles analisar, permanentemente, todos os relatórios ativos na empresa. A falta de monitorização contínua dos processos resulta na inexistência de deteção de falhas ou problemas, o que impossibilita a correção imediata, acabando por se produzir *scrap*.

Com a implementação do sistema de alarmes nos processos produtivos das extrusoras sete e nove, os colaboradores alcançam uma monitorização contínua em tempo real. Através dos parâmetros de "disparo" dos alertas. Os colaboradores são notificados sempre que um estado crítico é atingido. Essas notificações, enviadas por e-mail ou SMS, permitem que os mesmos reajam, imediatamente, possibilitando quer a correção dos perfis de paredes ou pisos, quer a reutilização da sua borracha, noutras etapas do processo de produção de pneus.

No entanto, apesar do sistema estar ativo ao longo do dia, verificando os parâmetros de "disparo" a cada vinte minutos, o termo "monitorização em tempo real" usado para descrevê-lo é um tanto subjetivo. Neste caso específico, devido à execução do ETL a cada quatro horas, é possível que os dados que disparam o alarme estejam relacionados com uma ordem de produção ocorrida há quatro horas. Devido à complexidade e duração do processo ETL, o intervalo entre as iterações do mesmo, precisa de ser longo para evitar a sobreposição entre iterações que resulta na perda de informação. Sugere-se uma análise criteriosa à possibilidade de reduzir o intervalo entre as iterações do ETL para duas horas, utilizando uma nova base de dados de teste. Dado que o tempo médio do ETL, tanto para a extrusora sete quanto para a nove, é inferior a cento e dez minutos, é provável que esse intervalo seja adequado.

9. CONCLUSÃO

Ao longo desta dissertação, foi explorada a importância da implementação de um sistema abrangente de monitorização e controlo de qualidade para os perfis de paredes e pisos utilizados na produção de pneus na Continental Mabor. Com base nos estudos realizados e nas abordagens adotadas durante o projeto, foi possível concluir que a implementação de tal sistema tem o potencial de trazer benefícios significativos para a empresa, tanto em termos de redução do desperdício quanto, como na melhoria da qualidade dos produtos finais.

Através do processo de Extração, Transformação e Armazenamento dos dados provenientes das máquinas extrusoras, tornaram-se visíveis e acessíveis informações valiosas, ocultadas, anteriormente, nos arquivos gerados por essas máquinas. Essa etapa foi fundamental para a análise e o acompanhamento dos indicadores de performance relacionados com perfis de paredes e pisos.

Com a utilização da ferramenta Power BI, conseguiu-se desenvolver relatórios que permitiram um monitoramento mais eficiente e detalhado dos KPI de interesse. Esses relatórios forneceram ao departamento de qualidade informações relevantes sobre a espessura, área e assimetria dos perfis, possibilitando tomadas de decisão sustentadas por informação real e concreta e, uma abordagem proativa na deteção de falhas durante o processo de produção.

A implementação de um sistema de alarmísticas também desempenhou um papel crucial na monitorização em tempo real dos dados de qualidade. Ao sinalizar desvios significativos nos KPI ou erros durante o processo de ETL, o sistema de alarme permite uma intervenção imediata, evitando perdas e garantindo a melhoria contínua da qualidade dos produtos.

Embora os resultados obtidos tenham sido promissores e tenham comprovado a eficácia do sistema implementado, é importante realçar que existe espaço para melhoria. Planeia-se realizar reuniões adicionais com os fornecedores das extrusoras para obter informações mais detalhadas sobre os dados não visíveis presentes nos arquivos, bem como expandir a implementação do processo de ETL para abranger todas as extrusoras em produção.

Por outro lado, recomenda-se uma análise mais aprofundada dos parâmetros do sistema de alarme, a fim de otimizar a sua configuração e aumentar a sua eficácia na deteção de desvios significativos face à qualidade do produto. Essas ações futuras contribuirão para uma monitorização ainda mais abrangente e uma tomada de decisão mais informada, no que diz respeito à qualidade dos perfis de paredes e pisos.

Em suma, esta dissertação demonstrou que a implementação de um sistema abrangente de monitorização e controle de qualidade é fundamental para reduzir o desperdício e melhorar a qualidade

dos produtos finais na produção de pneus, na Continental Mabor. A combinação do processo de ETL, dos relatórios e do sistema de alarme provou ser eficaz na identificação e correção de desvios, permitindo à empresa tomar medidas proativas para garantir a excelência da qualidade. O trabalho realizado abre caminho para futuras melhorias e aperfeiçoamentos contínuos, fortalecendo a competitividade da empresa no mercado e o seu compromisso com a satisfação dos clientes.

REFERÊNCIAS BIBLIOGRÁFICAS

- Alaqueel, T. A., & Suryanarayanan, S. (2018). A fuzzy Analytic Hierarchy Process algorithm to prioritize Smart Grid technologies for the Saudi electricity infrastructure. *Sustainable Energy, Grids and Networks*, *13*, 122–133. <https://doi.org/10.1016/j.segan.2017.12.010>
- Alcacer, V., & Cruz-Machado, V. (2019). Scanning the Industry 4.0: A Literature Review on Technologies for Manufacturing Systems. *Engineering Science And Technology - An International Journal*, *22*(3), 899–919. <https://doi.org/10.1016/j.jestch.2019.01.006>
- Ali, W., Shafique, M. U., Majeed, M. A., & Raza, A. (2019). Comparison between SQL and NoSQL Databases and Their Relationship with Big Data Analytics. *Asian Journal of Research in Computer Science*, 1–10. <https://doi.org/10.9734/ajrcos/2019/v4i230108>
- Alonso, S. (2022). *Você sabe o que é KPI e como aplicá-la em seu negócio?* Promadjr. <https://www.promadjr.com/post/voc%C3%AA-sabe-o-que-%C3%A9-kpi-e-como-aplic%C3%A1-la-em-seu-neg%C3%B3cio>
- Amirat, Y., Benbouzid, M. E. H., Al-Ahmar, E., Bensaker, B., & Turri, S. (2009). A brief status on condition monitoring and fault diagnosis in wind energy conversion systems. *Renewable and Sustainable Energy Reviews*, *13*(9), 2629–2636. <https://doi.org/10.1016/j.rser.2009.06.031>
- Bauer, H., Baur, C., Camplone, G., & George, K. (2015). *Industry 4.0 How to navigate digitization of the manufacturing sector*. McKinsey Digital. <https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/Operations/Our%20Insights/Industry%2040%20How%20to%20navigate%20digitization%20of%20the%20manufacturing%20sector/Industry-40-How-to-navigate-digitization-of-the-manufacturing-sector.ashx>
- Bean, R. (2017). *How Companies Say They're Using Big Data*. <https://Hbr.Org/2017/04/How-Companies-Say-Theyre-Using-Big-Data>.
- Biehn, N. (2015, January 10). *The missing V's in big data: Viability and value*. WIRED. <https://Www.Wired.Com/Insights/2013/05/the-Missing-vs-in-Big-Data-Viability-and-Value/>.
- Bontempo, C., & Zagelow, G. (1998). The IBM Data Warehouse Architecture. *Communications of the ACM*, *41*(9), 38–48. <https://doi.org/10.1145/285070.285078>
- Borges, D. (2020). *DIFERENÇA ENTRE OLTP E OLAP*. <https://Medium.Com/@david.Borges/Diferen%C3%A7a-Entre-Oltp-e-Olap-B6dab6f924b8>.
- Bosi, F., Corradi, A., Foschini, L., & Monti, S. (2019). *Cloud Enabled Smart Data Collection in Shop Floor Environments for Industry 4.0*. <https://dx.doi.org/10.1109/WFCS.2019.8757952>

