



Bruno Do Monte Costa Rêga

Licenciado em Ciências da
Engenharia Electrotécnica e de Computadores

**Desenvolvimento de uma plataforma de análise
de dados para monitorização de inventário de
processos intra-logísticos na indústria
automóvel**

Dissertação para obtenção do Grau de Mestre em
Engenharia Electrotécnica e de Computadores

Orientador: Ricardo Rosa Jardim Gonçalves, Professor Cate-
drático, FCT-UNL
Co-orientadores: Ruben Duarte Dias da Costa, Investigador, FCT-
UNL
Paulo Alves Figueiras, Investigador, FCT-UNL

Júri

Presidente: Rui Miguel Henriques Dias Morgado Dinis
Arguente: Carlos Eduardo Dias Coutinho
Vogal: Ruben Duarte Dias da Costa



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro, 2021

Desenvolvimento de uma plataforma de análise de dados para monitorização de inventário de processos intra-logísticos na indústria automóvel

Copyright © Bruno Do Monte Costa Rêga, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade NOVA de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

AGRADECIMENTOS

Gostava de agradecer ao meu orientador, o Professor Ricardo Gonçalves, pela oportunidade de trabalhar neste projeto. Quero também agradecer aos meus co-orientadores, Ruben Costa e Paulo Figueiras, pelo apoio e feedback que me deram ao longo do desenvolvimento desta dissertação.

Gostaria também de agradecer à Universidade Nova de Lisboa, à Faculdade de Ciências e Tecnologias e a todos os professores e monitores que me ajudaram neste trajeto, que transmitiram os seus conhecimentos e nos inspiraram a aprender e a melhorar todos os dias.

Muito obrigado aos meus colegas e amigos que me acompanharam nestes anos, que me ajudaram a completar o meu trajeto académico, e com quem passei grandes momentos. Estes são amigos que levo para a vida.

Por fim, agradecer aos meus pais, António e Emília, às minhas irmãs, Daniela e Raquel, e ao meu cunhado, Bruno, pelo apoio incondicional e pela motivação para continuar neste caminho, mesmo quando as coisas não me corriam como queria.

RESUMO

A indústria 4.0 colocou em movimento uma mudança de paradigma na produção industrial, introduzindo a digitalização e modularização na fábrica, tecnologias de Internet of Things (IoT) e novas tecnologias no campo dos *smart objects* aplicados a máquinas e produtos.

Neste contexto, é abordado, no âmbito desta dissertação, o desenvolvimento de uma plataforma de visualização de dados que permita visualizar de forma simples e intuitiva, os *key performance indicators* (KPIs) dos processos na fábrica e que nos dê *insights* sobre os processos de forma a otimizar constantemente os mesmos. Como tal, foi feito um estudo do estado de arte sobre os tipos de visualização e interação de dados e as plataformas existentes para criação de *dashboards*, de forma a criar uma boa plataforma de visualização de dados.

Esta dissertação insere-se no âmbito do projeto europeu de investigação Boost 4.0, criado com o objetivo de aproveitar as mais recentes tecnologias da indústria 4.0 e implementá-las num ambiente industrial, de forma a reduzir o trabalho manual, susceptível a erros devido à natureza cíclica e prolongada dos processos, e substituir este modelo por um modelo modular mais flexível e automatizado, por forma a minimizar o erro e otimizar os processos.

Dentro do projeto Boost 4.0, esta dissertação aborda o desenvolvimento de plataformas de visualização de dados para dois casos de estudo distintos, sendo eles o da Volkswagen Autoeuropa, no contexto da indústria automóvel, e o caso da Ria Stone, no contexto da indústria cerâmica.

Palavras-chave: Indústria 4.0, dados, visualização de dados, *dashboard*, KPIs

ABSTRACT

Industry 4.0 put in motion a change in paradigm of industrial production, introducing the digitalization and modularization of the factory, Internet of Things (IoT) technologies and new technologies in the field of smart objects applied to machines and products.

In this context, this thesis addresses the development of a data visualization platform that allows the visualization, in a simple and intuitive manner, of the factory processes' key performance indicators (KPIs), and provides insights of the processes, in order to constantly optimize them. We studied the state of the art of the types of data visualization and interaction, as well as the existing dashboard creation platforms, in order to develop a good data visualization platform.

This thesis is part of the European project Boost 4.0, developed with the goal of implementing the latest Industry 4.0 technologies in an industrial environment, in order to reduce manual labor, susceptible to errors due to the cyclic nature and long up times of the processes, and replace the current model with a modular, more flexible and automated one, so as to minimize errors and optimize the processes.

From the Boost 4.0 project, this thesis addresses the development of data visualization platforms for two distinct case studies, Volkswagen Autoeuropa, in the context of the automotive industry, and Ria Stone, in the context of ceramics industry.

Keywords: Industry 4.0, data, data visualization, dashboard, KPIs

ÍNDICE

Lista de Figuras	xiii
Lista de Tabelas	xvii
1 Introdução	1
1.1 Indústria 4.0	1
1.1.1 Conceitos Fundamentais	2
1.2 Caso de Estudo	3
1.3 Descrição do Problema	6
1.4 Solução Proposta	7
1.5 Pergunta de Investigação e Hipótese	7
1.6 Estrutura do Documento	8
2 Estado de Arte	11
2.1 Visualização	11
2.1.1 O Processo de Visualização	11
2.1.2 Porquê Interatividade?	12
2.2 Cor na Visualização	13
2.2.1 Modelos de Cor	13
2.2.2 Propriedades das Escalas de Cor	13
2.2.3 Componentes de Modelos de Cor	14
2.2.4 Escalas de Cor Redundantes	15
2.2.5 Escalas Divergentes	15
2.2.6 Características de Visualização e Seleção de Escalas de Cor	15
2.3 Técnicas de Visualização	22
2.3.1 Tipos de Datasets	23
2.3.2 Técnicas de Visualização	25
2.3.3 Técnicas de Interação	37
2.4 Dashboard	39
2.4.1 Microsoft Power BI	39
2.4.2 ThoughtSpot	41
2.4.3 Tableau	42
2.4.4 Qlik	42

2.4.5	Kibana	44
2.4.6	Grafana	44
2.5	Benchmarks	45
3	Desenvolvimento da plataforma	47
3.1	Definição da Plataforma	47
3.2	Arquitetura Conceptual	48
3.3	Caso de Estudo Volkswagen Autoeuropa	49
3.3.1	Análise de Dados e Seleção das Métricas	49
3.3.2	Criação do dashboard e visualizações	50
3.3.3	Implementação na plataforma Web	61
3.4	Caso de Estudo Ria Stone	63
3.4.1	Análise dos Dados e Seleção das Métricas	63
3.4.2	Criação das visualizações	64
4	Discussão de Resultados	77
4.1	Caso de Estudo Volkswagen Autoeuropa	77
4.2	Caso de Estudo Ria Stone	79
5	Conclusão e Trabalho Futuro	83
5.1	Conclusões	83
5.2	Trabalho Futuro	85
	Bibliografia	87

LISTA DE FIGURAS

1.1	Processo atual de recepção, armazenamento, transporte para o <i>point of fit</i> e sequenciação [2, 4].	4
1.2	Processo de recepção, armazenamento, transporte para o <i>point of fit</i> e sequenciação com integração de tecnologias de indústria 4.0 [2, 4].	5
2.1	Escalas de Cor. Da esquerda para a direita e de cima para baixo: <i>gray scale</i> , <i>rainbow scale</i> , escala <i>heated-object</i> e escala <i>linearized optimal</i> aplicadas a um <i>dataset</i>	16
2.2	Escala de cor divergente aplicada a um mapa da intensidade do campo magnético da terra [14].	16
2.3	Escalas de cor qualitativas disponíveis no site ColorBrewer 2.0 [15].	17
2.4	Escalas de cor sequenciais disponíveis pelo website ColorBrewer 2.0 [15].	17
2.5	Escalas de cor divergentes disponibilizadas pelo website ColorBrewer 2.0 [15].	18
2.6	Ilustrações do campo magnético superficial da Terra (em cima) e da observação de frações de nuvens (em baixo), para cinco níveis (à esquerda) e 10 níveis (à direita). Imagem retirada de [18].	19
2.7	Efeito da cor de fundo na percepção de tamanho e brilho de objetos. Imagem retirada de [21].	20
2.8	Efeito da cor de fundo na percepção da cor de um objecto. À esquerda, dois objetos com a mesma cor em fundos de cores diferentes parecem ter cores diferentes. À direita, dois objetos com cores diferentes em fundos de cores diferentes parecem ter a mesma cor. Imagem retirada de [21].	21
2.9	Mapa do metro de Londres. Cada linha possui uma cor única que a identifica facilmente das restantes linhas. Imagem retirada de [22].	22
2.10	Skitter Graph de 2017 ilustra o exemplo de um grafo.	25
2.11	Exemplo de tabela, retirado de [6].	26
2.12	Exemplo de um <i>pie chart</i> , retirado de [28].	27
2.13	Exemplo de um <i>doghnut chart</i> , de [28].	27
2.14	Exemplo de <i>multi level pie chart</i> , de [28]. Os segmentos são divididos em dois sub segmentos, para homem e mulher, de modo a poder comparar dentro de cada segmento a proporção de indivíduos de cada sexo.	28
2.15	Exemplo de um gráfico de barras simples, de [28].	29

2.16	Exemplo de um <i>floating column chart</i> , de [28].	29
2.17	Exemplo de histograma, de [6].	30
2.18	Exemplo de um scatter plot, de [35].	31
2.19	Exemplo de um gráfico de linha, de [28].	31
2.20	Exemplo de um gráfico de área, de [28].	32
2.21	Exemplo de um <i>bubble chart</i> , de [28].	33
2.22	Exemplo de um <i>multiple data series plot</i> , de [28].	33
2.23	Exemplo da técnica de coordenadas paralelas, retirada de [23].	34
2.24	Exemplo de um display de ícones, retirado de [23].	35
2.25	Exemplo de um <i>dense pixel display</i> com técnica de padrões recursivos, de [41].	35
2.26	Exemplo de um <i>dense pixel display</i> com técnica de segmentos de círculo, de [42].	36
2.27	Exemplo de um <i>Stacked display</i> , retirado de [23].	36
2.28	Exemplo de uma distorção de fisheye lens aplicada a uma matriz de <i>scatter plots</i> , retirada de [23].	38
2.29	<i>Gartner Magic Chart</i> para plataformas de <i>Business Intelligence</i>	40
3.1	Arquitetura conceptual da plataforma de visualização interativa de dados. .	48
3.2	Diagrama de entidades e relações da base de dados da Volkswagen Autoeuropa.	51
3.3	Tabela de <i>hourly_stock</i> , em formato CSV. Cada coluna representa o tipo de bateria utilizada, enquanto que as linhas representam o número de paletes para cada tipo de bateria, para a data e hora dada na coluna <i>start_datetime</i> . .	52
3.4	Número de paletes de baterias encomendadas, dividido por número de peça de cada tipo de bateria.	52
3.5	Número total de encomendas de paletes de baterias em percentagem para cada tipo de bateria utilizada na fábrica.	53
3.6	Gráfico do número de baterias encomendadas a cada fornecedor, ao longo da janela de tempo dos dados.	53
3.7	Gráficos para o stock atual de contentores de baterias, para cada tipo de bateria.	54
3.8	Gráfico do estado do stock de cada tipo de bateria. Este gráfico, no qual cada linha representa o stock de uma peça ao longo do tempo, utiliza um código de cor para identificar o número de paletes em stock para cada tipo de peça. . .	55
3.9	Gráfico do stock de paletes de baterias, para cada tipo de bateria.	55
3.10	Gráfico do <i>stock</i> diário de contentores em armazém com o número de peça "5TA915105B", da família das baterias.	56
3.11	Gráfico do <i>stock</i> diário de baterias, com thresholds de utilização média (vermelho) e margem de segurança (amarelo) para a peça "5TA915105B", da família das baterias.	57
3.12	Gráfico do stock diário e encomendas de baterias para a peça "5TA915105B"	57
3.13	Gráficos para as operações de transporte de peças do armazém para outros pontos da fábrica.	58

3.14	Número contentores de baterias sequenciadas, por tipo de bateria, ao longo da janela de tempo dos dados.	59
3.15	Tempos de sequenciação por turno, com os valores da sequenciação correta a verde e outliers a vermelho.	60
3.16	Tempos de sequenciação, com os valores da sequenciação correta a verde e com os valores de uma sequenciação errada a vermelho, e um <i>threshold</i> , marcado com uma linha horizontal a vermelho, que separa estas sequenciações erradas dos restantes valores.	60
3.17	Tempos de sequenciação por turno, apenas com os valores da sequenciação seguindo o protocolo correto.	60
3.18	Caixas de texto criadas para a seleção das peças a visualizar nos gráficos do <i>dashboard</i>	61
3.19	Página de login da plataforma de visualização do projeto Boost 4.0.	62
3.20	Página inicial da plataforma de visualização do projeto Boost 4.0, após login.	62
3.21	<i>Dashboard</i> do processo de <i>Receiving</i> na plataforma de visualização do projeto Boost 4.0.	62
3.22	<i>Dashboard</i> do processo de <i>Warehousing</i> na plataforma de visualização do projeto Boost 4.0.	63
3.23	Tabela do processo de avaliação de qualidade.	65
3.24	<i>Pie chart</i> que representa a qualidade dos produtos, em percentagem, por turno. A verde estão representados os produtos de qualidade mais alta, a amarelo os produtos de qualidade mais baixa que foram aprovados, e a azul os produtos que foram rejeitados.	66
3.25	Gráfico de barras que ilustra a qualidade média dos produtos por turno, no qual T1 representa o primeiro turno, T2 o segundo turno e T3 o terceiro turno de produção.	66
3.26	Gráfico de barras que ilustra o número total de pratos produzidos por turno, no qual T1 representa o primeiro turno, T2 o segundo turno e T3 o terceiro turno de produção.	67
3.27	Gráfico de barras que ilustra a qualidade média dos produtos por mesa de avaliação de qualidade, no qual os números de 1 a 8 correspondem ao número de cada mesa.	67
3.28	Gráfico de barras que ilustra a qualidade média dos produtos por cada cor de produto. Da esquerda para a direita temos: Bege, Branco, Cinza Escuro, Rosa, Azul Petróleo, Verde Água e Cinza	68
3.29	Gráfico de barras que ilustra a qualidade média dos produtos por tipo de produto. Os tipos de pratos representados no gráfico são, da esquerda para a direita: Prato DIN 26, Prato DIN 30x20, Prato FLI 26, Prato de sobremesa DIN 20, Prato de sobremesa FLI 20, Tigela DIN 14, Prato FAR 27, Prato de sobremesa FAR 21, Prato de sopa DIN 22, Tigela FAR 16, Prato de sopa FAR 24, Tigela FLI 14 e Prato de sopa FLI 22.	68