- Caetano, M. (2019). *Compostos de borracha*. Ctborracha. <https://www.ctborracha.com/borracha-sintese-historica/aplicacoes/pneus/o-fabrico-de-pneus/componentes/>
- Chandarana, P., & Vijayalakshmi, M. (2014). Big data analytics frameworks. *2014 International Conference on Circuits, Systems, Communication and Information Technology Applications, CSCITA 2014*, 430–434. <https://doi.org/10.1109/CSCITA.2014.6839299>
- Chaudhuri, S., & Dayal, U. (1997). *An overview of data warehousing and OLAP technology*. <https://doi.org/10.1145/248603.248616>
- Chenhall, R. H. (2005). Integrative strategic performance measurement systems, strategic alignment of manufacturing, learning and strategic outcomes: an exploratory study. *Accounting, Organizations and Society*, 30(5), 395–422. <https://doi.org/https://doi.org/10.1016/j.aos.2004.08.001>
- Codd, E. (1998). Providing OLAP to User-Analysts: An IT Mandate. *Hyperion Solutions Corporation*. http://www.estgv.ipv.pt/paginaspessoais/jloureiro/esi_aid2007_2008/fichas/codd.pdf
- Costa, M. (2022). *Estudo da estabilidade dimensional na co-extrusão de pisos para pneus* [Universidade do Minho]. <https://hdl.handle.net/1822/65495>
- Cristea, C., & Cristea, M. (2021). KPIs for operational performance assessment in flexible packaging industry. *Sustainability (Switzerland)*, 13(6). <https://doi.org/10.3390/su13063498>
- Cuzzocrea, A., Bellatreche, L., & Song, I.-Y. (2013). Data Warehousing and OLAP over Big Data: Current challenges and future research directions. *International Conference on Information and Knowledge Management, Proceedings*, 67–70. <https://doi.org/10.1145/2513190.2517828>
- Cuzzocrea, A., Song, I.-Y., & Davis, K. C. (2011). Analytics over large-scale multidimensional data: The big data revolution! *14th International Workshop on Data*, 101–104. https://www.researchgate.net/publication/312449460_Analytics_over_large-scale_multidimensional_data_The_big_data_revolution
- Doug, L. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, 6. <https://www.bibsonomy.org/bibtex/263868097d6e1998de3d88fcbb7670ca6/refshah>
- Elhuni, R. M., & Ahmad, M. M. (2017). Key Performance Indicators for Sustainable Production Evaluation in Oil and Gas Sector. *Procedia Manufacturing*, 11, 718–724. <https://doi.org/https://doi.org/10.1016/j.promfg.2017.07.172>
- Ellars, S. (2013, September 12). *Beyond Volume, Variety and Velocity is the Issue of Big Data Veracity*. Insidebigdata. <https://insidebigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/>

- Elmasri, R., & Navathe, S. (2000). *Fundamentals of Database Systems* (6th ed.).
- Fatima, H., & Wasnik, K. (2016, July 2). Comparison of SQL, NoSQL and NewSQL databases for internet of things. *IEEE Bombay Section Symposium 2016: Frontiers of Technology: Fuelling Prosperity of Planet and People, IBSS 2016*. <https://doi.org/10.1109/IBSS.2016.7940198>
- Ferraioli, J., & Burke, R. (2018). *Drowning in data but starving for insights*. Deloitte. https://www2.deloitte.com/content/dam/insights/us/articles/4404_Drowning-in-data-but-starving-for-insights/DI_Drowning-in-data.pdf
- Franco, T. (2019). *OLTP vs OLAP: what's the difference between them?* Imaginary Cloud. <https://www.imaginarycloud.com/blog/oltp-vs-olap/>
- Franks, B., & Davenport, T. (2012). *Taming The Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics*. Wiley.
- Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). *Knowledge Discovery in Databases: An Overview*. <https://doi.org/10.1609/aimag.v13i3.1011>
- Fujishima, M., Ohno, K., Nishikawa, S., Nishimura, K., Sakamoto, M., & Kawai, K. (2016). Study of sensing technologies for machine tools. *CIRP Journal of Manufacturing Science and Technology*, 14, 71–75. <https://doi.org/10.1016/J.CIRPJ.2016.05.005>
- Gardner, S. R. (1998). Building the data warehouse. *Communications of the ACM*, 41(9), 52–60. <https://doi.org/10.1145/285070.285080>
- Gent, A., & Walter, J. (2006). *Mechanical Engineering Faculty Research* (Vol. 854). http://ideaexchange.uakron.edu/mechanical_ideashttp://ideaexchange.uakron.edu/mechanical_ideas/854
- Godinho, E., & Piedade, M. (2022). *ACCEPT – Web Applications for visualization and analytics of shop floor data* (pp. 1–6). https://iconline.ipleiria.pt/bitstream/10400.8/6221/1/ACCEPT__Web_Applications_for_visualization_and_analytics_of_shop_floor_data.pdf
- Goes, P. B. (2014). Big Data and IS Research. *MIS QUARTERLY*, 38(3), III–VIII. https://www.researchgate.net/publication/293109828_Big_Data_and_IS_Research
- Hackathorn, R. (2004). The BI watch real-time to real-value. *DM Review*. https://www.researchgate.net/publication/228498840_The_BI_watch_real-time_to_real-value
- Han, J., E, H., Le, G., & Du, J. (2011). Survey on NoSQL database. *2011 6th International Conference on Pervasive Computing and Applications*, 363–366. <https://doi.org/10.1109/ICPCA.2011.6106531>

- Höller, Jan. (2014). *From machine to machine to the internet of things : introduction to a new age of intelligence*. Academic Press. <https://doi.org/10.1016/C2012-0-03263-2>
- Hozdić, E. (2015). Smart Factory For Industry 4.0: A Review. In *International Journal of Modern Manufacturing Technologies: Vol. VII* (Issue 1). https://www.researchgate.net/publication/282791888_Smart_factory_for_industry_40_A_review
- IDC. (2020). *Worldwide Global DataSphere Forecast, 2020*. Market Research. <https://www.marketresearch.com/IDC-v2477/Worldwide-Global-DataSphere-Forecast-Keeps-14315439/>
- Inmon, W. H. (1992). Building the Data Warehouse. In *New York: Warehouse Wiley* (4th ed.). WILEY.
- Inmon, W. H. (1996). Data warehouse and data mining. *Communications of the ACM*, 39(11), 49–50. <https://doi.org/10.1145/240455.240470>
- Jukic, N., Sharma, A., Nestorov, S., & Jukić, B. (2015). Augmenting Data Warehouses with Big Data. *Information Systems Management*, 32(3), 200–209. <https://doi.org/10.1080/10580530.2015.1044338>
- Jukic, N., Vrbsky, S., & Nestorov, S. (2016). *Database Systems: Introduction to Databases and Data Warehouses* (1st ed.). Prospect Press.
- Juran, J. M., & Feo, J. A. De. (2010). *Juran's Quality Handbook: The Complete Guide to Performance Excellence, Sixth Edition* (6th ed.). McGraw-Hill Education. <https://www.accessengineeringlibrary.com/content/book/9780071629737>
- Juran, J. M., & Godfrey, B. (1999). *Juran's Quality Handbook*. McGraw-Hill.
- Kagermann, H., Wahlster, W., & Helbig, J. (2013). *Securing the future of German manufacturing industry - Recommendations for implementing the strategic initiative INDUSTRIE 4.0*. [https://www.scirp.org/\(S\(lz5mqp453edsnp55rrgjt55.\)\)/reference/referencespapers.aspx?referenceid=2966479](https://www.scirp.org/(S(lz5mqp453edsnp55rrgjt55.))/reference/referencespapers.aspx?referenceid=2966479)
- Karasakal, E., & Aker, P. (2017). A multicriteria sorting approach based on data envelopment analysis for R&D project selection problem. *Omega*, 73, 79–92. <https://doi.org/https://doi.org/10.1016/j.omega.2016.12.006>
- Király, T., Primusz, P., & Tóth, C. (2022). Simulation of Static Tyre–Pavement Interaction Using Two FE Models of Different Complexity. *Applied Sciences (Switzerland)*, 12(5). <https://doi.org/10.3390/app12052388>