3.30	Gráfico de linha que representa visualmente a velocidade média da linha de produção para os pratos produzidos ao longo de um mês.	69
3.31	Gráfico de linha que representa visualmente a velocidade média do torno para os pratos produzidos ao longo de um mês.	69
3.32	Gráfico de linha que representa visualmente a viscosidade média do vidro aplicado no processo de <i>glazing</i> , ao longo de um mês. A verde temos a viscosidade do vidro aplicado nos pratos, e a amarelo a viscosidade do vidro dentro do reservatório.	70
3.33	Gráfico de linha que representa visualmente a densidade média do vidro utilizado no processo de <i>glazing</i> , ao longo de um mês.	70
3.34	Gráfico de linha que representa visualmente a temperatura média do vidro utilizado no processo de <i>glazing</i> , ao longo de um mês. A verde temos a temperatura atual para cada prato vidrado no processo, enquanto que a amarelo temos a temperatura média do vidro utilizado no processo.	70
3.35	Densidade do líquido de <i>glazing</i> por cada turno de produção.	71
3.36	Temperatura do líquido de <i>glazing</i> por cada turno de produção.	71
3.37	Densidade do líquido de <i>glazing</i> por hora do dia.	71
3.38	Temperatura do líquido de <i>glazing</i> por hora do dia.	72
3.39	Visualização gráfica da matriz de correlação dos parâmetros do processo de <i>glazing</i> com a percentagem de qualidade.	72
3.40	Visualização gráfica da matriz de correlação dos parâmetros do processo de <i>glazing</i> com a percentagem de qualidade, para os dados anteriores ao dia 19.	74
3.41	Gráfico da correlação entre os valores de temperatura do vidro (a verde) e os valores de percentagem de qualidade (a azul).	75
3.42	Gráfico da correlação entre os valores da densidade do vidro (a verde) e os valores de percentagem de qualidade (a roxo).	75

LISTA DE TABELAS

2.1	Tabela dos requisitos necessários que apresenta quais os requisitos que cada solução cumpre.	45
3.1	Análise de correlação dos parâmetros do processo de <i>glazing</i> com o parâmetro de percentagem de qualidade dos pratos, do processo de avaliação de qualidade.	73
3.2	Comparação das análises de correlação dos parâmetros do processo de <i>glazing</i> com o parâmetro de percentagem de qualidade dos pratos, do processo de avaliação de qualidade, para o mês completo e para os dados entre o dia 1 e o dia 19.	74

INTRODUÇÃO

1.1 Indústria 4.0

Desde o início da industrialização que os vários avanços nas tecnologias deram origem ao que chamamos de "revoluções industriais". A primeira revolução industrial deu-se devido aos avanços na área da mecanização, a segunda revolução industrial ocorreu devido à utilização de energia elétrica, e a terceira revolução ocorreu com os avanços da digitalização. Com a evolução da digitalização nas fabricas, combinado com as tecnologias de Internet of Things (IoT) e com o aparecimento de futuras tecnologias no campo dos *"smart"objects* aplicadas a máquinas e produtos, está em movimento uma nova mudança de paradigma na produção industrial, a quarta revolução industrial ou, como é mais conhecida, "Indústria 4.0". A Indústria 4.0 baseia-se na digitalização da fábrica, modularização sistemas de manufatura, e na ideia de que os produtos controlam o processo de manufatura, com o objetivo de criar produtos personalizados com base na preferência de cada cliente, sem no entanto abdicar dos princípios da produção em massa [1].

Segundo Lasi et al. [1], existem dois fatores que motivaram o desenvolvimento da Indústria 4.0, sendo o primeiro a necessidade de alteração das condições de operação. Em particular:

- Curtos períodos de desenvolvimento: A necessidade de redução do tempo de desenvolvimento e inovação, que garantem o sucesso das empresas.
- Individualização *on demand*: A mudança do paradigma do mercado, de *seller driven* para *buyer driven*, que permite a personalização dos produtos por parte dos compradores.
- Descentralização: Devido à necessidade de uma tomada de decisão mais rápida, é necessário a redução de hierarquias organizacionais.

- **Eficiência de Recursos:** Devido ao aumento da escassez de recursos e, como tal, ao aumento do preço dos mesmos, em conjunto com o contexto ecológico atual, é necessário garantir a sustentabilidade no contexto da indústria, e aumentar a eficiência.

O outro fator que motiva o surgimento da Indústria 4.0 é existência de novas tecnologias que foram lançadas para o mercado industrial. Estas novas tecnologias tinham já entrado no mercado privado mas não eram, no entanto, ainda muito utilizadas no mercado industrial [1]. Estas tecnologias permitem:

- **Aumento da mecânização e automação:** No processo industrial, existem mais apoios ao trabalho físico. Soluções automáticas vão realizar a execução de operações versáteis, através de células de manufatura autónomas que controlam e otimizam o processo de manufatura.
- **Digitalização e *networking*:** Com o aumento da digitalização da manufatura e das ferramentas de manufatura, existe um aumento da quantidade de dados sensoriais e de atuação que permitem o suporte de funções de análise e controlo. O *networking* de componentes permite o desenvolvimento de processos digitais, que em conjunto com a digitalização dos produtos e dos serviços, permitem a completa digitalização da indústria.
- **Miniaturização:** Computadores e dispositivos com grande poder de processamento ocupam poucos centímetros cúbicos, o que permite a sua utilização no contexto da logística e da produção industrial.

1.1.1 Conceitos Fundamentais

O termo Indústria 4.0 refere-se a um conjunto de conceitos cuja extensão e compreensão é de difícil definição [1].

Portanto, apenas é possível listar os conceitos fundamentais, que são os seguintes:

- ***Smart Factory*:** Os equipamentos de manufatura estão equipados com sensores, atuadores e sistemas autónomos. Utiliza modelos digitais de produtos e fabricas e aplica tecnologias de *Ubiquitous Computing* para criar uma *smart factory* controlada autonomamente.
- **Sistemas ciber-físicos:** Os espaços físico e digital fundem-se num só, de forma que os novos sistemas possuem representações digitais que não podem ser diferenciados do espaço físico.
- **Auto organização:** Os sistemas de manufatura estão cada vez mais descentralizados, com o abandono da hierarquia clássica da indústria a favor de uma auto organização de cada modulo independente dentro da fábrica.

- Novos sistemas na distribuição e aquisição: A distribuição e aquisição serão cada vez mais individualizados.
- Novos sistemas no desenvolvimento de produtos e serviços: Os produtos e serviços serão cada vez mais individualizados.
- Adaptação às necessidades humanas: Novos sistemas de manufatura adaptados às necessidades humanas, em vez do oposto.
- Responsabilidade social corporativa: Aumento do foco na sustentabilidade e eficiência de recursos.

1.2 Caso de Estudo

A Volkswagen Autoeuropa, pertencente ao grupo Volkswagen, é uma fábrica automóvel localizada em Palmela, Portugal, fundada em 1995, que emprega atualmente cerca de 6000 funcionários na fábrica, e possui cerca de 800 fornecedores de material, peças e componentes [2]. Em 2019, a produção de automóveis chegou às 254,600 mil unidades [3]. No entanto, apesar destes valores, a linha de produção não está ainda otimizada, requerendo ainda bastante intervenção humana, que produz atrasos significativos na produção e diminui a produção.

É aqui que surge o consorcio Boost 4.0, que tem como objetivo a implementação das tecnologias emergentes da indústria 4.0 de modo a otimizar a produção da fábrica, reduzir a dependência de mão-de-obra manual e aumentar a automação e flexibilidade modular [4].

No processo logístico atual, representado na Figura 1.1, existem dados das várias fases do processo, no entanto, nem todas as fases, como é o caso dos transportes, possuem dados, e os dados que existem não estão completamente integrados. Na Figura 1.2, página 5, é possível ver o objetivo final da integração 4.0, no qual existe uma completa integração dos dados do processo, desde a receção de peças ao *point of fit*.

Os benefícios da implementação destas tecnologias traduzem-se em redução de gastos, aumento da flexibilidade da linha, melhoria das condições ergonómicas para os operadores, minimização da dependência humana para operações manuais, com consequência final o aumento da eficiência do processo logístico.

O objetivo do projeto é:

- Estabelecer a automação do armazém (*put away and picking*) e entrega automática à área de comissionamento (SUMA).
- Estabelecer um ambiente interativo de sequenciação de material para a linha de montagem. O sistema de sequenciação permite à Volkswagen Autoeuropa sequenciar as peças pela ordem de montagem, para enviar para o *point of fit*.

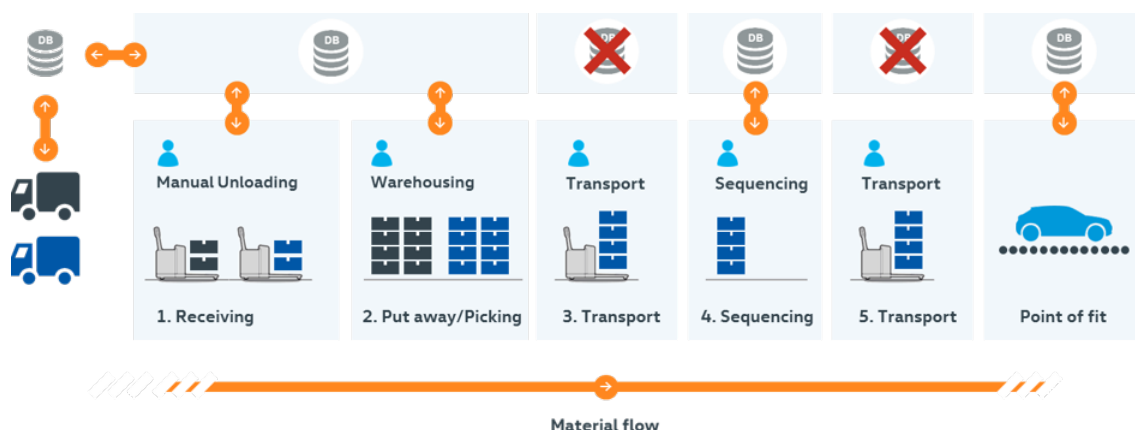


Figura 1.1: Processo atual de recepção, armazenamento, transporte para o *point of fit* e sequenciação [2, 4].

O projecto tem como objetivo permitir à Volkswagen Autoeuropa automatizar todo o processo, desde o armazém à linha de montagem, incluindo o processo de sequenciação e transporte das paletes para o *point of fit*.

O processo logístico atual pode ser descrito nos seguintes passos [2]:

- Recepção e Descarregamento manual (*Receiving and Unloading*): Os camiões chegam à baía de descarregamento com um ASN (*advance shipping notice*) do material, um operador faz o *scan* de cada unidade e cada uma é verificada de acordo com as normas da Volkswagen. O sistema de armazenamento atribui o destino de cada unidade, correspondente aos pontos de armazenamento corretos. Este processo é em grande parte manual.
- Armazenamento (*Warehousing*): As peças são armazenadas num armazém ou enviadas diretamente para a linha de produção, no caso de serem encomendas *just in time* (JIT).
- Transporte para sequenciação: Transporte das peça em carrinhos moveis, com recurso a reboques, desde o armazém até à estação de sequenciação.
- Sequenciação (*Sequencing*): Atualmente, a sequenciação é feita manualmente, apoiado por um sistema que ajuda o operador a colocar as peças, que serão montadas nos carros no *point of fit*, na sequência correta de acordo com a sequência dos modelos dos carros que estão na linha de montagem, de modo a que não seja necessário perder tempo a procurar a peça certa para cada carro, tornando o processo mais eficiente. Quando o contentor está cheio, com todas as partes sequenciadas, o operador passa à estação de sequenciação seguinte e repete o processo.
- Transporte para o *point of fit*: Os contentores com as peças sequenciadas são transportados para a linha de montagem, com recurso a AGVs (Automated Guided Vehicles) ou reboques.

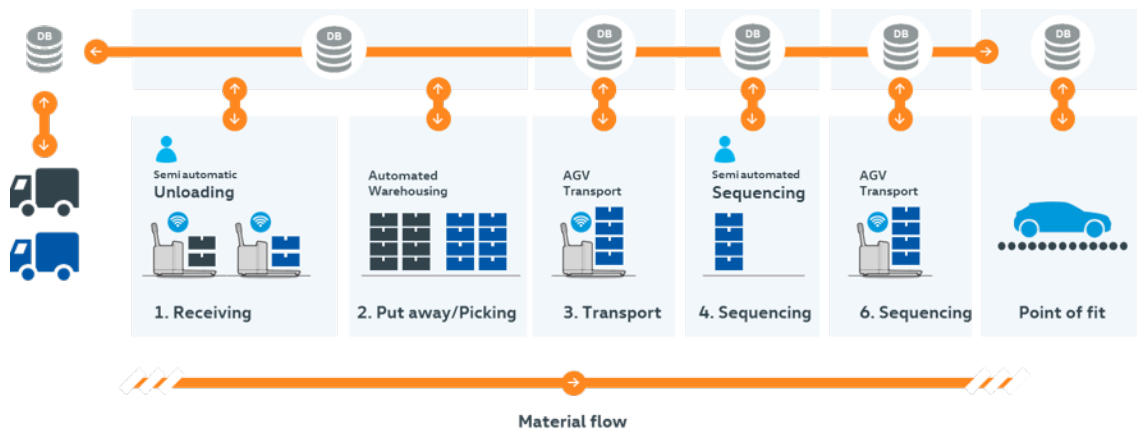


Figura 1.2: Processo de recepção, armazenamento, transporte para o *point of fit* e sequenciação com integração de tecnologias de indústria 4.0 [2, 4].

No contexto deste projecto, e visto que esta dissertação está inserida num extenso projecto de introdução de tecnologias de indústria 4.0 no processo de produção da Volkswagen Autoeuropa, e não seria possível abordar todas as fases deste projecto num espaço de tempo tão reduzido, o foco da mesma será a criação de uma plataforma para visualização de dados e KPIs de cada fase do processo logístico na fábrica da Volkswagen Autoeuropa.

Será abordado ainda um segundo caso de estudo, na empresa Ria Stone, um caso de estudo de replicação, de modo a testar a solução no contexto da indústria cerâmica, um contexto bastante distinto da indústria automóvel.

A Ria Stone é uma fábrica de da indústria cerâmica do grupo Visabeira, situada em Ílhavo, que produz anualmente cerca de 30 milhões de peças de louça que são comercializadas pela IKEA, empresa sueca de retalho, especializada na venda de artigos para a casa. A Ria Stone realizou uma expansão da sua fábrica e implementou um processo automatizado de produção de louça de mesa [5]. Neste contexto, a empresa necessita de uma plataforma de visualização de dados que permita monitorizar os parâmetros relevantes na linha de produção, de modo identificar a relação entre estes e os defeitos encontrados nas peças produzidas, assim como para diminuir estes defeitos e aumentar a qualidade das peças.

Com estes problemas em mente, o objetivo para este caso de estudo é aplicar os mesmos métodos utilizados para o caso de estudo da Volkswagen Autoeuropa, adaptados para o caso da Ria Stone, para os seus dados e objetivos, de modo a obter uma plataforma que permita a visualização dos dados de dois processos da fábrica, os processos de *glazing* (processo de vidragem dos pratos) e de avaliação de qualidade, e da correlação dos dados destes dois processos, de forma a tentar descobrir alguma correlação entre os seus parâmetros e os dados de qualidade.

1.3 Descrição do Problema

Temos dois casos de estudo diferentes, pelos que teremos também dois problemas distintos, para cada um dos casos de estudo.

No processo intra-logístico da indústria automóvel, mais especificamente do nosso caso de estudo, a Volkswagen Autoeuropa, existe a necessidade uma plataforma que permita a visualização, análise e monitorização dos dados de encomendas, armazenamento, transporte e utilização de peças, que permita construir uma imagem da operação da fábrica e identificar possíveis anomalias nos seus processos.

A Volkswagen Autoeuropa não possui nenhuma ferramenta robusta que permita a monitorização do processo intra-logístico. Sem esta plataforma, não é possível a visualização dos dados, que são essenciais na análise de padrões e tendências dos dados do processo intra-logístico, o que dificulta o processo de tomada de decisão, e reduz a eficiência de produção. Utilizando um exemplo do caso de estudo da Volkswagen Autoeuropa, sem uma plataforma de visualização de dados, não é possível monitorizar o volume do stock de baterias. Sem esta ferramenta, não é possível garantir a eficiência do stock, não sendo possível a encomendada de baterias em número suficiente para garantir o volume de stock desejado, sem encomendar em excesso.

Outro factor determinante para a necessidade de uma ferramenta como esta prende-se com o facto de que a informação da fábrica está distribuída, não concentrada num único local, e, devido à falta de integração dos dados, não existe neste momento nenhum suporte de tomada de decisão.

Para a demonstração do projecto numa primeira fase, a uma escala mais reduzida, será abordado o processo intra-logístico completo das baterias, dado o seu fácil armazenamento, transporte e instalação, e, visto que as baterias são instaladas em todos os carros, cobrem toda a operação da Volkswagen Autoeuropa.

No caso de estudo da Ria Stone, foram implementados recentemente sensores na linha de produção da fábrica, de forma a recolher dados da linha de produção e fazer monitorização e otimização da mesma. Como tal, é necessária uma plataforma de visualização de dados que permita a monitorização dos parâmetros da linha de produção, e a correlação destes dados com os dados de qualidade, de modo a identificar relações entre estes parâmetros e os defeitos dos produtos, e permitir a eliminação destes defeitos e a otimização da produção.

O grande problema que queremos abordar no âmbito desta tese é o desenvolvimento desta plataforma de visualização de dados. Outro problema, que não será abordado mas que está relacionado com o desenvolvimento da plataforma, nomeadamente na correlação dos dados da linha de produção com os dados de qualidade, é a falta de rastreabilidade no processo logístico da fábrica, que irá dificultar a correlação dos dados.

Mais que apenas uma plataforma de visualização, é necessário que esta plataforma seja simples, intuitiva e fácil de utilizar, para permitir a que qualquer pessoa com acesso à plataforma a consiga utilizar sem grandes dificuldades. É também necessário que a

plataforma permita a análise dos dados a qualquer altura que esta seja necessária.

1.4 Solução Proposta

Com este desafio em mente, o objectivo desta dissertação é o desenvolvimento uma plataforma *dashboard* que permita a visualização fácil e eficaz dos dados logísticos da Volkswagen Autoeuropa. Esta plataforma será uma página web, acessível pelo pessoal autorizado com recurso a credenciais de *login*, na qual serão integradas as visualizações criadas com recurso a uma das soluções existentes para criação dos *dashboards*. Esta plataforma permitirá também filtrar as visualizações para visualizar apenas os dados pretendidos. Os dados são recolhidos por sensores nas linhas de produção e registos de stock, que são depois armazenados na base de dados. É depois feita a gestão dos dados de modo a retirar os dados indesejados, como *missing values* e *outliers*, para garantir a qualidade dos dados analisados.

Como tal, foram definidos alguns passos de forma a desenvolver uma solução que satisfaça as necessidades apresentadas pelo problema. Será então necessário:

- Explorar as técnicas de visualização existentes;
- Estudar as soluções existentes para plataformas de criação de *dashboards*;
- Realizar uma análise dos dados do processo intra-logístico para perceber quais as técnicas de visualização mais adequadas ao tipo de dados em estudo;
- Implementar a plataforma de visualização de dados;
- Definir os serviços para alimentar a plataforma com os dados do processo intra-logístico;
- Definir e implementar os KPIs associados ao processo.

1.5 Pergunta de Investigação e Hipótese

Embora os problemas abordados na secção 1.3 não serão completamente resolvidos pela solução proposta, existem algumas questões importantes que se pretendem responder com esta dissertação. Estas são as perguntas de investigação, para as quais as respostas são dadas pelas hipóteses de investigação apresentadas abaixo.

Pergunta de Investigação: Como é que um sistema de visualização e análise de dados poderá detectar padrões nos dados e correlações entre dados de produção e qualidade em grande escala, de modo a facilitar a deteção de problemas de qualidade e de optimização de stock, a partir da visualização dos dados de produção?

Hipótese de Investigação: Se este sistema utilizar tecnologias de análise e visualização de dados por forma a descobrir estes padrões e correlações, e se esses padrões e

correlações entre os dados existirem, então será possível correlacionar estes dados de qualidade e de produção e detectar problemas de qualidade e optimização da produção a partir dos dados de produção disponibilizados.

Pergunta de Investigação: Como será desenvolvida esta plataforma e quais as técnicas e tecnologias existentes para visualização de grandes volumes de dados que podem ser aplicadas no contexto da industria?

Hipótese de Investigação: Se iremos trabalhar sobre grandes volumes de dados, com o objectivo de visualizar e analisar os mesmos, então uma possível abordagem será o desenvolvimento de uma plataforma que integra as tecnologias de *Big Data* e visualização interativa de dados, em plataforma web.

1.6 Estrutura do Documento

Esta dissertação encontra-se dividida em 4 capítulos (introdução, estado de arte, desenvolvimento da plataforma e conclusão), cada capítulo com várias secções e sub-secções. As figuras e tabelas apresentadas ao longo do documentos estão numeradas na forma $(x.y)$, onde x refere-se ao capítulo onde esta é utilizada e y é a ordem em que aparece no capítulo.

Um resumo das temáticas abordadas em cada um dos capítulos é apresentado em seguida:

- O capítulo 2 apresenta um resumo da literatura estudada sobre o estado de arte, e está dividida em 4 secções. A primeira aborda a visualização, o seu processo e a necessidade de interatividade na visualização para uma mais fácil análise dos dados. A segunda secção aborda a utilização da cor na visualização e como utilizar cor de uma forma eficaz para transmitir uma ideia ou facilitar a aquisição de novos conhecimentos acerca dos dados estudados. A terceira secção apresenta vários exemplos de técnicas de visualização de dados, desde as técnicas mais simples às mais complexas, e os cenários onde estas ajudam a facilitar a visualização e interpretação dos dados. O quarto capítulo aborda algumas das plataformas *dashboard* para visualização de dados existentes no mercado, com alguns prós e contras em relação à sua utilização.
- O capítulo 3 apresenta o desenvolvimento da plataforma de visualização de dados, assim como o desenvolvimento das visualizações. Este capítulo aborda dois cenários diferentes de produção com objetivos semelhantes: a industria automóvel no caso da Volkswagen Autoeuropa e a industria cerâmica no caso da Ria Stone.
- O capítulo 4 apresenta uma análise e discussão dos resultados obtidos no capítulo anterior, o que correu bem e o que correu mal, assim como as conclusões a que as visualizações criadas nos permitiram chegar.

- O capítulo 5 apresenta as conclusões desta dissertação assim como algum trabalho futuro.

ESTADO DE ARTE

Este capítulo apresenta uma síntese da literatura estudada sobre o estado de arte das técnicas e tecnologias de análise e visualização de dados, e está dividido em 4 secções. A secção 2.1 aborda os conceitos básicos da visualização e a importância da interatividade com a visualização para uma análise e interpretação mais simples e intuitiva dos mesmos. A secção 2.2 explica a utilização de cor na visualização e como utilizar a mesma para transmitir uma ideia ou conhecimentos acerca dos dados mais fácil e eficazmente. A secção 2.3 apresenta vários exemplos de técnicas de visualização utilizadas para vários tipos de dados, e quando se deve utilizar cada tipo de visualização. Na secção 2.4 são abordadas algumas plataformas de visualização de dados para *Business Intelligence*, com os seus pros e contras, consoante o tipo de dados, fontes de dados e análises pretendidas.

2.1 Visualização

Visualização é o ato de interpretar dados em termos visuais, normalmente com recurso a computador e de um modo interativo, de modo a contribuir para um aumento do conhecimento. A visualização é uma representação visual ou imagem mental de dados, cenas ou imagens, e como tal, pode ser considerada semelhante a percepção. É utilizada para representar ideias ou dados de maneira mais clara e precisa possível, de modo a ser mais fácil entender e fazer interpretações sobre os mesmos. O objetivo da visualização é permitir uma fácil análise e exploração de dados, e aquisição de novos conhecimentos sobre os dados visualizados [6].

2.1.1 O Processo de Visualização

Segundo Chittaro [7], o processo de visualização pode ser dividido em 6 passos: Seleção, Mapeamento, Apresentação, Interatividade, Usabilidade e Avaliação.

- *Mapeamento*: Codificação de informação de forma a possibilitar a sua visualização, utilizando features gráficas (formas, cores, textura, posição...) para diferenciar valores ou tipos diferentes de dados. Um bom mapeamento produz uma boa representação visual dos dados em estudo e é alcançado quando existe uma relação precisa entre os dados e a sua representação visual.
- *Seleção*: O passo mais importante da visualização. Neste passo seleciona-se, dos dados disponíveis, os dados que são relevantes ao problema em questão. É considerado o passo mais importante pois a seleção de dados errados ou desnecessários pode levar os utilizadores a enganos em decisões ou que pode levar a aquisição de conhecimento errado.
- *Apresentação*: Uma boa apresentação dos dados implica uma boa organização e gestão do espaço disponível para utilização. É também bastante importante que os dados sejam apresentados de forma clara e inteligível.
- *Interatividade*: Interatividade na visualização é a existência de ferramentas que permitem uma melhor exploração dos dados, através do rearranjo e reorganização dos dados. Uma boa interatividade permite aos utilizadores uma maior capacidade de explorar e entender a informação contida nos dados apresentados e, portanto, é também uma parte importante na visualização.
- *Fatores humanos*: Os fatores humanos têm que ser também abordados, visto que o utilizador tem um grande impacto no design da visualização. Fatores como a facilidade de utilização, e a adaptação a deficiências visuais, como o daltonismo, são fatores a ter em atenção aquando do design da visualização de dados. Por exemplo, uma má utilização de cores pode impossibilitar uma pessoa daltónica de perceber os dados. Para obter uma boa visualização, é necessário ter em conta o conceito de percepção, e o efeito da percepção na visualização.
- *Avaliação*: Depois de criada a visualização é necessário que esta seja testada em utilizadores de modo a se perceber se a visualização é eficaz, clara e consegue transmitir o conhecimento que se pretende transmitir.

2.1.2 Porquê Interatividade?

A visualização estática de dados oferece uma visão previamente definida dos dados, e como tal, por vezes são necessárias várias destas para que se tenha uma visão mais completa da informação que se quer passar. O número de dimensões que se conseguem representar numa tela é limitado, e visualizar dados multi-dimensionais numa imagem estática é complicado, portanto é necessário que existam ferramentas que nos permitam explorar de uma maneira mais simples e eficaz este tipo de dados. Estas ferramentas permitem que pessoas comuns possam explorar os dados sozinhas. As funções de visualização interativa de dados foram propostas em 1996 por Bem Shneiderman, da Universidade de

Maryland, e pouco mudaram até agora, com a máxima "*overview first, zoom and filter, then details-on-demand*". Esta visualização oferece uma visão geral dos dados, em conjunto com ferramentas que permitem explorar em detalhe diferentes aspectos dos dados apresentados. Esta visualização torna os dados acessíveis a várias audiências, desde as pessoas sem experiência com tratamento de dados a investigadores familiarizados com os dados. Este tipo de visualização também encoraja a exploração dos dados, pois tornam a exploração em algo semelhante a um jogo, o que chama a atenção para audiências que não tenham à partida grande interesse nos dados [8].

2.2 Cor na Visualização

O mapeamento de dados com recurso à cor é uma técnica muito importante de visualização, no entanto, a escolha da escala de cores utilizada para representar os dados tem um papel muito importante no apoio à interpretação os mesmos. Portanto, não se deve escolher a escala de cores apenas por ser “bonita”, pois adicionar à visualização uma cor que não acrescente informação útil pode criar confusão nos utilizadores. É, então, necessário escolher as cores de forma a que se represente a informação a transmitir de uma maneira clara [9].

2.2.1 Modelos de Cor

O objetivo de um modelo de cor é padronizar a especificação das cores. É uma especificação de um sistema de coordenadas em que cada cor é representa um único ponto [10]. Segundo Silva et al. [9] Existem duas classes de modelos de cor:

- *Device-dependent*, em que o modelo permite a representação da gama de cores de cada dispositivo e a mesmas coordenadas podem ser representadas por cores ligeiramente diferentes de dispositivo para dispositivo. Um exemplo disto é o sistema de cores RGB, em que cada cor é representada por três coordenadas de cores primárias: vermelho (R), verde (G) e azul (B). Como não existe um valor único de cromaticidade destas três cores primárias, a mesma coordenada RGB pode dar origem a cores diferentes em dispositivos diferentes.
- *Device-independant* ou modelo perceptualmente uniforme, em que o modelo fornece uma representação de cor com um sistema de coordenadas independente dos dispositivos. Num modelo de cor perceptualmente uniforme, a distância euclidiana entre um par de cores está diretamente relacionada com a sua distância perceptual. O primeiro modelo perceptualmente uniforme foi o modelo de Munsell [11].

2.2.2 Propriedades das Escalas de Cor

Tendo uma sequência ordenada de valores representados por cores, é possível identificar as características desejadas para a escala de cor utilizada:

- **Ordem** – As cores escolhidas para representar os valores têm que ser entendidas como tendo a mesma ordem que os valores que estas representam. Por exemplo, na representação de uma escala de temperaturas, a utilização de cores mais frias para temperaturas mais baixas, e cores mais quentes para temperaturas mais altas.
- **Uniformidade e distância representativa** – A cor utilizada para representar dois valores deve transmitir a noção de distância entre os mesmos. Cores que representem valores igualmente distantes devem parecer igualmente diferentes. Valores claramente diferentes devem ser representados por cores visivelmente diferentes, e valores próximos devem ser representados por cores mais parecidas entre elas.
- **Limites** – Se existirem ou não limites nos dados representados, as cores selecionadas devem transmitir esta noção de existência ou não de limites, ou seja, a escala de cores tem que ser capaz de representar valores contínuos ou discretos.

2.2.3 Componentes de Modelos de Cor

Quando se usa uma escala de cores para representar dados univariados, cada cor representa um único valor. De seguida apresentam-se alguns exemplos de escalas de cores utilizados para representar valores contínuos de dados apresentados por Rheingans [12].

- *Grey scale* – Mapeamento de cada valor a um valor de brilho. A escala varia entre o preto e o branco, normalmente com os valores mais baixos dos dados a serem representados por cores mais perto do preto e os valores mais altos dos dados representados por cores mais perto do branco. É fácil de perceber a ordem dos dados, e tem como desvantagem o baixo contraste entre cores, que impede a sua utilização em *datasets* com grandes quantidades de dados. Pode ser observado na figura 2.1.
- *Rainbow scale* – Mapeamento de valores às várias cores do arco-íris, variando o tom e mantendo a saturação e brilho constantes. Este modelo possui, no entanto, alguns problemas. Para utilizadores que não estão familiarizados com a ordem das cores no espectro de luz, é difícil interpretar a ordem dos dados. Outro problema é que algumas cores, como por exemplo o amarelo, chamam mais a atenção os utilizadores pois são percebidas como tendo mais brilho que outras cores, o que pode interferir com a representação de valores nos extremos da escala, que são representados por cores percebidas como tendo menos brilho. Outro problema deste modelo é que, segundo o modelo de Munsell, o amarelo tem o menor número de níveis de saturação perceptíveis, o que torna mais difícil distinguir pequenas diferenças de saturação do amarelo comparando com outras cores. Os seus extremos da escala são as cores vermelho e violeta, o que pode gerar confusão, visto que são cores parecidas. Uma maneira de contornar este problema é não utilizar os extremos do espectro. Exemplificado na figura 2.1.