- Klingenberg, C. O., Borges, M. A. V., & Antunes Jr., J. A. V. (2021). Industry 4.0 as a data-driven paradigm: a systematic literature review on technologies. *Journal Of Manufacturing Technology Management*, 32(3, SI), 570–592. <https://doi.org/10.1108/JMTM-09-2018-0325>
- Krish Krishnan. (2013). *Data Warehousing in the Age of Big Data* (1st ed.). Morgan Kaufmann Publishers Inc.
- Lee, E. (2008). *Cyber Physical Systems: Design Challenges*. <http://www.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-8.html>
- Li, H., Ni, F., Dong, Q., & Zhu, Y. (2018). Application of analytic hierarchy process in network level pavement maintenance decision-making. *International Journal of Pavement Research and Technology*, 11(4), 345–354. <https://doi.org/https://doi.org/10.1016/j.ijprt.2017.09.015>
- Lu, B., Li, Y., Wu, X., & Yang, Z. (2009). A review of recent advances in wind turbine condition monitoring and fault diagnosis. *2009 IEEE Power Electronics and Machines in Wind Applications*, 1–7. <https://doi.org/10.1109/PEMWA.2009.5208325>
- Márquez, G., Tobias, A., & Roberts, C. (2010). Railway point mechanisms: Condition monitoring and fault detection. *Proceedings of The Institution of Mechanical Engineers Part F - Journal of Rail and Rapid Transit*. <https://doi.org/10.1243/09544097JRRT289>
- McAfee, A., & Brynjolfsson, E. (2012). Strategy & Competition Big Data: The Management Revolution. *Harvard Business Review*, 90(10), 60+.
- Min, S., Lidde, J., Raue, N., & Dornfeld, D. (2011). Acoustic emission based tool contact detection for ultra-precision machining. *CIRP Annals*, 60(1), 141–144. <https://doi.org/https://doi.org/10.1016/j.cirp.2011.03.079>
- Mishra, S. (2017, May 19). *Unsupervised Learning and Data Clustering*. <https://Towardsdatascience.Com/Unsupervised-Learning-and-Data-Clustering-Eeeeb78b422a>.
- Moniruzzaman, A. B. M., & Hossain, S. A. (2013). NoSQL Database: New Era of Databases for Big data Analytics-Classification, Characteristics and Comparison. In *International Journal of Database Theory and Application* (Vol. 6, Issue 4). <http://hortonworks.com/blog/7-key-drivers-for-the-big-data-market/>
- Mourtzis, D., Milas, N., & Vlachou, A. (2018). An Internet of Things-Based Monitoring System for Shop-Floor Control. *JOURNAL OF COMPUTING AND INFORMATION SCIENCE IN ENGINEERING*, 18(2). <https://doi.org/10.1115/1.4039429>

- Mourtzis, D., Vlachou, E., & Milas, N. (2016). Industrial Big Data as a Result of IoT Adoption in Manufacturing. *Procedia CIRP*, 55, 290–295. <https://doi.org/https://doi.org/10.1016/j.procir.2016.07.038>
- Myriantous, G. (2022). *Fact vs Dimension Tables*. <https://Towardsdatascience.com/Star-Schema-924b995a9bdf>.
- Nagorny, K., Lima-Monteiro, P., Barata, J., & Colombo, A. W. (2017). Big Data Analysis in Smart Manufacturing. *International Journal of Communications, Network and System Sciences*, 10(03), 31–58. <https://doi.org/10.4236/ijcns.2017.103003>
- Nakayama, R. S., de Mesquita Spínola, M., & Silva, J. R. (2020). Towards I4.0: A comprehensive analysis of evolution from I3.0. *Computers & Industrial Engineering*, 144, 106453. <https://doi.org/10.1016/J.CIE.2020.106453>
- Navathe, S., & Elmasri, R. (2005). *Sistemas de banco de dados* (1st ed.). Editora Afiliada.
- NEXEN TIRE. (2017). *Manufacturing Process*. Nexen. https://www.nexentire.com/international/information/tire_information/basic_sense/process.php
- O'Brien, R. (1998). *An Overview of the Methodological Approach of Action Research*. <http://www.web.ca/~robrien/papers/arfinal.html>
- Oliveira, V. F., Pessoa, M. A. de O., Junqueira, F., & Miyagi, P. E. (2022). Sql and nosql databases in the context of industry 4.0. *Machines*, 10(1). <https://doi.org/10.3390/machines10010020>
- Prathima, B., Sudha, P., & Suresh, P. (2020). Shop floor to cloud connect for live monitoring the production data of CNC machines. *International Journal of Computer Integrated Manufacturing*, 33(2), 142–158. <https://doi.org/10.1080/0951192X.2020.1718762>
- Prince, G., & Shortland, M. (1987). Condition monitoring: The automated manufacturing perspective. *British Journal of Non-Destructive Testing*. <https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=7523151>
- Ramakrishnan, R. (2008). *Sistemas de Gerenciamento de Bancos de Dados* (Vol. 1). AMGH.
- Rostek, K. (2010). Data Analytical Processing in Data Warehouses. *Foundations of Management*, 2(1), 99–116. <https://doi.org/10.2478/v10238-012-0023-x>
- Rubmann, M., & Lorenz, M. (2015). *Industry 4.0: The Future of Productivity and Growth in Manufacturing Industries*. BCG. https://www.bcg.com/pt-br/publications/2015/engineered_products_project_business_industry_4_future_productivity_growth_manufacturing_industries

- Sadalage, P., & Fowler, M. (2013). *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence* (Vol. 1). Addison-Wesley Professional.
- Shi, D., Lee, Y., Duan, X., & Wu, Q. H. (2001). Power system data warehouses. In *IEEE Computer Applications in Power* (Vol. 14, Issue 3, pp. 49–55). <https://doi.org/10.1109/MCAP.2001.952937>
- Shi, Z., Xie, Y., Xue, W., Chen, Y., & Fu Liulu and Xu, X. (2020). Smart factory in Industry 4.0. *SYSTEMS RESEARCH AND BEHAVIORAL SCIENCE*, 37(4, SI), 607–617. <https://doi.org/10.1002/sres.2704>
- Spitzer, D. (2007). *Transforming Performance Measurement: Rethinking the Way We Measure and Drive Organizational Success*. AMACOM.
- Srivastava, J., & Chen, P. Y. (1999). Warehouse creation a potential roadblock to data warehousing. *IEEE Transactions on Knowledge and Data Engineering*, 11(1), 118–126. <https://doi.org/10.1109/69.755620>
- Stock, T., & Seliger, G. (2016). Opportunities of Sustainable Manufacturing in Industry 4.0. In G. Seliger, H. Kohl, & J. Mallon (Eds.), *13TH GLOBAL CONFERENCE ON SUSTAINABLE MANUFACTURING - DECOUPLING GROWTH FROM RESOURCE USE* (Vol. 40, pp. 536–541). <https://doi.org/10.1016/j.procir.2016.01.129>
- Stonebraker, M. (2011). Stonebraker on data warehouses. *Communications of the ACM*, 54(5), 10–11. <https://doi.org/10.1145/1941487.1941491>
- StreamSets. (2021). *Explained: Stream Processing, Streaming Data, and Streaming Data Pipelines*. StreamSets. <https://streamsets.com/learn/stream-processing-streaming-data/>
- Susman, G. (1978). *An Assessment of the Scientific Merits of Action Research*. Sage Publications, Inc. <https://www.jstor.org/stable/2392581>
- Taticchi, P., Tonelli, F., & Cagnazzo, L. (2010). Performance measurement and management: a literature review and a research agenda. *Measuring Business Excellence*, 14(1), 4–18. <https://doi.org/10.1108/13683041011027418>
- Teorey, T., Sam Lightstone, S., Nadeau, T., & Jagadish, H. (2011). *Database Modeling and Design* (5th ed.). Logical Design.
- Teti, R., Mourtzis, D., D'Addona, D. M., & Caggiano, A. (2022). Process monitoring of machining. *CIRP Annals*, 71(2), 529–552. <https://doi.org/10.1016/j.cirp.2022.05.009>
- Unutmaz Durmuşoğlu, Z. D. (2018). Assessment of techno-entrepreneurship projects by using Analytical Hierarchy Process (AHP). *Technology in Society*, 54, 41–46. <https://doi.org/https://doi.org/10.1016/j.techsoc.2018.02.001>