- Escala de saturação – Mapeamento de valores à saturação, mantendo o tom. Os valores mais altos têm valores de saturação maiores. A vantagem desta escala de cores é que é simples e intuitiva. No entanto, tem muitos poucos níveis perceptíveis de cor, o que limita a sua utilização a baixas quantidades de dados.

2.2.4 Escalas de Cor Redundantes

Utiliza vários parâmetros de cor para representar dados, de modo a poder representar diferentes tipos de informação [12]. Utiliza os benefícios das várias características dos parâmetros da cor para representar informação: o brilho é melhor para representar formas enquanto que o tom é melhor para representar distinguíveis níveis. Estas escalas de cor permitem que um parâmetro ambíguo (causado, por exemplo, por deficiências visuais) seja compensado pelos outros, permitindo na mesma que a informação seja transmitida. Permite também melhor distinção entre valores, por aumentar a diferença visual entre valores [9].

Alguns exemplos de escalas de cor redundantes, ilustrados na figura 2.1:

- Componentes de modelos de cor redundantes – Uma escala de cor redundante simples pode ser construída mapeando os dados ao brilho e tom. Pode ser utilizado por pessoas com daltonismo dicromático.
- Escala *heated-object* – Uma junção das escalas grey scale e rainbow scale que vai de preto a branco, e passa por laranja e amarelo. Permite uma melhor percepção de ordem pois tem um aumento monotónico de brilho.
- Escala de cor *Linearized Optimal* – Uma escala proposta por Levkowitz et al. [13] que maximiza o número de diferenças perceptíveis (just noticeable differences) e preserva a ordem da *rainbow-scale*.

2.2.5 Escalas Divergentes

Criadas através da junção de duas escalas monotónicas que aumentam e se juntam num ponto comum. As duas escalas têm cores diferentes e terminam geralmente numa cor neutra como o branco ou cinzento. Um exemplo de uma escala Divergente é uma escala de polaridade, na qual a cor vermelha é associada à polaridade Norte e o azul à polaridade Sul, em que ambas têm a mesma grandeza, mas sinais opostos [9].

2.2.6 Características de Visualização e Seleção de Escalas de Cor

As escalas de cor utilizadas para construir uma visualização dependem de vários fatores, como características dos dados, os objetivos da visualização e o público alvo.

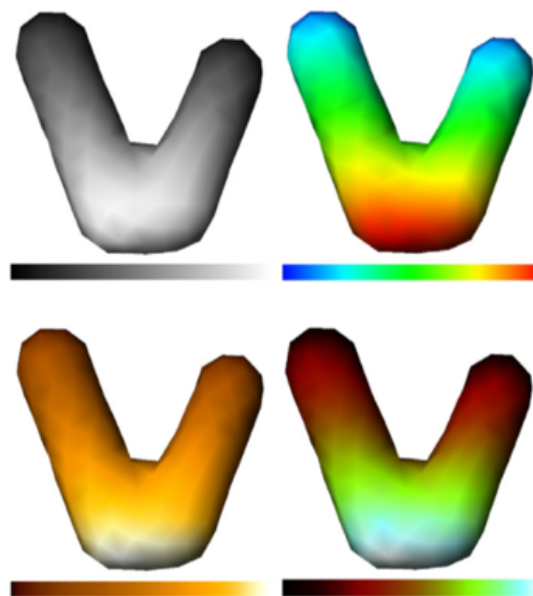


Figura 2.1: Escalas de Cor. Da esquerda para a direita e de cima para baixo: *gray scale*, *rainbow scale*, escala *heated-object* e escala *linearized optimal* aplicadas a um *dataset*.

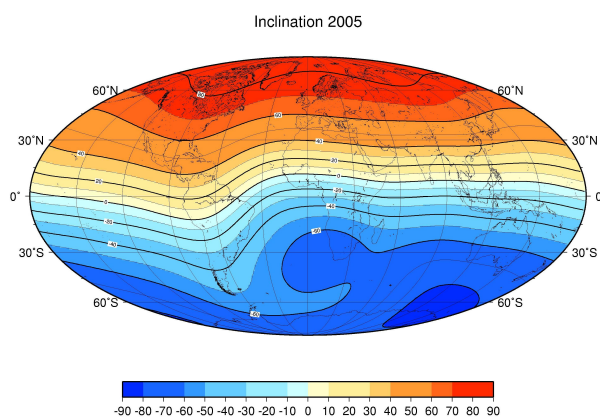


Figura 2.2: Escala de cor divergente aplicada a um mapa da intensidade do campo magnético da terra [14].

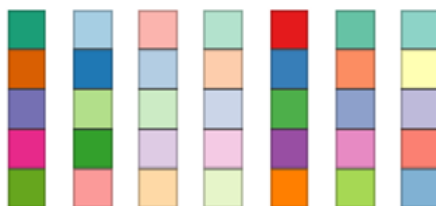


Figura 2.3: Escalas de cor qualitativas disponíveis no site ColorBrewer 2.0 [15].

2.2.6.1 Tipos de Dados

Quando se desenvolve uma visualização, é importante que esta represente as *features* mais importantes dos dados e esta que não atraia a atenção para as *features* menos importantes, para que não se retire atenção das mais importantes. É importante escolher escalas de cor apropriadas para os tipos de dados que se querem representar.

- Dados Nominais – Como os dados representam nomes, é impossível aplicar a este tipo de dados operações matemáticas. Como tal, a representação dos mesmos não deve implicar uma ordem dos dados. Deve-se utilizar escalas de cor segmentadas como as visíveis na figura abaixo. O número de cores utilizadas não deve passar de sete, devido à habilidade dos utilizadores de diferenciar cores e de se lembrar dos significados de cada uma.
- Dados Ordinais – Os dados possuem uma ordem que é representada pelos valores atribuídos, sem, no entanto, ter informação sobre o espaçamento entre elementos. A escala de cores escolhida deve permitir a diferenciação entre cores e a percepção de ordem dos valores. Um bom exemplo de escala de cores para este tipo de dados é uma escala monotónica com variação de brilho ou multitónica com ordem perceptível, como ilustrado na figura 2.4.

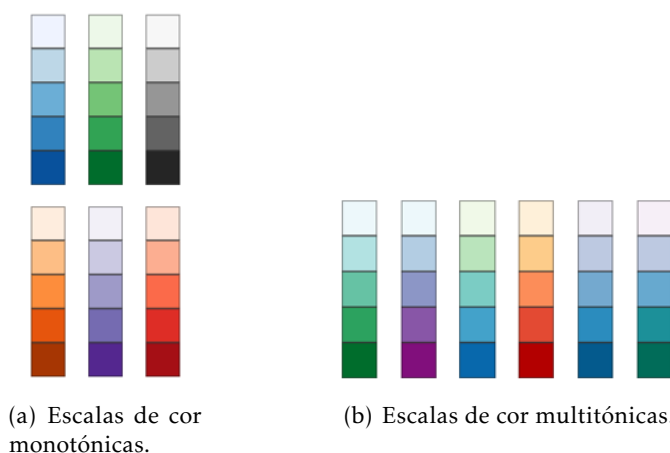


Figura 2.4: Escalas de cor sequenciais disponíveis pelo website ColorBrewer 2.0 [15].

- Dados Intervalares – Neste tipo de dados, a distância entre os valores é igual. Como tal, estas distâncias devem ser representadas por diferenças iguais entre a magnitude visível das cores que representam cada um dos valores consecutivos na escala.
- Dados de Razão – Os dados de razão são dados quantitativos que consistem em valores que aumentam a partir de um zero absoluto e a razão entre valores é igual. Pode ser representado com recurso a escalas sequenciais. Se o zero absoluto for um ponto central da escala, pode-se usar escalas divergentes para representar este tipo de dados, como as escalas da figura 2.5.

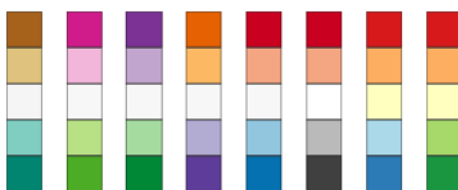


Figura 2.5: Escalas de cor divergentes disponibilizadas pelo website ColorBrewer 2.0 [15].

2.2.6.2 Objetivos

A escolha da escala de cores utilizada deve ter em conta o objetivo da visualização. Por exemplo, quando o objetivo é analisar quantidades métricas nos dados, uma escala de cores que não varie monotonicamente no canal oposto (brilho, red-green, yellow-blue) funciona bem. Quando a tarefa envolve a análise qualitativa da distribuição de valores, escalas de cor que variem sistematicamente em brilho são mais adequadas, permitindo à nossa visão aplicar os mecanismos de familiar shade-from-shaping [12].

Em situações nas quais o tamanho de um objeto é um fator importante, há que ter também atenção às cores utilizadas na sua representação. Um estudo de Tedford et al. [16] mostrou que retângulos com o mesmo tamanho, saturação e brilho eram vistos como tendo tamanhos diferentes consoante a sua cor. Retângulos com cores mais quentes (amarelo, laranja, vermelho) parecem maiores que os retângulos de cores frias (azul, verde).

Um estudo de Cleveland e McGill [17] investigou este fenómeno de ilusão de cor-tamanho em mapas estatísticos. Utilizando um mapa do Nevada, no qual condados como aproximadamente a mesma área foram coloridos de verde e vermelho. Foi pedido aos sujeitos da experiência que respondessem qual a cor que corresponde à maior área, ao qual responderam mais vezes que as áreas dos condados de cor vermelha eram maiores, quando comparado com os que responderam que as áreas dos condados verdes eram maiores. A experiência foi repetida para os mesmos tons de verde e vermelho, mas com menos saturação, sendo que não foi verificada a tendência da primeira experiência, o que sugere que a saturação das cores tem um forte efeito na perceção de tamanho de um objeto. Portanto, para tarefas em que o tamanho de um objeto é importante, não se devem utilizar cores muito saturadas.

Segundo Rogowitz e Treinish [18], as regras usadas para criação de escalas de cor isomórficas utilizadas em dados de razão e intervalares são também bastante eficazes para segmentar dados. Tal como mencionado em [19], a componente de luminância é utilizada para representar monotonia para altas frequências espaciais enquanto a componente de saturação é utilizada para representar monotonia para baixas frequências espaciais. No entanto, como neste caso os níveis estão bem definidos, é possível também usar a componente de luminância em baixas frequências espaciais para representar monotonia. Nestas escalas de cor segmentadas, é importante que se consigam discriminar as várias cores entre si, o que limita o número de níveis que podem ser representados. Segundo o estudo [18], mais níveis podem ser discriminados a baixas frequências espaciais do que a altas frequências. Na figura abaixo, é possível ver e ilustração disto. À esquerda temos uma escala de cores segmentada com cinco níveis e à direita uma escala de cores segmentada com dez níveis. Para o caso do campo magnético da Terra (em cima), no qual a frequência espacial é baixa, mais níveis dão-nos mais informação. No caso da representação das observações de frações de nuvens (em baixo), que possui alta frequência espacial, mais níveis não trazem mais informação, antes pelo contrário, diminuem a segmentação.

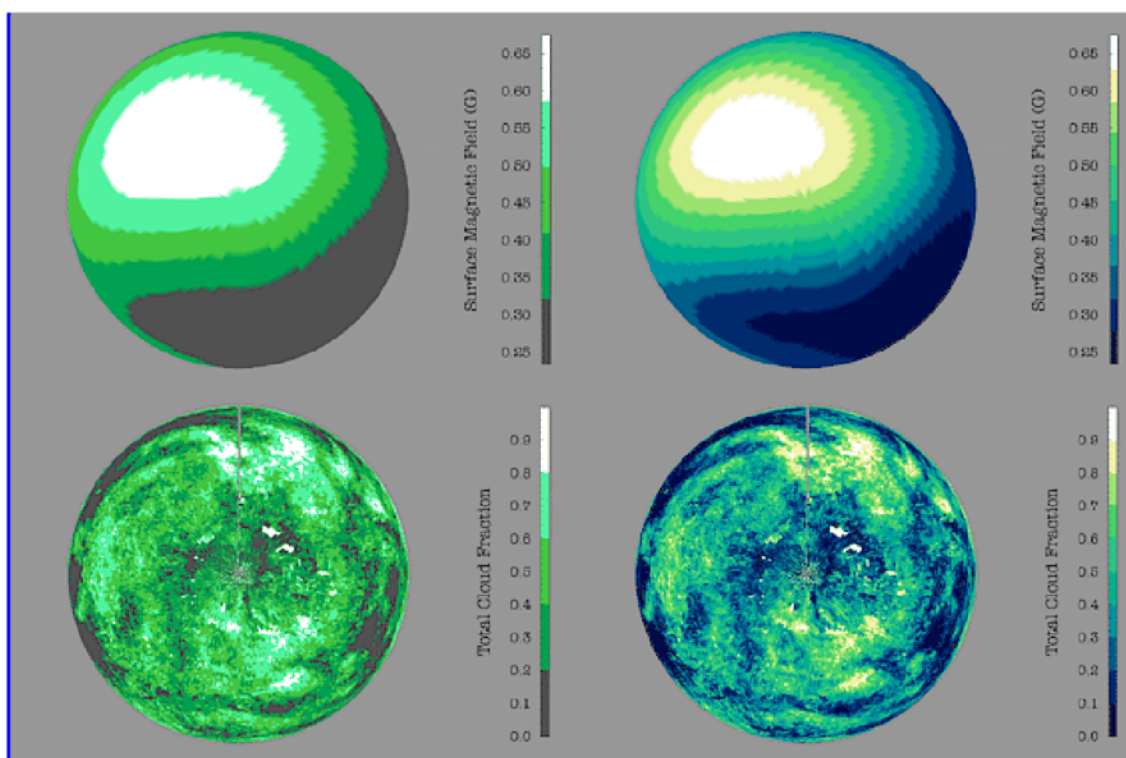


Figura 2.6: Ilustrações do campo magnético superficial da Terra (em cima) e da observação de frações de nuvens (em baixo), para cinco níveis (à esquerda) e 10 níveis (à direita). Imagem retirada de [18].

Segundo Wang et al. [20], cores mais vivas (cores saturadas e brilhantes) dão mais nas vistas e chamam a atenção para uma *feature* em particular, criando um efeito de *pop-out*, podendo, portanto, ser utilizadas quando se quer chamar a atenção para uma ou mais

features dos dados. É também mencionado que objetos com tons de cor semelhantes são interpretados como pertencendo a um mesmo grupo, enquanto que objetos com tons de cor diferentes têm o efeito oposto, e são interpretados como pertencendo a grupos diferentes. A cor pode ser utilizada para organizar a tela em partes perceptualmente semelhantes, vistas pré-atentivamente. Partes da imagem formam áreas distintas e se a mesma cor aparece em partes diferentes da imagem, então estas têm uma ligação. Este efeito pode ser utilizado para detetar grupos de dados semelhantes dentro do *dataset* total.