- Ventura, A. (2012). *Aplicação do método Action Research ao processo de apoio à decisão com o uso do sistema VIP Analysis*. <https://core.ac.uk/download/pdf/19133892.pdf>
- Victor, K. (2018). *An evaluation on how interactive visualizations can enhance a brewer's cognition of brewing data*. <https://www.diva-portal.org/smash/get/diva2:1233704/FULLTEXT01.pdf>
- Wan, J., Cai, H., & Zhou, K. (2015). Industrie 4.0: Enabling technologies. *Proceedings of 2015 International Conference on Intelligent Computing and Internet of Things*, 135–140. <https://doi.org/10.1109/ICAIOT.2015.7111555>
- Wang, L., & Gao, R. (2006). *Condition Monitoring and Control for Intelligent Manufacturing*. Springer. <https://doi.org/10.1007/1-84628-269-1>
- Wang, S., Wan, J., Li, D., & Zhang, C. (2016). Implementing Smart Factory of Industrie 4.0: An Outlook. *International Journal Of Distributed Sensor Networks*. <https://doi.org/10.1155/2016/3159805>
- Watson, H., Ariyachandra, T., & Matyska, R. J. (2001). Data warehousing stages of growth. *Information Systems Management*, 18(3), 42–50. <https://doi.org/10.1201/1078/43196.18.3.20010601/31289.6>
- Wohlers, B., Dziwok, S., & Bremer, C. (2017). *Improving the product control of mechatronic systems using key performance indicators*. <https://doi.org/10.12783/dtetr/icpr2017/17630>
- Wohlers, B., Dziwok, S., Pasic, F., Lipsmeier, A., & Becker, M. (2020). Monitoring and control of production processes based on key performance indicators for mechatronic systems. *International Journal of Production Economics*, 220. <https://doi.org/10.1016/j.ijpe.2019.07.025>

ANEXO A

Tabela 19 - Escala Valorativa do AHP. Retirado de: (Cristea & Cristea, 2021)

Numeric Scale	Definition
1	Equal importance
3	Moderate importance
5	Strong importance
7	Very strong importance
9	Extreme importance
2, 4, 6, 8	Intermediate values between the two adjacent judgements

APÊNDICE A – ESQUEMA DAS TABELAS

Tabela 20 - Esquema da tabela *Header*

Atributo	Descrição	Tipo de Dados	UN.
ID	ID incrementável	int	-
Upload_Time	Instante de Armazenamento para cada linha	<i>Datetime</i>	-
Work_Center	Posto de trabalho	varchar(50)	-
Production_ID	Chave criada pelo <i>developer</i> da script de modo a facilitar JOINS na base de dados	varchar(50)	-
Perfils_Dimension_ID	Tipo de perfil	varchar(50)	-
Date_Time	Instante temporal em que o perfil foi analisado pelo sensor	<i>Datetime</i>	-
Has_Out	Sinaliza os ficheiros TSX que representam estados críticos	varchar(50)	-
Is_Beginning_Mid_End	Indica se o instante representa o início, meio ou fim da corrida	varchar(50)	-
Denotation	Denotação da máquina	varchar(50)	-
soft_Total_Width_mm	Largura total do perfil real	float	mm
Nominal_Total_Width_mm	Largura total do perfil <i>target</i>	float	mm
py_Left_Width_mm	Largura do lado esquerdo do perfil real	float	mm
Nominal_Left_Width_mm	Largura do lado esquerdo do perfil <i>target</i>	float	mm
py_Right_Width_mm	Largura do lado direito do perfil real	float	mm
Nominal_Right_Width_mm	Largura do lado direito do perfil <i>target</i>	float	mm
soft_Total_Area_mm2	Área total do perfil medida pelo <i>software TRISCAN</i>	float	mm
soft_Left_Area_mm2	Área do lado esquerdo do perfil medida pelo <i>software TRISCAN</i>	float	mm

soft_Total_Area_mm2	Área do lado direito do perfil medida pelo <i>software TRISCAN</i>	float	mm
py_Total_Area_mm2	Área total do perfil medida pela script em python	float	mm
py_Left_Area_mm2	Área do lado esquerdo do perfil medida pela script em python	float	mm
py_Right_Area_mm2	Área do lado direito do perfil medida pelo script em python	float	mm
Nominal_Total_Area_mm2	Área total do perfil <i>target</i>	float	mm
Total_dif_perc_area_Total	Diferença percentual entre a área do perfil real e a área do perfil <i>target</i> – corrida inteira	float	mm
Total_dif_perc_area_1_4	Diferença percentual entre a área do perfil real e a área do perfil <i>target</i> – 1ª parte da corrida	float	mm
Total_dif_perc_area_2_4	Diferença percentual entre a área do perfil real e a área do perfil <i>target</i> – 2ª parte da corrida	float	mm
Total_dif_perc_area_3_4	Diferença percentual entre a área do perfil real e a área do perfil <i>target</i> – 3ª parte da corrida	float	mm
Total_dif_perc_area_4_4	Diferença percentual entre a área do perfil real e a área do perfil <i>target</i> – 4ª parte da corrida	float	mm
Left_dif_perc_area_Total	Diferença percentual entre a área do perfil real e a área do perfil <i>target</i> – corrida inteira	float	mm
Left_dif_perc_area_1_4	Diferença percentual entre a área do perfil real e a área do perfil <i>target</i> – 1ª parte da corrida	float	mm
Left_dif_perc_area_2_4	Diferença percentual entre a área do perfil real e a área do perfil <i>target</i> – 2ª parte da corrida	float	mm
Left_dif_perc_area_3_4	Diferença percentual entre a área do perfil real e a área do perfil <i>target</i> – 3ª parte da corrida	float	mm
Left_dif_perc_area_4_4	Diferença percentual entre a área do perfil real e a área do perfil <i>target</i> – 4ª parte da corrida	float	mm
Right_dif_perc_area_Total	Diferença percentual entre a área do perfil real e a área do perfil <i>target</i> – corrida inteira	float	mm
Right_dif_perc_area_1_4	Diferença percentual entre a área do perfil real e a área do perfil <i>target</i> – 1ª parte da corrida	float	mm
Right_dif_perc_area_2_4	Diferença percentual entre a área do perfil real e a área do perfil <i>target</i> – 2ª parte da corrida	float	mm
Right_dif_perc_area_3_4	Diferença percentual entre a área do perfil real e a área do perfil <i>target</i> – 3ª parte da corrida	float	mm
Right_dif_perc_area_4_4	Diferença percentual entre a área do perfil real e a área do perfil <i>target</i> – 4ª parte da corrida	float	mm
soft_Area_Difference_mm2	Diferença entre a área real e a área do perfil <i>target</i> usando as medidas calculadas pelo <i>software TRISCAN</i>	float	mm
py_Area_Difference_mm2	Diferença entre a área real e a área do perfil <i>target</i> usando as medidas calculadas pela script python	float	mm

Tabela 21 - Esquema da tabela *Contour*

Atributo	Descrição	Tipo de Dados	UN.
ID	ID incrementável	int	-
Upload_Time	Instante de Armazenamento para cada linha	<i>Datetime</i>	-
Work_Center	Posto de trabalho	varchar(50)	-
Production_ID	Chave criada pelo <i>developer</i> da script de modo a facilitar JOINS na base de dados	varchar(50)	-
Perfils_Dimension_ID	Tipo de perfil	varchar(50)	-
Date_Time	Instante captado pelo sensor	<i>Datetime</i>	-
Cumulative_Width_mm	Indica cumulativamente a largura perfil ao longo dos pontos	float	mm
Contour_Thickness_mm	Espessura do perfil no ponto em questão	float	mm
Nominal_Thickness_mm	Espessura desejada do perfil no ponto em questão	float	mm
Is_Middle	Indica se este é ponto que representa o ponto que divide o perfil ao meio	varchar(50)	-
Is_Beginning_Mid_End	Indica se o instante representa o início, meio ou fim da corrida	varchar(50)	-
Is_Contour	Indica se aquela linha representa um ponto real do perfil	varchar(50)	-
What_Side	Indica a que lado representa este ponto do perfil	varchar(50)	-
assimetry_differences_ Pvalue_Paired_T_Test	Indica o <i>P-value</i> do teste de assimetria <i>T-Test</i> realizado para o instante em questão	float	-
assimetry_differences_ Pvalue_Paired_Wilcoxon	Indica o <i>P-value</i> do teste de assimetria <i>Wilcoxon</i> realizado para o instante em questão	float	-

Tabela 22 - Esquema da tabela *Date*

Coluna	Descrição	Tipo de Dados	UN.
ID	ID incrementável	int	-
Upload_Time	Instante de Armazenamento para cada linha	<i>Datetime</i>	-
Work_Center	Posto de trabalho	varchar(50)	-
Production_ID	Chave criada pelo <i>developer</i> da script de modo a facilitar JOINS na base de dados	varchar(50)	-
Perfils_Dimension_ID	Tipo de perfil	varchar(50)	-
Date_Time	Instante captado pelo sensor	<i>Datetime</i>	-
Year	Ano	float	-
Month	Mês	float	-
Day	Dia	float	-
Hour	Hora	float	-
Minute	Minutos	float	-
Seconds	Segundos	float	-
Thousandth	Milésima	float	-

Tabela 23 - Esquema da tabela *Dimensions*

Coluna	Descrição	Tipo de Dados	UN.
ID	ID incrementável	int	-
Upload_Time	Instante de Armazenamento para cada linha	<i>Datetime</i>	-
Work_Center	Posto de trabalho	varchar(50)	-

Production_ID	Chave criada pelo developer da script de modo a facilitar JOINS na base de dados	varchar(50)	-
Perfils_Dimension_ID	Tipo de perfil	varchar(50)	-
Date_Time	Instante captado pelo sensor	<i>Datetime</i>	-
Number_Marks	Número de pontos para o instante em questão	float	-
Cumulative_Width_mm	Indica, cumulativamente, a largura do perfil ao longo dos pontos	foat	mm
Thickness_mm	Espessura do perfil no ponto em questão	float	mm
Total_Area_mm2	Área total do perfil resumido , representado nesta tabela	float	mm
Left_Area_mm2	Área do lado esquerdo do perfil resumido , representado nesta tabela	float	mm
Right_Area_mm2	Área do lado direito do perfil resumido , representado nesta tabela	float	mm

Tabela 24 - Esquema da tabela *Nominal*

Coluna	Descrição	Tipo de Dados	UN.
ID	ID incrementável	int	-
Upload_Time	Instante de Armazenamento para cada linha	<i>Datetime</i>	-
Work_Center	Posto de trabalho	varchar(50)	-
Perfils_Dimension_ID	Tipo de perfil	varchar(50)	-
Cumulative_Width_mm	Indica, cumulativamente, a largura do perfil ao longo dos pontos	float	mm
Thickness_mm	Espessura do objetivo do perfil no ponto em questão	float	mm

Tabela 25 - Esquema da tabela *Section*

Coluna	Descrição	Tipo de Dados	UN.
ID	ID incrementável	int	-
Upload_Time	Instante de Armazenamento para cada linha	<i>Datetime</i>	-
Work_Center	Posto de trabalho	varchar(50)	-
Perfils_Dimension_ID	Tipo de perfil	varchar(50)	-
Date_Time	Instante captado pelo sensor		
Left_Width_mm	Indica a largura do perfil no lado esquerdo	float	mm
Right_Width_mm	Indica a largura do perfil no lado direito	float	mm

Tabela 26 - Esquema da tabela *Overlay*

Coluna	Descrição	Tipo de Dados	UN.
ID	ID incrementável	int	-
Upload_Time	Instante de Armazenamento para cada linha	<i>Datetime</i>	-
Work_Center	Posto de trabalho	varchar(50)	-
Production_ID	Chave criada pelo <i>developer</i> da script de modo a facilitar JOINS na base de dados	varchar(50)	-
Perfils_Dimension_ID	Tipo de perfil	varchar(50)	-
Beg_what_side	Indica a que lado pertence o ponto do perfil (o perfil em questão representa o início da corrida)	varchar(50)	-
Beg_Thickness_mm	Espessura do perfil no ponto em questão	float	mm
Mid_what_side	Indica a que lado pertence o ponto do perfil (o perfil em questão representa o meio da corrida)	varchar(50)	-
Mid_Thickness_mm	Espessura do perfil no ponto em questão	float	mm
End_what_side	Indica a que lado pertence o ponto do perfil (o perfil em questão representa o fim da corrida)	varchar(50)	-

End_Thickness_mm	Espessura do perfil no ponto em questão	float	mm
Nominal_Thickness_mm	Espessura desejada do perfil no ponto em questão	float	mm
Cumulative_Width_mm	Indica, cumulativamente, a largura perfil ao longo dos pontos	float	mm

Tabela 27 - Esquema da tabela *Assimetry_Differences*

Coluna	Descrição	Tipo de Dados	UN.
ID	ID incrementável	int	-
Upload_Time	Instante de Armazenamento para cada linha	<i>Datetime</i>	-
Work_Center	Posto de trabalho	varchar(50)	-
Production_ID	Chave criada pelo <i>developer</i> da script de modo a facilitar JOINS na base de dados	varchar(50)	-
Perfils_Dimension_ID	Tipo de perfil	varchar(50)	-
Date_Time	Instante captado pelo sensor	<i>Datetime</i>	-
is_Beginning_Mid_End	Indica se o instante representa o início, meio ou fim da corrida	varchar(50)	mm
Assimetry_Differences_mm	Indica a diferença da espessura para cada par de assimetria	float	-
assimetry_differences_Pvalue_Paired_T_Test	Indica o <i>P-value</i> do teste de assimetria <i>T-Test</i> realizado para o instante em questão	float	-
assimetry_differences_Pvalue_Paired_Wilcoxon	Indica o <i>P-value</i> do teste de assimetria <i>Wilcoxon</i> realizado para o instante em questão	float	-

Tabela 28 - Esquema da tabela *Assimetry_Differences_BME*

Coluna	Descrição	Tipo de Dados	UN.
ID	ID incrementável	int	-
Work_Center	Posto de trabalho	varchar(50)	-
Production_ID	Chave criada pelo <i>developer</i> da script de modo a facilitar JOINS na base de dados	varchar(50)	-
Perfis_Dimension_ID	Tipo de perfil	varchar(50)	-
Cumulative_Width_mm	Indica, cumulativamente, a largura perfil ao longo dos pontos	float	-
Beg_Assimetry_Differences_mm	Indica a diferença da espessura para cada par de assimetria, no perfil que representa a iteração inicial da corrida	float	mm
Mid_Assimetry_Differences_mm	Indica a diferença da espessura para cada par de assimetria, no perfil que representa a iteração do meio da corrida	float	mm
End_Assimetry_Differences_mm	Indica a diferença da espessura para cada par de assimetria, no perfil que representa a iteração final da corrida	float	mm

APÊNDICE B – RELATÓRIOS EM POWER BI



WELCOME TO THE PROFILOMETER EXE07 - LAST 15 DAYS BI DASHBOARD!

DESCRIPTION

This BI Dashboard was designed to facilitate data-driven decision-making for the Quality Department at Continental Mabor. It is a component of the larger Business Analysis Project, which aims to create methodologies and frameworks that enable various Continental Mabor departments to perform their duties using data-driven insights. This Power BI dashboard presents only the most critical Key Performance Indicators (KPIs) to prevent overwhelming the quality engineer with too much information. Instead, it ensures that the most important metrics are monitored, enabling quick and easy analysis of critical data to support both daily operational and strategic decisions.

This dashboard differs from the Profilometer - BME BI Dashboard in that it analyzes data only from a 15-day spectrum. However, it enables the user to examine any of the machine's run iterations.

Walkthrough - Pages:

Thickness - This page focuses on the Perfil's Thickness, several metrics are applied in order to compare the real Perfil with the target one.

Area - This page focuses on the Perfil's Area, several metrics are applied in order to compare the real Perfil with the target one.

Assimetry - This page focuses on the Perfil's Assimetry levels, several metrics are applied in order to compare the real Perfil with the target one.