2.2.6.3 Tipos de Visualização

É importante ter em consideração o tipo de visualização que vai ser utilizada quando se escolhe uma escala de cor, pois os vários tipos de visualizações têm diferentes limitações. Visualização 3D tem limitações diferentes de visualização 2D. Um exemplo destas é o problema que surge com o shading: se os utilizadores usarem o shading como método de identificação da forma 3D de um objeto, uma escala de cores a variar no brilho pode interferir com os valores de brilho que resultam do shading [9].

Devido aos mecanismos de visão humana, o nosso cérebro preenche a área dos objetos de acordo com o contraste nas arestas, portanto, a aparência da cor de um objeto é afetada pela luminosidade e cor do fundo. As cores tendem a parecer mais escuras e pequenas num fundo branco, e mais claras e maiores em fundos escuros [21], como mostra a figura 2.7.

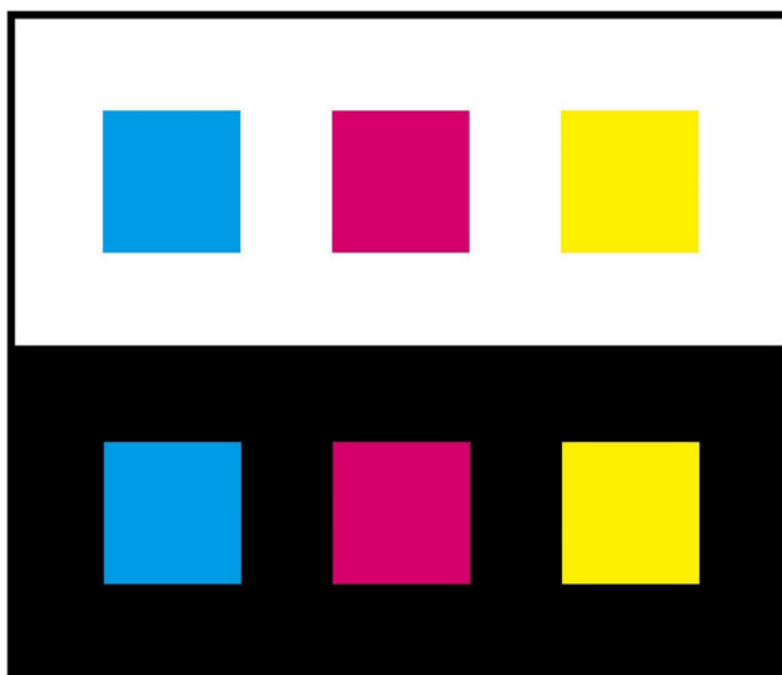


Figura 2.7: Efeito da cor de fundo na percepção de tamanho e brilho de objetos. Imagem retirada de [21].

Fundos de cor podem alterar a percepção de uma cor, parecendo esta tingida com o tom do fundo, num efeito conhecido como indução cromática. A indução cromática pode fazer com que a mesma cor pareça diferente em fundos diferentes e também com que cores diferentes pareçam iguais. Este efeito pode fazer com que se vejam cores de maneira diferente do pretendido [21].



Figura 2.8: Efeito da cor de fundo na percepção da cor de um objecto. À esquerda, dois objetos com a mesma cor em fundos de cores diferentes parecem ter cores diferentes. À direita, dois objetos com cores diferentes em fundos de cores diferentes parecem ter a mesma cor. Imagem retirada de [21].

No caso de interfaces gráficas de utilizador (GUI), é importante que o seu design não distraia o utilizador do seu conteúdo. As frames de janelas, caixas, sliders, botões e ícones devem ser monocromáticos, com cor a ser utilizada para acentuações, de forma discreta. O design de ícones exige clareza, simplicidade e consideração pelo que o utilizador consegue ver a uma distância normal de visão, dados o seu reduzido tamanho. A sua forma é mais importante que a cor, que deve ser utilizada apenas para reforçar simbolismos [21].

Em publicidade, o uso de cores caras em fundos escuros é bastante utilizado para captar a atenção e excitar o observador, da mesma forma que os sinais néon ou luzes de discoteca. Este efeito é chamado “Efeito Las Vegas” [21].

Em texto, o critério mais importante é a legibilidade. A legibilidade é determinada pelo contraste entre o plano de fundo e as letras, que deve ser no mínimo de 3:1 e preferivelmente de 10:1. Deve-se também evitar largas áreas de branco no ecrã, pois estas causam fadiga visual e aumentam as chances de o utilizador ver flicker nas periferias do texto. As combinações de cor mais legíveis são as de maior contraste, como preto ou azul sobre fundo branco ou amarelo, e vice-versa. Vermelho, verde e magenta tornam a leitura difícil e desconfortável. Não se deve também utilizar texto colorido sobre fundo colorido, optando por colorir apenas o texto ou o fundo [21].

A cor pode ser um meio bastante eficaz de aumentar a informação ou torná-la mais fácil de entender. No entanto, esta tem limitações no número de cores que o utilizador consegue diferenciar e no número de cores das quais o utilizador consegue memorizar o significado. Portanto, deve-se utilizar cor com moderação [21]. Um exemplo da utilização

de cor como facilitador de interpretação de informação é o mapa do metro de Londres, no qual cada linha tem uma cor que a identifica, permitindo distinguir as linhas numa rede complexa de linhas de metro.

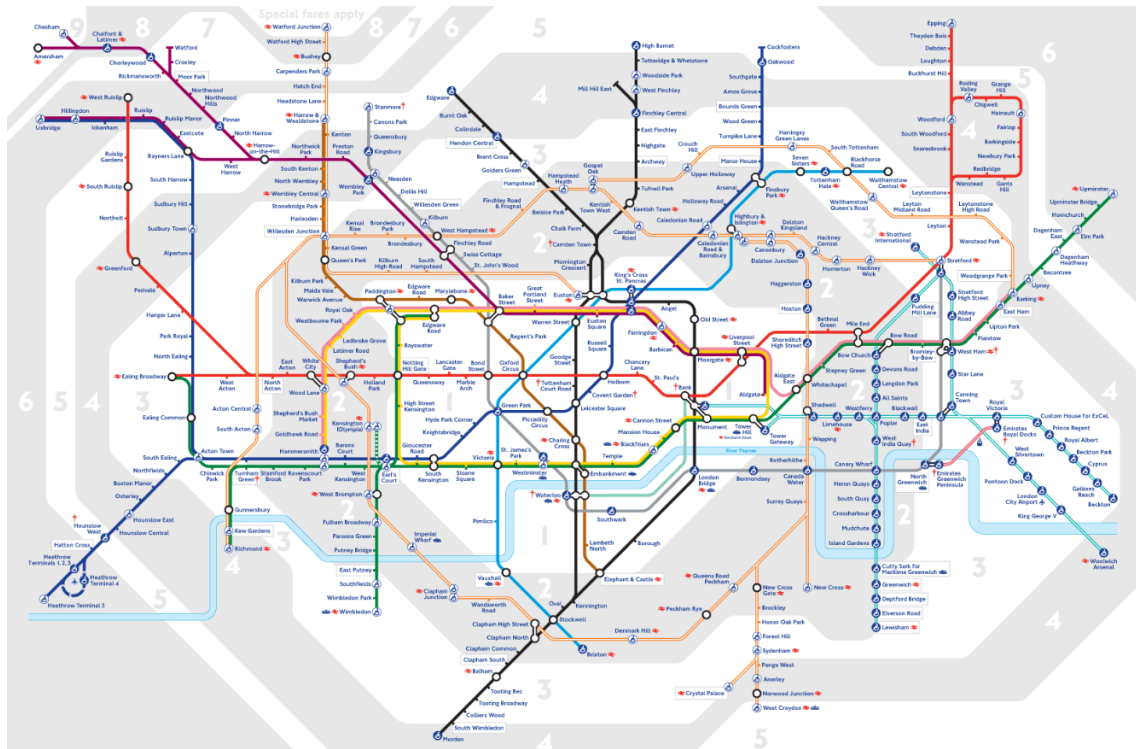


Figura 2.9: Mapa do metro de Londres. Cada linha possui uma cor única que a identifica facilmente das restantes linhas. Imagem retirada de [22].

2.3 Técnicas de Visualização

Segundo o livro *Análise Inteligente de Dados*, de Berthold et al. [23], a visualização de informação foca-se principalmente em *datasets* sem uma semântica 2D ou 3D implícita, e, portanto, não têm também um mapeamento definido dos dados para o ecrã. Existem muitas técnicas de visualização, como por exemplo, gráficos x-y, gráficos de linhas e histogramas. No entanto estas técnicas estão muito limitadas a *datasets* de baixas dimensões.

As técnicas de visualização podem ser classificadas em três critérios: Os dados a ser visualizados, a técnica de visualização e a técnica de interação usada [23].

Os tipos de dados podem ser:

- Dados uni-dimensionais, como por exemplo, dados temporais.
- Dados bidimensionais, como por exemplo, mapas geográficos.
- Dados multi-dimensionais, como por exemplo, tabelas relacionais.

As técnicas de visualização utilizadas podem ser:

- Displays 2D/3D (gráficos de barras, gráficos x-y).
- Displays transformados geometricamente (paisagens, coordenadas paralelas).
- Displays baseados em ícones (needle e star icons).
- Dense pixel displays (padrões recursivos e circle segments).
- Stacked displays (treemaps e dimensional stacking).

As técnicas de interação permitem aos utilizadores navegar e modificar as visualizações, assim como selecionar conjuntos de dados específicos para análise [23]. Alguns exemplos de técnicas de interação incluem:

- *Projeção Dinâmica*, permite uma navegação mais suave pelos dados.
- *Filtragem Interativa*, permite aos utilizadores isolar conjuntos de dados para análise.
- *Zooming*, permite aumentar os dados para análise mais detalhada.
- *Distorção*, permite aumentar o espaço no ecrã alocado aos dados de interesse, preservando o *dataset* inteiro.
- *Linking e Brushing*, permite ao utilizador selecionar os dados de interesse numa view e vê-los realçados noutras *views*.

2.3.1 Tipos de Datasets

Na visualização de informação, os *datasets* consistem em largos valores de dados, cada um deles possuindo um número de variáveis e dimensões. Estes dados correspondem a observações, medidas, transações ou valores físicos obtidos por medições com sensores, guardados num *dataset*. O número de variáveis dos dados corresponde à dimensionalidade do *dataset*. Existem por vezes distinções entre dimensões densas e dimensões com possíveis valores arbitrários e dependendo do número de dimensões com valores arbitrários, os dados podem ser chamados de uni-variados, bi-variados ou multi-variados [23].

2.3.1.1 Datasets Uni-dimensionais

Os *datasets* uni-dimensionais possuem normalmente uma única dimensão densa. Um exemplo destes *datasets* é o de dados temporais, no qual um ou mais valores dos dados podem estar associados a um ponto no tempo [23].

2.3.1.2 *Datasets* Bidimensionais

Um exemplo de *datasets* bidimensionais é o de dados geográficos, em que as duas dimensões representam a latitude e longitude dos dados. Um dos métodos mais conhecidos de visualização de dados bidimensionais são os gráficos x-y, sendo que os mapas são uma versão dos gráficos x-y em que os dados geográficos estão mapeados em x pela longitude e em y pela latitude [23].

2.3.1.3 *Datasets* Multi-dimensionais

Muitos *datasets* possuem mais do que duas dimensões e, portanto, os métodos de visualização 2D e 3D convencionais não são os mais adequados para visualizar este tipo de dados. Exemplos de *datasets* multi-dimensionais são as tabelas de bases de dados relacionais, que normalmente têm várias colunas com valores. Como estes tipos de dados não permitem um mapeamento simples de todas as dimensões, são necessárias técnicas de visualização mais complexas. Uma destas técnicas utilizadas para *datasets* multi-dimensionais são os gráficos de coordenadas paralelas, no qual um segmento de reta interseca eixos paralelos entre si na posição correspondente ao valor dos dados para cada dimensão (cada coluna das tabelas) [23].

2.3.1.4 Texto e *Hypertext*

Nem todos os tipos de dados têm uma dimensionalidade definida. Texto, *hypertext* e conteúdos de websites são exemplo disso. Como tal, e visto que estes dados não podem ser representados por números, a maior parte das técnicas de visualização não pode ser utilizada para os representar.

Para se poder visualizar estes dados, é necessária uma transformação dos dados em vetores de descrição para que se possam aplicar a estes as técnicas de visualização utilizadas noutros tipos de dados. Um exemplo de transformação é a contagem de palavras, que é normalmente acompanhada de uma análise de componente principal ou escalamento multidimensional para reduzir a dimensionalidade do *dataset* [23].

2.3.1.5 Hierarquias e Grafos

Os dados de um *dataset* têm muitas vezes relações entre si, que podem ser ordenadas, hierárquicas ou arbitrárias. Estas relações são comumente representadas por grafos. Grafos consistem numa rede que liga um conjunto de objetos chamados nós através de *edges* ou *links*. Um exemplo de um grafo pode ser visto na imagem seguinte [23].

2.3.1.6 Algoritmos e Software

Trabalhar em projetos de largas dimensões é por vezes um grande desafio, com grandes equipas de programadores e desenvolvedores, perceber os algoritmos e fazer *debug* de código não é uma tarefa fácil. O objetivo da visualização de software é ajudar a perceber

2.3.2.1 Tabela

As tabelas são técnicas de visualização vastamente utilizadas pela sua facilidade de utilização, interpretação e capacidade de representação de dados. É estruturada em linhas e colunas, onde cada linha representa contém um vetor ou registo de dados, as colunas contêm valores, atributos ou propriedades dos dados, e é acompanhada ainda de *meta-data*, como por exemplo as *labels* que identificam as linhas e colunas [25]. Hoffman [26] introduziu o conceito de “*Table Visualization*” para descrever a visualização de *datasets* com recurso a tabelas, e utiliza o termo “*dimensions*” para descrever as linhas (variáveis independentes) e o termo “*variates*” para se referir às colunas (variáveis dependentes).

Paragraphs	C/L Without Spaces	C/L With Spaces	C/L in Inches
Paragraphs 1	77	93(±5)	5.25
Paragraphs 2	111	133(±5)	9.13
Paragraphs 3	116	141(±5)	9.5
Paragraphs 4	102	122(±5)	10.74
Paragraph 5	74	87(±5)	4.87
Paragraph 6	96	115(±5)	8.13
Paragraph 7	114	134(±5)	9.4
Paragraph 8	96	111(±5)	10

Figura 2.11: Exemplo de tabela, retirado de [6].

2.3.2.2 Pie Chart

Também chamado de gráfico circular, este gráfico é formado por um círculo, dividido em setores comumente chamados de “fatias”, em que cada segmento representa uma proporção da quantidade total, representada pelo círculo completo [27]. Cada fatia contém uma fração de dados que tem features de interesse comuns entre si.

Duas variações comuns de *pie charts*: *Doghnut charts* e *Exploding charts*. Os *doghnut charts*, tal como o nome indica, têm um “buraco” no centro, enquanto os *exploding charts* têm os segmentos separados, ao contrário de um *pie chart* comum, onde todas as fatias estão juntas [28].

Existem também *multi-level pie charts*, que possuem vários níveis, onde o círculo central representa o segmento *parent* e os círculos do exterior contêm os segmentos *child*, que dividem o segmento do círculo central em sub segmentos [29]. Estes gráficos estão ilustrados na figura 2.14.