CHANGE LOG

27.04.2023 - templated used to create the Profilometer Dashboard

14.05.2020 - template created

Author: Gonçalo Pereira

Reuestor: Ricardo Ferrão

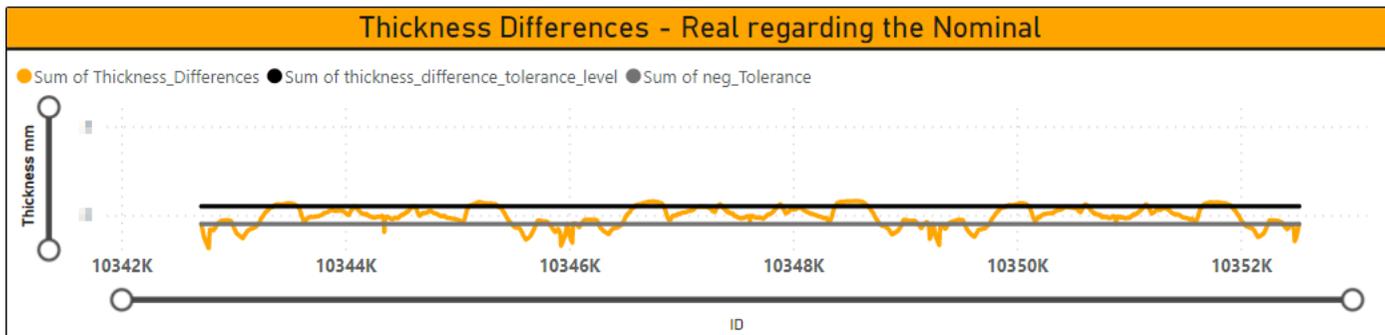
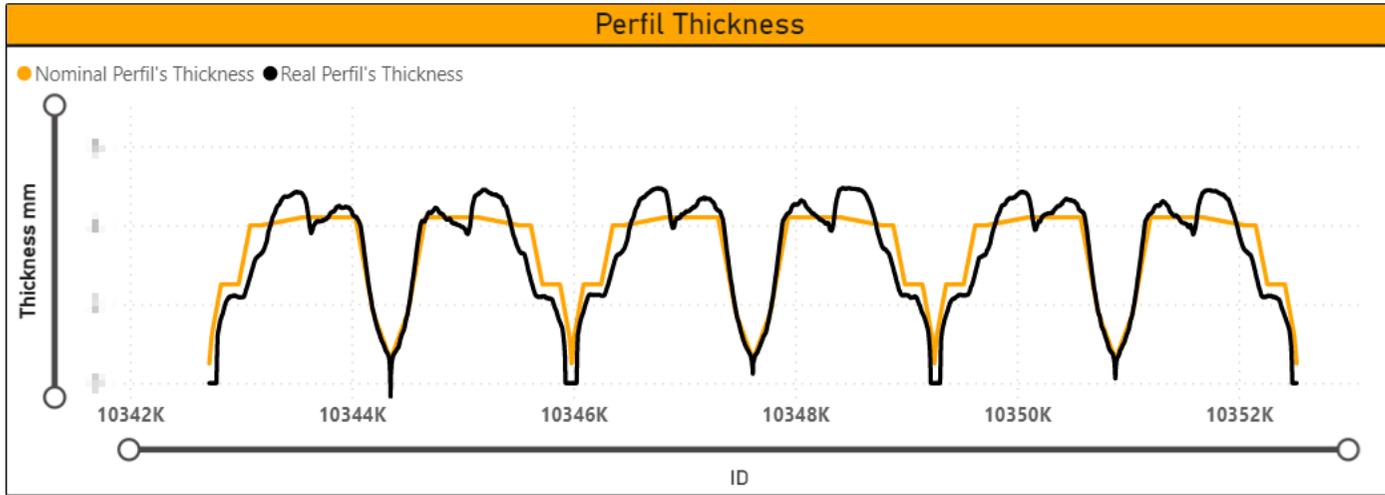
Figura 47 – Introdução, *Last 15 Days*

Perfil's Type Selected

Perfil's Side Selected

Y_Axis_Graph1

Y_Axis_Graph2



Average Difference (mm)



Standard Deviation (mm)



Maximum Difference (mm)



Dots Above Tolerance Level



Figura 48 - Espessura parte 1, Last 15 Days



Figura 49 - Espessura parte 2, Last 15 Days

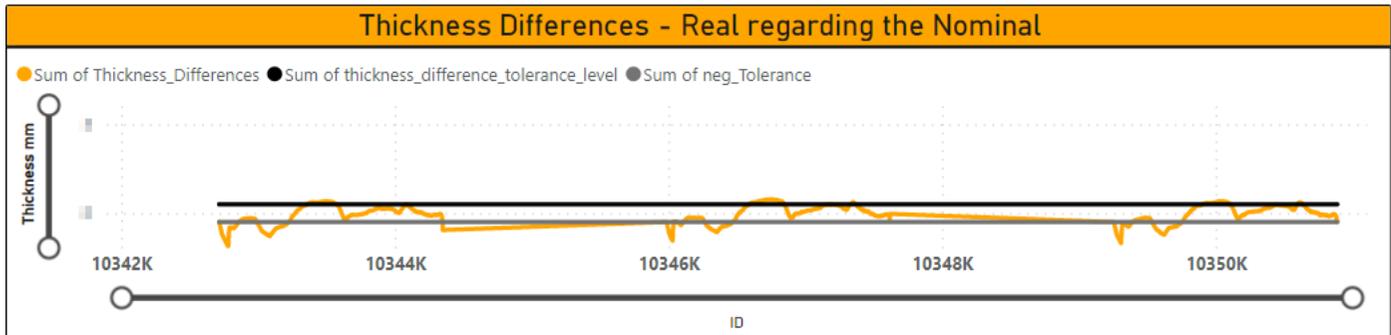
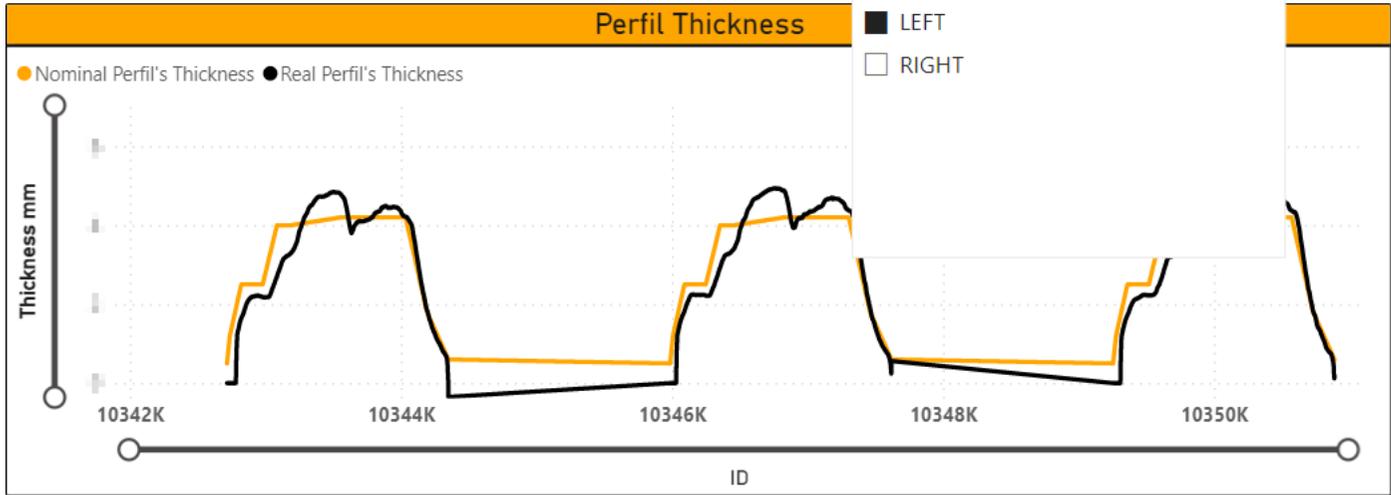
Perfil's Type Selected
 Multiple selections v

Perfil's Side Selected
 Multiple selections ^

Y_Axis_Graph1
T

Y_Axis_Graph2
+

- Select all
- LEFT
- RIGHT



Average Difference (mm)



Standard Deviation (mm)



Maximum Difference (mm)



Dots Above Tolerance Level

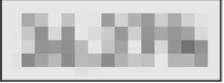


Figura 50 - Espessura parte 3, Last 15 Days

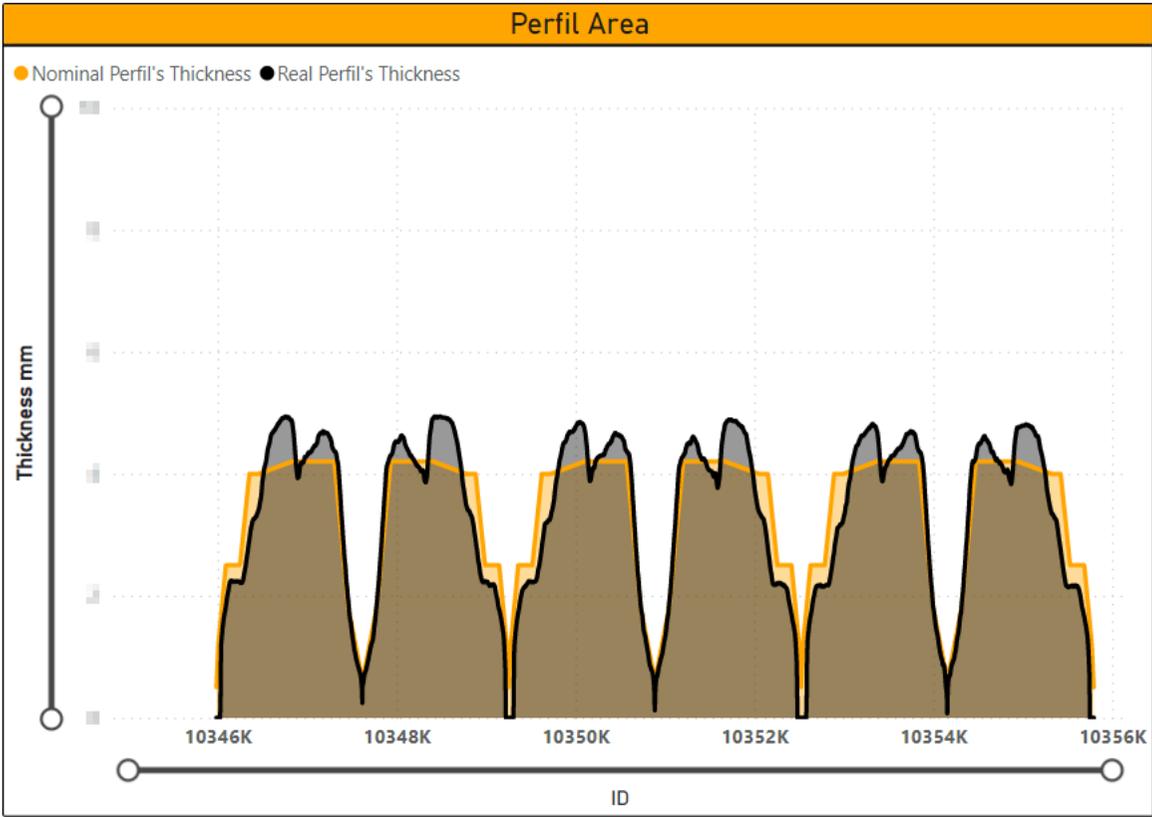
REAL
NOMINAL
R and N

FULL PERFIL
LEFT SIDE
RIGHT SIDE

Y_Axis_zoom

10

Left Area (mm ²)	Right Area (mm ²)
Nominal Left Area (mm ²)	Nominal Right Area (mm ²)
Left Real Width (mm)	Right Real Width (mm)
Left Nominal Width (mm)	Right Nominal Width (mm)



Area Difference Real vs Nominal (mm ²)
Area Difference Real vs Nominal (%)
Difference Width Real vs Nominal (mm)
Difference Width Real vs Nominal (%)

Figura 51 - Área, Last 15 Days

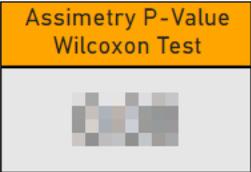
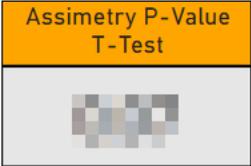
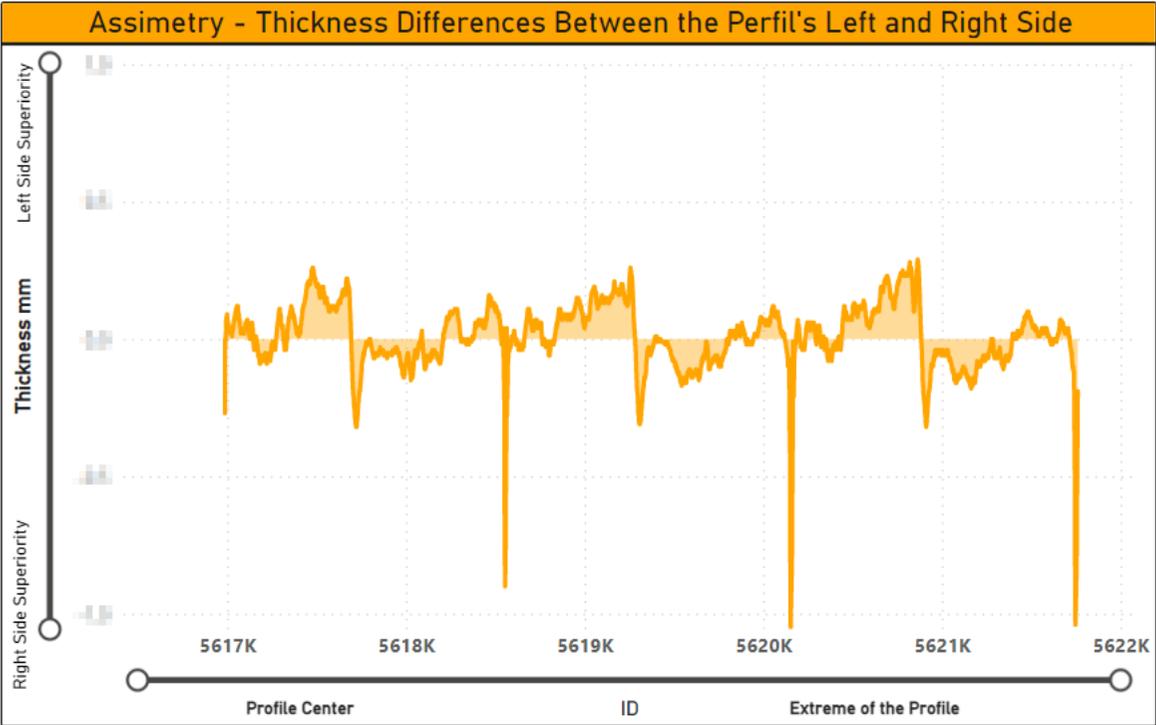
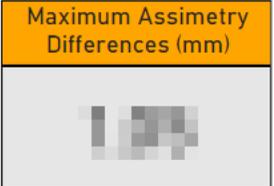
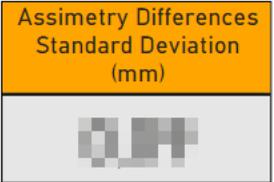
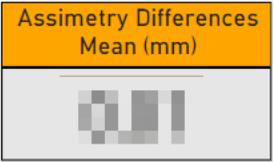
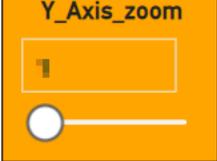
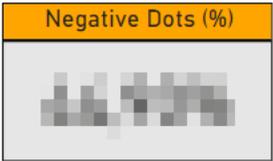
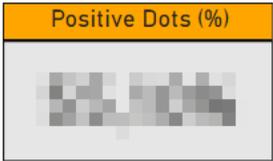


Figura 52 - Assimetria, Last 15 Days



WELCOME TO THE PROFILOMETER EXE07 - BME BI DASHBOARD!

DESCRIPTION

This BI Dashboard was designed to facilitate data-driven decision-making for the Quality Department at Continental Mabor. It is a component of the larger Business Analysis Project, which aims to create methodologies and frameworks that enable various Continental Mabor departments to perform their duties using data-driven insights. This Power BI dashboard presents only the most critical Key Performance Indicators (KPIs) to prevent overwhelming the quality engineer with too much information. Instead, it ensures that the most important metrics are monitored, enabling quick and easy analysis of critical data to support both daily operational and strategic decisions.

This dashboard differs from the Perfilometer - last 15 days BI Dashboard in that it analyzes data from the whole historical spectrum. However, it only enables the user to examine any three of the machine's run iterations: The first one, the middle one and the last one.

Walkthrough - Pages:

Thickness - This page focuses on the Perfil's Thickness, several metrics are applied in order to compare the real Perfil with the target one.