Os *pie charts* são bastante eficazes quando utilizados para comparações entre segmentos. No entanto, é difícil retirar valores precisos, comparar segmentos de tamanhos semelhantes e segmentos de diferentes *pie charts* [30].

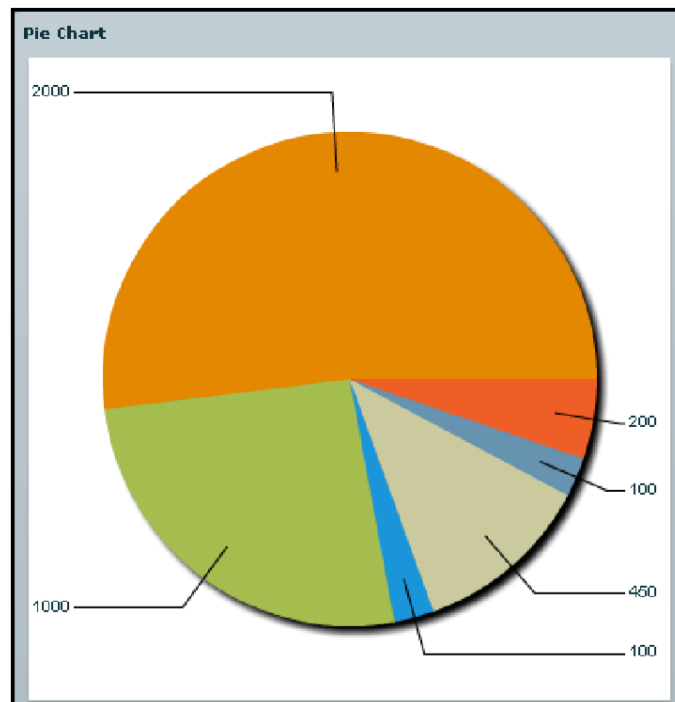


Figura 2.12: Exemplo de um *pie chart*, retirado de [28].

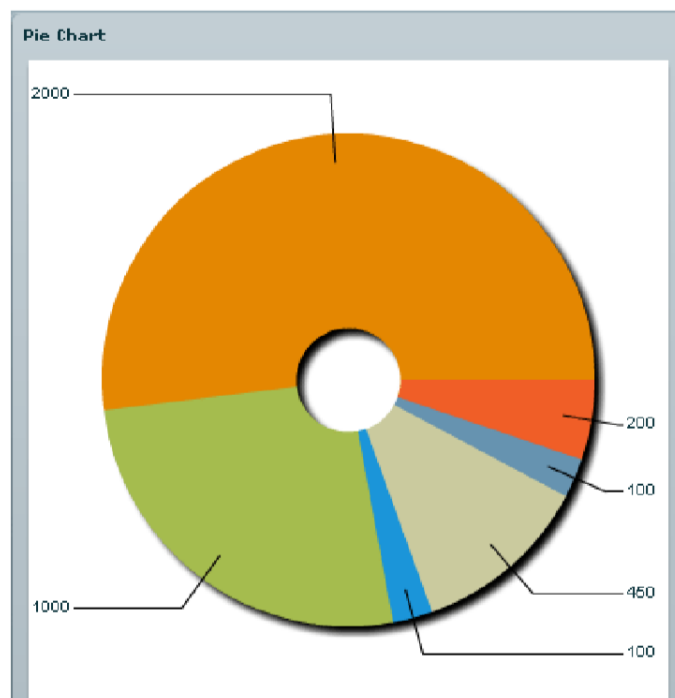


Figura 2.13: Exemplo de um *doghnut chart*, de [28].

Passenger Class and Gender

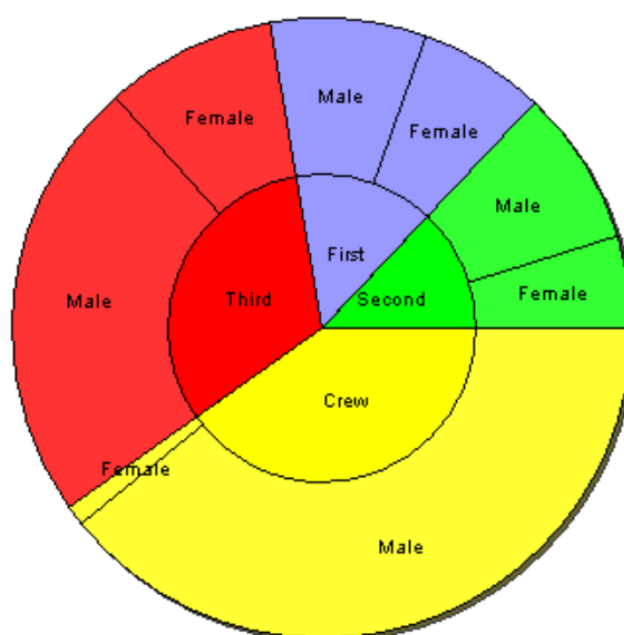


Figura 2.14: Exemplo de *multi level pie chart*, de [28]. Os segmentos são divididos em dois sub segmentos, para homem e mulher, de modo a poder comparar dentro de cada segmento a proporção de indivíduos de cada sexo.

2.3.2.3 Gráfico de Barras

Um dos métodos de visualização mais utilizados, os gráficos de barras são utilizados para representar dados discretos, em que, geralmente, o eixo X corresponde a um parâmetro, e o eixo Y os valores dos parâmetros. Todas as barras têm a mesma largura e a altura é proporcional ao valor de cada parâmetro [31]. Podem ser representados na horizontal ou na vertical, sendo as representações equivalentes.

Existem vários tipos de gráficos de barras, sendo que os mais comuns têm o mesmo ponto de origem, sendo apenas limitados superiormente. No entanto, existem outros tipos, como os *floating column chart* (2.16) ou *High Low Open Close charts*, e *Candlestick charts*, nos quais as barras têm limites superiores e inferiores [28].

2.3.2.4 Histograma

Os histogramas são uma das técnicas de visualização muito importantes em estatística e *data analytics*, introduzidos por Karl Pearson. Representam a distribuição de dados. Semelhante a gráficos de barras, mas normalmente utilizado para representar dados contínuos [32], o histograma utiliza barras horizontais adjacentes, normalmente da mesma largura, que representam classes a que cada *datapoint* pertence [31].

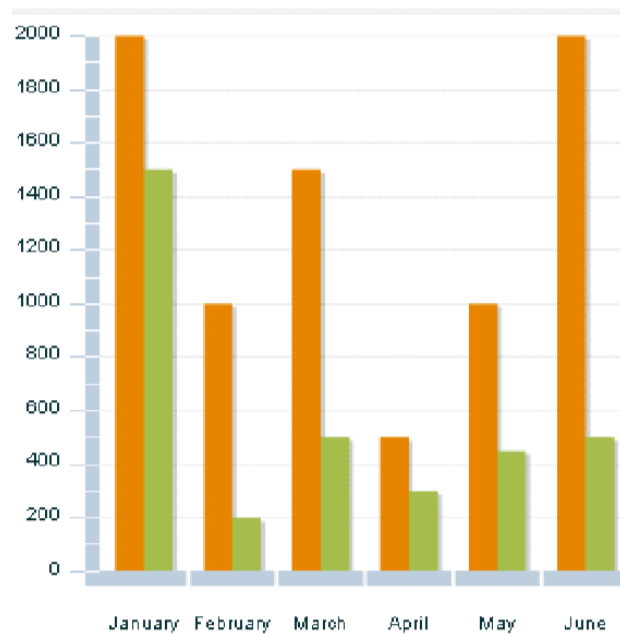


Figura 2.15: Exemplo de um gráfico de barras simples, de [28].

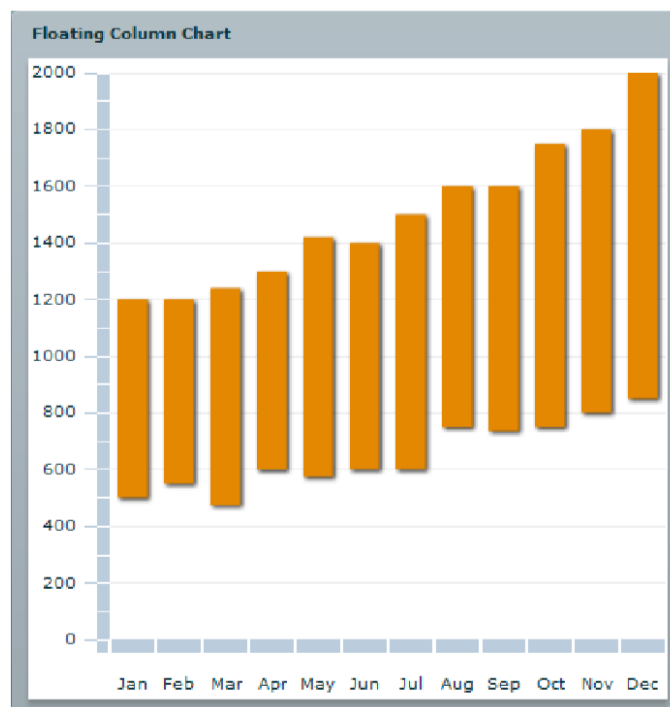


Figura 2.16: Exemplo de um *floating column chart*, de [28].

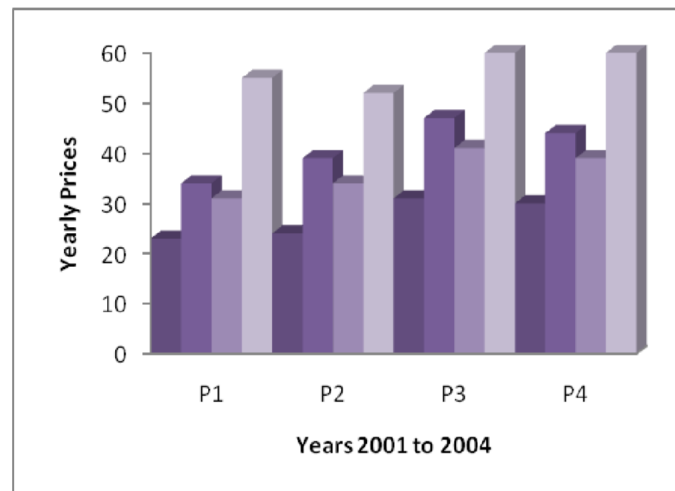


Figura 2.17: Exemplo de histograma, de [6].

2.3.2.5 Scatter Plot

Os *scatter plots* são uma representação gráfica de um *dataset* no plano cartesiano, e ilustra a relação entre duas variáveis, uma delas independente, normalmente representada no eixo horizontal, e a outra dependente da primeira, normalmente representada no eixo vertical [31, 33]. São a representação mais simples de relação entre 2 variáveis não contínuas [34]. Os *scatter plots* são utilizados para avaliar as relações entre dados, analisar a dispersão dos dados e detectar *outliers* [35], valores nos dados bastante diferentes dos restantes, normalmente associados a erros de leituras de sensores ou falta de dados. Na figura 2.18 pode-se ver um exemplo de um *scatter plot*. Os *scatter plots* são visualizações simples e eficazes, tornam a interpretação dos dados fácil, pois permitem uma fácil busca e memória visual. No entanto, com o aumento das dimensões e do número de *data points*, a sua eficácia diminui [36, 37]. Uma forma de contornar esta limitação é a utilização de uma matriz de *scatter plots*, que emparelha apenas duas variáveis do *dataset* em cada um dos *scatter plots*. As variáveis do *dataset* formam as linhas e colunas da matriz, e a relação entre cada variável é observável em cada um dos *scatter plots* da matriz [36]. É possível observar um exemplo de uma matriz de *scatter plots* na Figura 2.28, da página 38.

2.3.2.6 Gráfico de Linha

Os gráficos de linha são semelhantes a *scatter plots*, com a diferença que os pontos estão unidos por segmentos de reta contínuos. São utilizados para representar variáveis contínuas ou variáveis amostradas de um espaço contínuo [31]. Os gráficos de linha podem ser representados por ícones ligados por segmentos de reta, ou simplesmente por segmentos de reta que intercetam os pontos que correspondem aos dados, sem ícones [28]. São utilizados, por exemplo, para visualizar tendências e comportamentos dos dados ao longo de um espaço de tempo [34, 38].

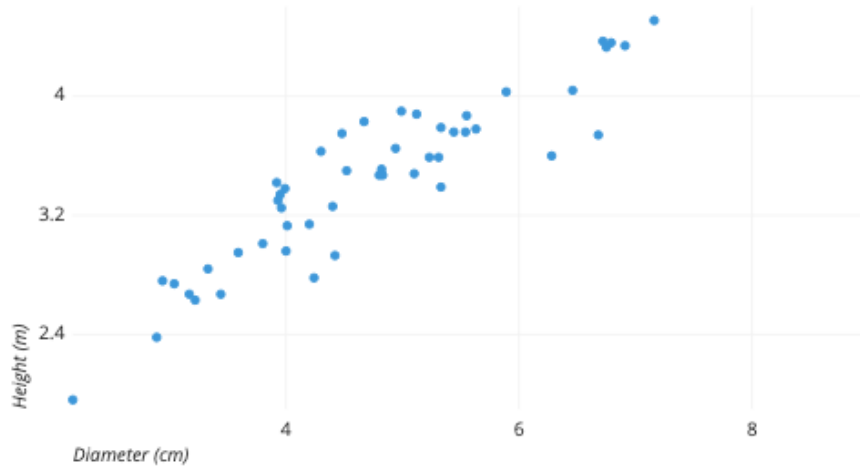


Figura 2.18: Exemplo de um scatter plot, de [35].

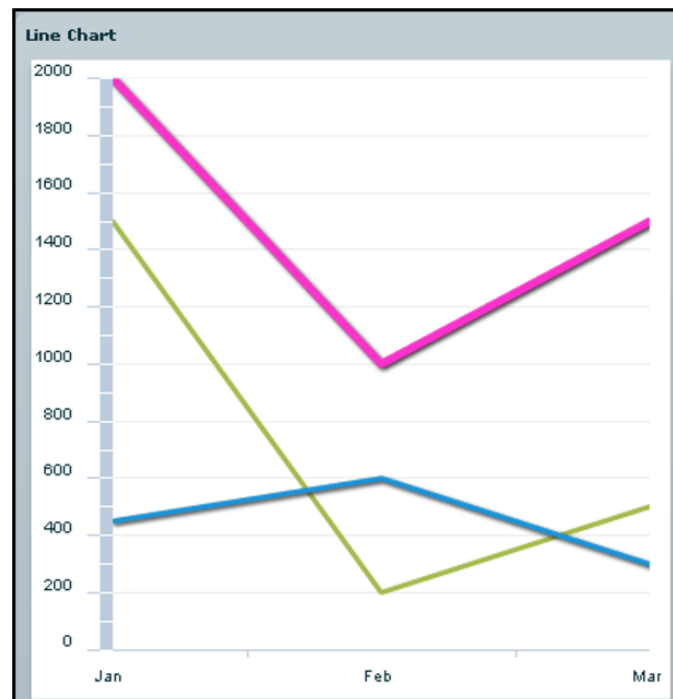


Figura 2.19: Exemplo de um gráfico de linha, de [28].

2.3.2.7 Gráfico de Área

Os gráficos de área são utilizados para representar dados quantitativos. Utiliza um gráfico de linha para definir o limite superior da área e, portanto, tem as mesmas características. A área abaixo do gráfico de linha é preenchida com uma cor ou padrão [28].