Area - This page focuses on the Perfil's Area, several metrics are applied in order to compare the real Perfil with the target one.

Assimetry - This page focuses on the Perfil's Assimetry levels, several metrics are applied in order to compare the real Perfil with the target one.

CHANGE LOG

27.04.2023 - templated used to create the Perfilometer Dashboard

14.05.2020 - template created

Author: Gonçalo Pereira

Created: 27.04.2023

Requestor: Ricardo Ferrão

Updated: 02.05.2023

Figura 53 - Introdução, BME

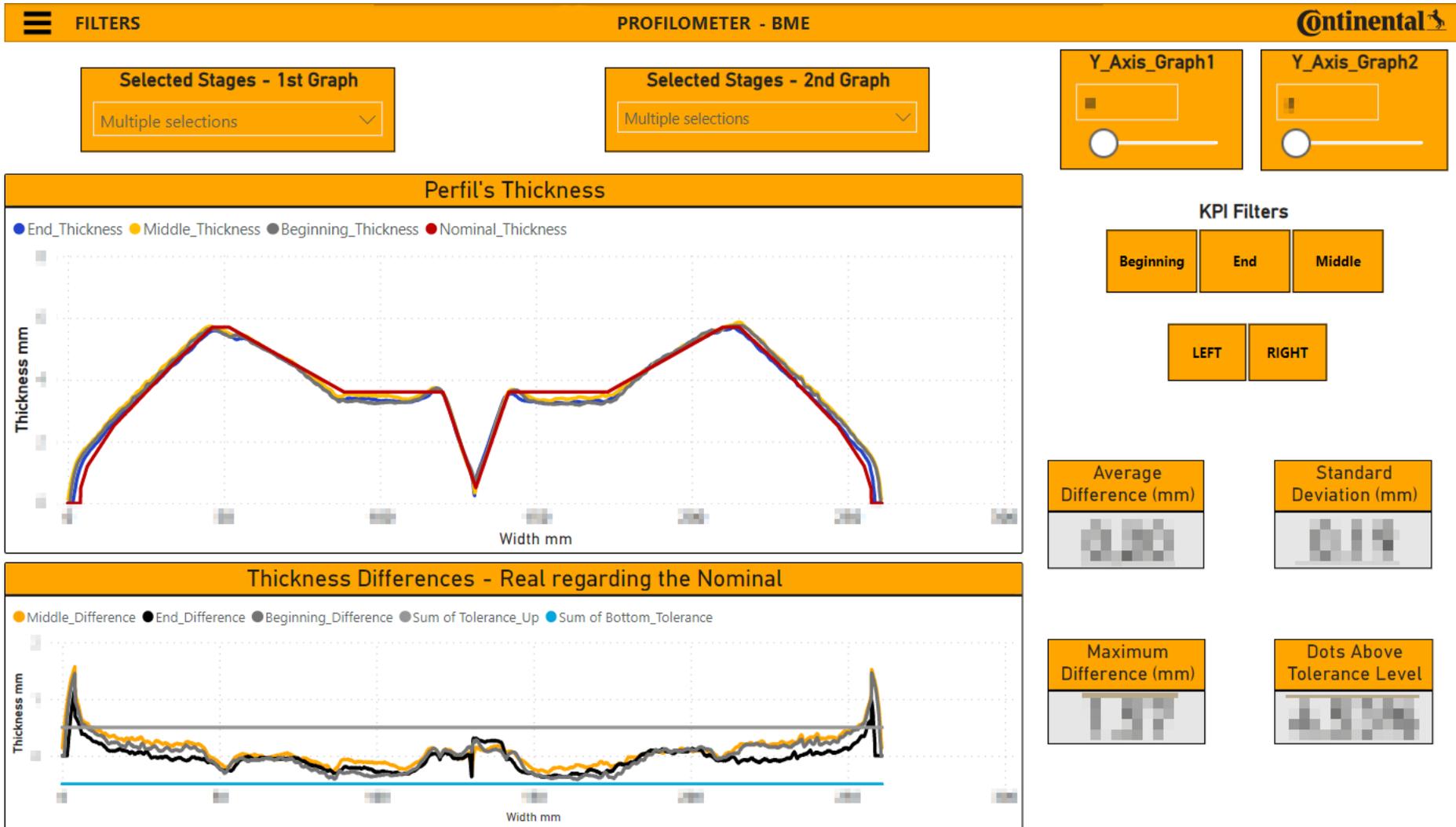


Figura 54 – Espessura parte 1, BME



Figura 55 - Espessura parte 2, BME

KPI Filters

Beginning

Middle

End

Selected Stages - Graph 1

Multiple selections ▼

Y_Axis_zoom

T

Left Area (mm²)

Right Area (mm²)

Nominal Left Area (mm²)

Nominal Right Area (mm²)

Left Real Width (mm)

Right Real Width (mm)

Left Nominal Width (mm)

Right Nominal Width (mm)

Perfil Area

● End_Thickness
 ● Middle_Thickness
 ● Beginning_Thickness
 ● Nominal_Thickness

Area Difference (mm²)

Area Difference (%)

Difference Width (mm)

Difference Width (%)

Figura 56 - Área, BME

KPI Filters

Beginning Middle

End

Selected Stages

Multiple selections

Positive Dots (%)

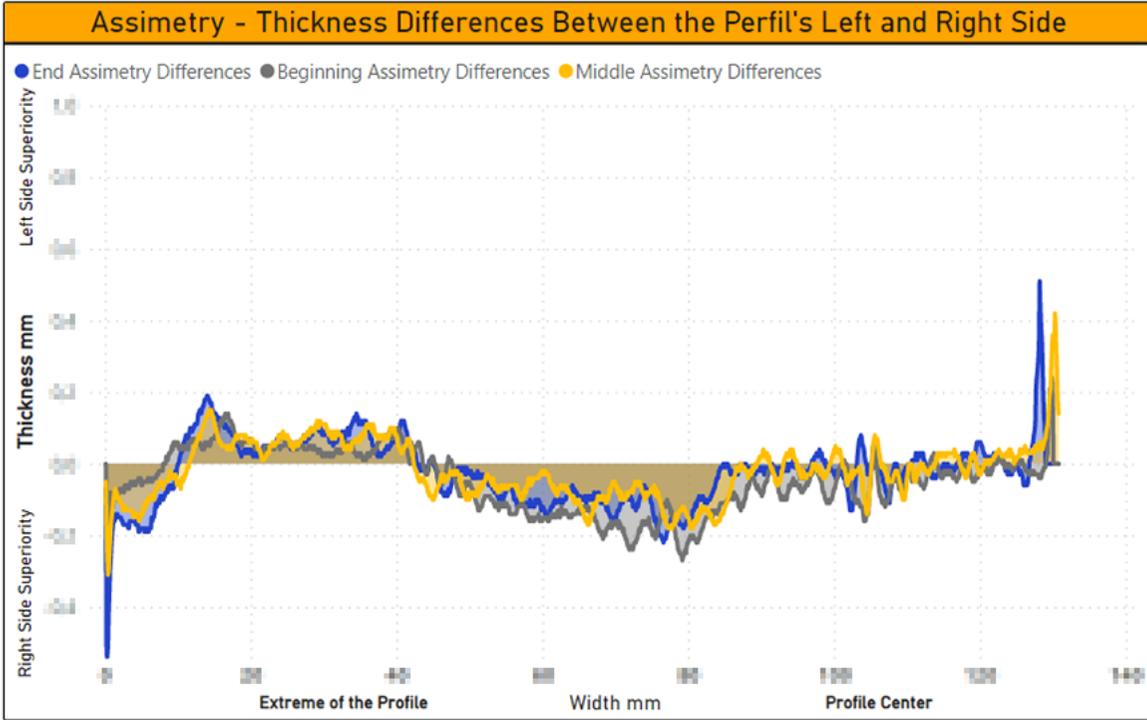
Negative Dots (%)

Y_Axis_zoom

Assimetry Differences Mean (mm)

Assimetry Differences Standard Deviation (mm)

Maximum Assimetry Differences (mm)



Assimetry P-Value T-Test

Assimetry P-Value Wilcoxon Test

Figura 57 - Assimtria, BME