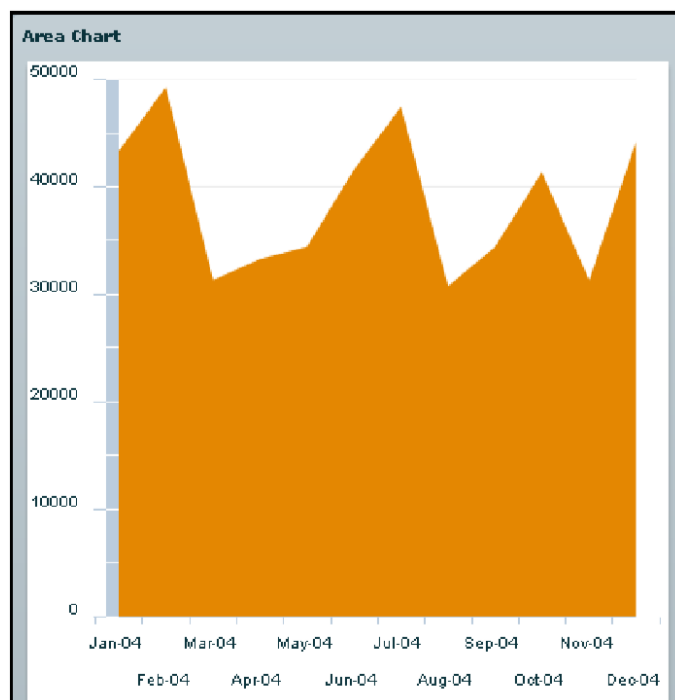


Figura 2.20: Exemplo de um gráfico de área, de [28].

2.3.2.8 Bubble Chart

Os *bubble charts* podem ser utilizados para representar *datasets* em que cada *datapoint* tem três valores. Estes 3 valores dão a posição da bolha em x, a posição em y e o tamanho da bolha [28]. O *bubble chart* é uma variação do *scatter plot*, que substitui os pontos por bolhas. Na figura 2.21 pode-se ver um exemplo de *bubble chart*.

2.3.2.9 Multiple Data Series Chart

Combina vários tipos de gráficos num só, de modo a melhorar e facilitar a visualização de certos aspectos dos dados. Na figura 2.22, é possível ver um *multiple data series chart* com a representação de duas *stock charts*, uma delas representada por um gráfico de barras e a outra representada por um gráfico de linhas.

2.3.2.10 Displays geometricamente transformados

Os *displays* geometricamente transformados têm como objetivo encontrar boas transformações geométricas de *datasets* multi-dimensionais [23]. Estas técnicas de visualização

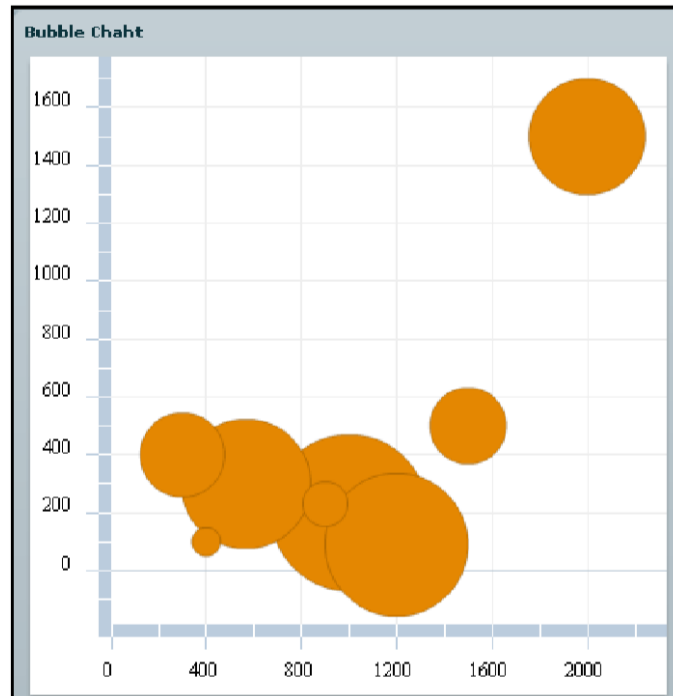


Figura 2.21: Exemplo de um *bubble chart*, de [28].

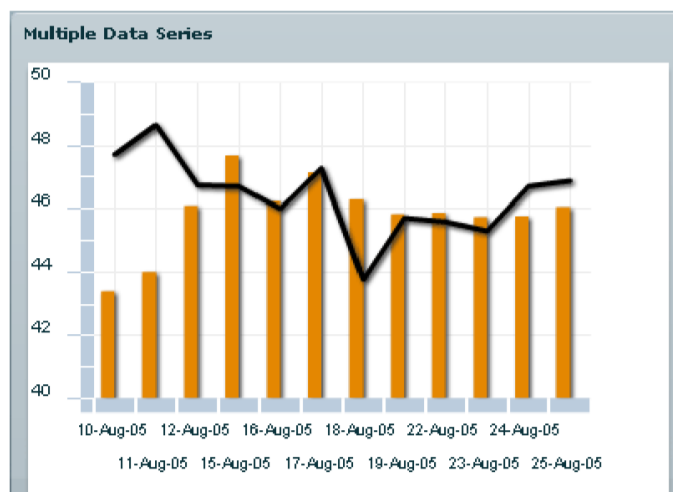


Figura 2.22: Exemplo de um *multiple data series plot*, de [28].

incluem técnicas de exploração estatística como matrizes de *scatter plots* [24], “*projection pursuit*”, uma classe de técnicas que procura projeções com qualidades de interesse, *Prosection Views*, uma técnica onde apenas os dados selecionados pelo utilizador são projetados e a técnica de coordenadas paralelas (Figura 2.23), uma técnica bastante conhecida onde um *dataset* com n dimensões é mapeado num *display 2D* utilizando n eixos paralelos entre si, com escalas lineares que vão de um valor mínimo ao valor máximo. Cada *datapoint* é representado por um conjunto de segmentos de reta interligados que intercetam os eixos na localização correspondente ao valor de cada dimensão [23, 37, 39, 40].

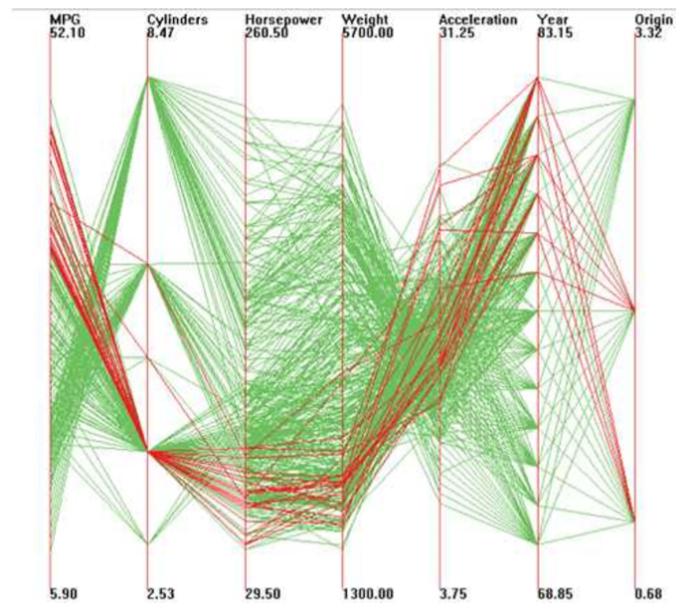


Figura 2.23: Exemplo da técnica de coordenadas paralelas, retirada de [23].

2.3.2.11 Display de Ícones

Mapeia os atributos de cada *datapoint* de um *dataset* às características de um ícone. Por exemplo, no caso de o ícone ser uma *stick figure*, as dimensões do *datapoint* podem ser mapeadas para a posição, tamanho do ícone, ângulos dos membros e tamanho dos membros do ícone. Os ícones podem ser escolhidos arbitrariamente [23, 24]. A figura 2.24 ilustra um exemplo de um *display* de ícones de estrela, em que cada dimensão controla o tamanho dos raios que partem do centro do ícone.

2.3.2.12 Dense Pixel Display

Mapeia cada dimensão de um *datapoint* a um pixel e agrupa os pixels pertencentes a cada dimensão em áreas contíguas. Permite a visualização de uma grande quantidade de dados, visto que cada valor dos dados é mapeado para um único pixel do *display*. Dois exemplos deste tipo de técnicas de visualização são as técnicas de padrões recursivos e de segmentos de círculo [23, 24, 39].

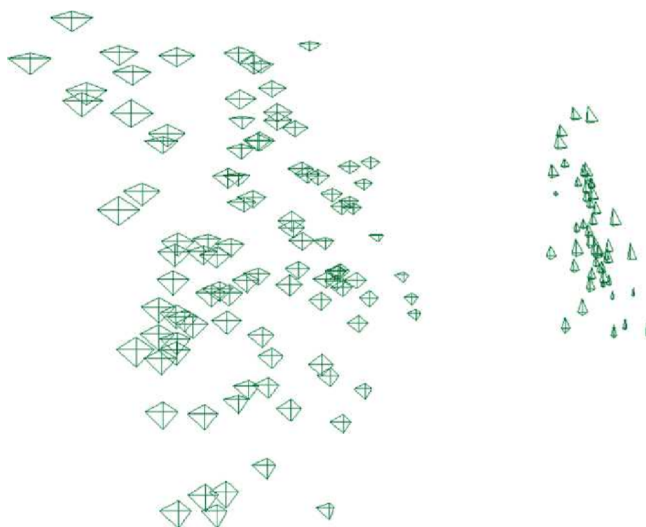


Figura 2.24: Exemplo de um display de ícones, retirado de [23].

A técnica de padrões recursivos arranja os pixels num padrão recursivo em forma de S, e é utilizado para representar *datasets* com uma ordem natural aparente relativamente a um atributo (Figura 2.25) [41].

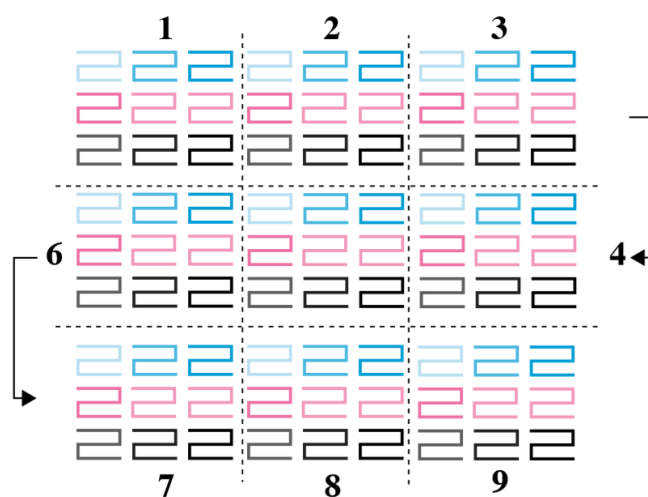


Figura 2.25: Exemplo de um *dense pixel display* com técnica de padrões recursivos, de [41].

A técnica de segmentos de círculo representa os dados num círculo dividido em segmentos, um para cada atributo. Tal como na técnica de padrões recursivos, cada pixel representa um único valor dos dados. A distribuição dos dados começa no centro do círculo, até ao exterior movendo-se ortogonalmente, também em S, a uma linha que passa no centro do segmento e vai também do centro do círculo ao seu perímetro (Figura 2.26) [42].

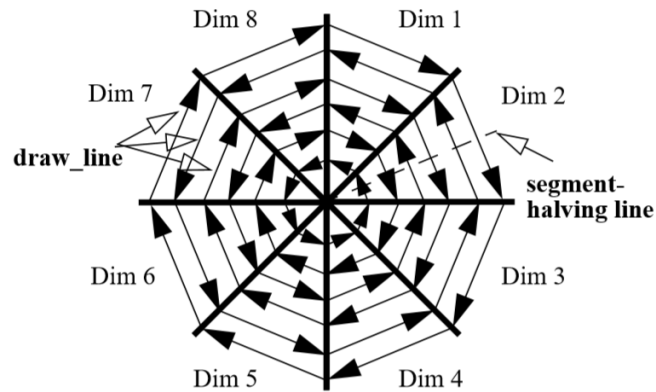


Figura 2.26: Exemplo de um *dense pixel display* com técnica de segmentos de círculo, de [42].

2.3.2.13 *Stacked displays*

Apresentam os dados de uma forma segmentada e hierárquica [39]. No caso dos *datasets* multi-dimensionais, selecionam-se as dimensões mais adequadas para segmentar os dados e construir a hierarquia. Um exemplo de técnicas de *stacked displays* é o *dimensional stacking* [43], que possui sistemas de coordenadas dentro de outros sistemas de coordenadas. Para criar uma visualização *dimensional stacking*, divide-se o nível exterior de coordenadas em células retangulares, e em cada célula são utilizados dois atributos como coordenadas da célula. A segmentação pode ser realizada várias vezes, dependendo da dimensionalidade do *dataset*. A utilização deste tipo de visualização depende da distribuição dos dados, e das dimensões escolhidas para o nível exterior, e é, portanto, necessário escolher com cuidado as dimensões utilizadas [23]. Uma solução para este problema é selecionar as dimensões mais importantes primeiro. Um exemplo de *dimensional stacking* é apresentado na figura seguinte.

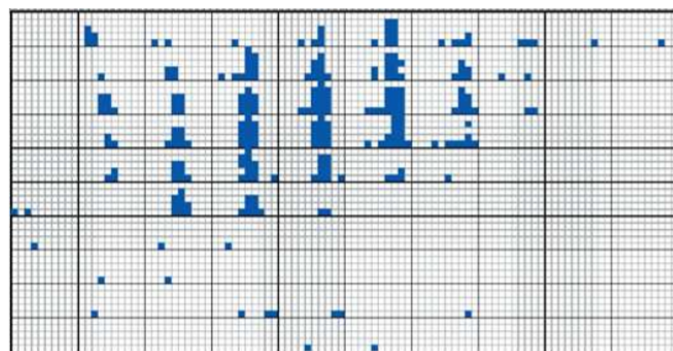


Figura 2.27: Exemplo de um *Stacked display*, retirado de [23].

2.3.3 Técnicas de Interação

Uma parte importante da exploração de dados, em conjunto com as técnicas de visualização, são as técnicas de interação. Estas permitem ao utilizador interagir com os dados e alterar dinamicamente as visualizações de modo a se adequarem aos objetivos da exploração. As técnicas de interação permitem também relacionar várias visualizações independentes [23].

As técnicas de interação podem ser categorizadas com base no efeito que têm na visualização. Técnicas de navegação permitem modificar a projeção de dados no monitor. Os métodos de melhoria de visualização permitem ajustar o nível de detalhe dos dados, seja em parte da visualização ou no todo, e permitem destacar uma porção dos dados. Técnicas de seleção dão ao utilizador a capacidade de isolar um conjunto de dados de interesse para realizar outras operações como *highlighting*, filtragem e análise quantitativa [23].

2.3.3.1 Projeção Dinâmica

Projeção dinâmica é uma operação de navegação automática, que permite mudar as projeções dinamicamente de modo a melhor explorar um *dataset* multidimensional. Um exemplo de uma projeção dinâmica é o sistema *GrandTour* [44] que tem como objetivo mostrar todas as projeções bidimensionais que sejam interessantes para a exploração do *dataset*, utilizando uma série de *scatter plots* [23].

2.3.3.2 Filtragem Interativa

Quando se exploram *datasets*, é importante segmentar os dados de forma a poder analisar *subsets* interessantes. No entanto quando se trata de *datasets* muito grandes, é difícil procurar e selecionar *subsets* de dados interessantes, e *querying* não produz, normalmente, os resultados pretendidos. A filtragem interativa é uma combinação de técnicas de seleção e melhoramento de visualização interativa que permite uma melhor filtragem dos dados. Um exemplo destas técnicas é o *Magic Lens* [45, 46], que funciona como uma lupa que filtra os dados diretamente na visualização. Estes são processados e são depois visualizados de uma forma diferente do resto do *dataset*, que não é afetado [23].

2.3.3.3 Zooming

Zooming é uma técnica de modificação de visualização bastante utilizada em várias aplicações. No caso de análise e tratamento de dados de um *dataset* muito grande, é importante ter uma visão geral dos dados e, ao mesmo tempo, poder variar a resolução da visualização, de modo a poder visualizar subconjuntos de dados mais pequenos com mais detalhe. O *zooming* permite exatamente isto, pois mostra não só os objetos a uma escala maior, como altera a visualização para mostrar mais detalhes dos dados, sendo que os detalhes aumentam com o nível de zoom. Os dados podem, por exemplo, ser representados por pixels únicos com pouco zoom, e como objetos com descrições a níveis de zoom mais alto.

Um exemplo da aplicação desta técnica é o *TableLens* [47], como mostra a figura seguinte. O *TableLens* representa cada valor numérico com uma barra. Cada barra tem um pixel de largura e o comprimento é determinado pelo valor do atributo. A visualização inicial permite uma visão geral dos dados, e observar correlações e *outliers* nos dados. O zoom permite visualizar os valores com mais detalhe e de forma contextual [23].

2.3.3.4 Distorção

Distorção é uma técnica de visualização que preserva a visão geral dos dados. Permite mostrar segmentos de dados com mais detalhes, e os dados restantes são apresentados com um nível de detalhe mais baixo. Uma técnica muito utilizada é o *Graphical Fisheye View* [48, 49], normalmente são utilizadas em hierarquias e grafos, mas também aplicável a outros tipos de visualizações [23], como exemplificado na imagem 2.28, na qual é aplicada uma *fish-eye lens* a uma matriz de *scatter plots*.

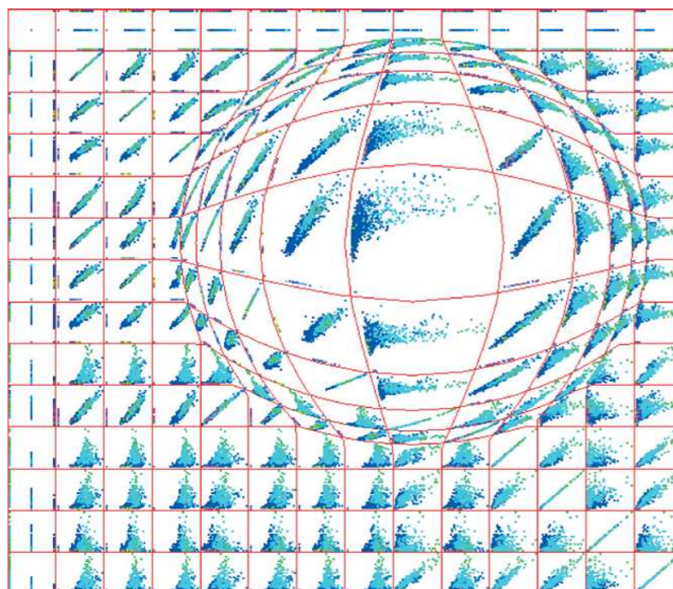


Figura 2.28: Exemplo de uma distorção de fisheye lens aplicada a uma matriz de *scatter plots*, retirada de [23].

2.3.3.5 Brushing e Linking

Brushing é uma técnica de seleção interativa normalmente acompanhada de *Linking*, um processo de comunicação dos dados selecionados entre as várias visualizações do *dataset*. O objetivo do *brushing* e *linking* é ultrapassar as fraquezas individuais de cada tipo de visualização. Por exemplo, *scatter plots* de projeções diferentes podem ser combinados com *brushing* e *linking*, através do uso de cor para realçar padrões em *subsets* nas diferentes projeções. Este exemplo pode ser também aplicado a outros tipos de visualizações. Os *datapoints* que são *brushed* ficam realçados em todas as visualizações, o que permite a detecção de dependências e correlações. Alterações feitas numa visualização são automaticamente

aplicadas às outras visualizações. *Linking* e *brushing* interativo permite, portanto, extrair mais informação do que a análise individual de cada uma das visualizações. O *linking* e *brushing* são normalmente utilizados para combinar vários *scatter plots*, gráficos de barras, gráficos de coordenadas paralelas, *dense pixel displays* e mapas [23].

2.4 Dashboard

Os dashboards são painéis que permitem ao utilizador visualizar métricas e valores importantes de forma a compreender e tomar decisões com base nos dados visualizados[50]. Os dashboards surgiram como uma ferramenta para apoio a empresas que trabalham com um grande volume de dados, para permitir que os vários setores consigam monitorizar os resultados, com recurso a *key performance indicators* (KPIs) e, com base nestes, tomar decisões [50]. O objetivo de um dashboard é apresentar informação numa forma que seja fácil e rápida de interpretar [30]. De seguida vão ser abordados algumas soluções existentes no mercado de plataformas para a criação de dashboards.

Estas soluções foram escolhidas de acordo com as funcionalidades disponíveis e com recurso ao *Gartner Magic Quadrant* para *business intelligence* (BI). O *Gartner Magic Quadrant* classifica o posicionamento das soluções existentes, num mercado que é altamente competitivo, em 4 quadrantes distintos:

- Os *Leaders* são boas soluções, bem posicionadas para evoluírem no futuro.
- Os *Visionaries* têm um bom entendimento do mercado ou uma visão para mudar o mercado, mas não executam, ainda, tão bem as suas soluções.
- Os *Niche Players* focam-se num pequeno segmento com sucesso, ou não inovam nem superam a performance da competição.
- Os *Challengers* executam a sua solução bem ou dominam um grande segmento, mas não possuem uma boa visão do futuro e direção do mercado.

A Figura 2.29 apresenta as plataformas de visualização de dados para *Business Intelligence*, distribuídas entre os 4 quadrantes.

Tendo em conta a classificação do *Gartner Magic Chart*, serão abordadas as quatro plataformas líderes (Tableau, Microsoft Power BI, ThoughtSpot e Qlikview) e ainda outras duas alternativas *open source*, o Grafana e o Kibana.

2.4.1 Microsoft Power BI

O Microsoft Power BI é uma plataforma *cloud-based* de análise de dados de BI, que fornece uma visão geral dos dados importantes. Tem versão *Desktop* e *Mobile*, que permite a visualizar os dados em qualquer lugar [51].

De acordo com a análise do website BetterBuys [51], as funcionalidades principais do Microsoft Power BI são:



Figura 2.29: Gartner Magic Chart para plataformas de Business Intelligence

- Datasets: Permite a conexão de várias fontes de dados, reunindo-os numa só plataforma, tornando mais fácil o acesso e análise de todos os dados de uma empresa, o que por sua vez, torna mais fácil a identificação de tendências e padrões nos dados, e a tomada de decisão.
- Dashboards: Dashboards compostos por painéis interativos que, por sua vez, contêm displays chamados *tiles* que fazem correlação entre os dados das diferentes fontes. Cada dashboard está ligado a um *dataset* e a plataforma permite a partilha destes dashboards.
- Relatórios: Permite a criação e partilha de relatórios sobre a informação recolhida dos dados.

Possui ainda funcionalidades de suporte às três funcionalidades principais, como um painel de navegação que permite navegar entre datasets, dashboards e relatórios, uma caixa de perguntas que permite fazer *querying* dos dados com perguntas, botões de ajuda e *feedback* e um *launcher* de aplicações do Office 365 integrado. No entanto, como esta

plataforma está ainda em desenvolvimento, possui algumas falhas, nomeadamente na publicação de relatórios e na função de partilha [51].

2.4.2 ThoughtSpot

O ThoughtSpot é uma plataforma de análise de dados que possui uma caixa de pesquisa no qual os utilizadores podem escrever uma pergunta, à qual a plataforma apresenta respostas relevantes e apresenta as melhores visualizações para análise dos dados, criação de *dashboards* e relatórios, com recurso a um motor de inteligência artificial chamado SpotIQ. O ThoughtSpot é capaz de receber e combinar dados de várias origens, desde cloud, big data ou *desktop data*. Utiliza AI para descobrir padrões, tendências e novos conhecimentos sobre os dados [52].

De acordo com a análise do website BetterBuys [52], as funcionalidades principais do ThoughtSpot são:

- Motor de Busca Relacional: Utiliza as pesquisas para construir a visualização dos dados.
- Motor de Inteligência Artificial SpotIQ: permite executar vários algoritmos de análise de dados em grandes quantidades de dados e descobrir novos conhecimentos de interesse, e sugere novas pesquisas de acordo com as últimas pesquisas realizadas.
- Servidor de Visualização/BI: permite gerar queries utilizando linguagem natural, criação automática e personalizada de visualizações interativas e a criação de modelos de metadata partilhados entre vários *datasets* que tornam as respostas das pesquisas consistentes em todos os *datasets* que partilham o mesmo modelo.
- Motor de cálculo em memória: permite analisar grandes volumes de dados em segundos, não requer otimização de dados para melhorar a performance, o que permite manter a fidelidade dos dados e reunir todos os dados da empresa numa única cache dedicada, de modo a aumentar a velocidade das queries.
- *Distributed Cluster Manager*: Permite escalar a quantidade de dados sem comprometer a performance, impede a perda de dados e realiza operações de auto-reparação automaticamente.
- Segurança e Administração Empresarial: Permissões adequadas a cada utilizador, de modo que só os utilizadores autorizados têm acesso a certos *datasets*. Administração centralizada dos dados. Capacidade de rastrear as origens dos dados, garantindo a fidelidade dos dados.
- Ligações de dados: Permite aceder a dados *on-premise* ou na cloud e a capacidade de realizar queries em tempo real.
- Análise de dados com Busca Relacional que utiliza e visualizações *embedded*.

- Mobile: Aplicação mobile com todas as funcionalidades da solução Desktop.

Um contra do ThoughtSpot é a sua curva de aprendizagem inicial elevada, devido ao motor de busca e ao motor de inteligência artificial [52].

2.4.3 Tableau

O Tableau é uma plataforma de visualização de dados de *business intelligence* que fornece ferramentas intuitivas que tem como objetivo facilitar a compreensão dos dados. Possui funcionalidades *drag-and-drop* que ajudam os utilizadores na análise de dados, criação de visualizações e relatórios e partilha de conhecimento. Oferece soluções Desktop, Online, via cloud, e Mobile [53].

De acordo com a análise do website BetterBuys [53], as funcionalidades principais do Tableau são:

- Várias ligações de dados: Permite a ligação a várias fontes de dados sem necessidade de programação.
- Dados em memória ou em tempo real: Permite receber os dados armazenados em memória ou em tempo real, com *refresh* automático e avisos de falhas de ligação às fontes de dados.
- Colaboração segura: Partilha de conhecimento e colaboração através do Tableau Server ou Tableau Online.
- Otimização para Mobile: Permite a criação e personalização de dashboards adaptados para computador, telemóvel e tablet.
- Dashboards *embedded*: Permite a criação de dashboards embutidos em outras aplicações.
- Integração *drag-and-drop*: Permite criar e integrar visualizações com propriedades *drag-and-drop*.

O Tableau tem como contras uma curva de aprendizagem elevada e o preço, considerado demasiado alto para pequenas empresas [53].

2.4.4 Qlik

O Qlikview é uma plataforma de análise e visualização de dados para *business intelligence* que oferece um motor de inferência único em relação à concorrência, pois faz as associações de dados automaticamente, em vez de utilizar ferramentas de query. Possui um motor de busca que devolve resultados para a pesquisa, assim como resultados relacionados com a pesquisa. A visualização do Qlikview é simples e intuitiva. Os dados

relevantes, relações entre dados, padrões e tendências são coloridos, de modo a ficarem realçados, enquanto que os dados irrelevantes são apresentados a cinzento [54].

De acordo com a análise dos websites BetterBuys [54] e GetApp [55], as funcionalidades principais do Qlikview são:

- Integração de análise de dados, permite calculos avançados com Python e outras ferramentas externas.
- Dashboards interativos e aplicações dinâmicas.
- Capacidade de integração em APIs e Aplicações.
- Espaço de trabalho colaborativo e ferramentas de colaboração.
- Permite colaboração segura em tempo-real.
- União de todas as fontes de dados numa só plataforma.
- Análise de dados em tempo real.
- Permissões de acesso aos dados de acordo com a função.
- Criação e personalização de relatórios.
- Motor de busca.

Segundo alguns utilizadores, existem algumas dificuldades com a ferramenta de filtragem de dados [54].

A Qlik oferece ainda outra opção de ferramenta de visualização de dados, o Qlik Sense. O Qlik Sense é uma ferramenta mais recente desenvolvida pela Qlik, e uma evolução do QlikView, que tal como este, é uma ferramenta de visualização de dados que oferece a simplificação da exploração dos dados, o que permite a sua utilização a utilizadores que não têm experiência na área de *data analytics*, com recurso a ferramentas de *drag and drop* e que permitem explorar os dados em mais detalhe. A sua API (*application program interface*) permite a integração dos seus *dashboards* na aplicações do utilizador [56].

Algumas das principais funcionalidades, listadas pelo website BetterBuys [56] são apresentadas em seguida:

- *Insight Advisor* - Esta funcionalidade utiliza machine learning para permitir que, utilizando linguagem natural, sejam feitas perguntas que devolvem respostas com novos conhecimentos sobre os dados.
- *Self-service creation* - Criação de *dashboards* com recurso a funcionalidade *drag and drop*, sem utilizar scripts, queries ou joins.
- *Centralized management* - O Qlik Sense é uma central para o desenvolvimento de apps e conhecimento sobre os dados.

- *Data Integration* - O Qlik Sense permite a unificação de diferentes tipos de dados na plataforma, e não esconde os dados irrelevantes, em vez disso, estes são mostrados a cinzento, de modo a que sejam possíveis detetar tendências nestes.
- *Interactive Visualizations* - O Qlik Sense oferece a criação de visualizações que responde à mudança de dimensão e do contexto dos dados.
- *Data Storytelling* - Ilustra vários pontos de vista diferentes dos dados, atribuindo-lhes um contexto. Apresenta também a análise original de modo a fazer uma análise em mais detalhe e alterar o contexto dos dados.
- *Responsive Design* - O design adapta-se ao meio onde é utilizado, seja em computador, tablet ou smartphone.

Alguns utilizadores queixam-se da personalização limitada, que só pode ser ultrapassada com recurso a aplicações *third-party*, não oficiais [56].

2.4.5 Kibana

O Kibana é uma plataforma *open source* de visualização cuja funcionalidade principal é a análise e *query* de dados. Permite a pesquisa de dados para análise e diagnóstico. Com base nestas pesquisas, o Kibana fornece uma visualização dos dados de várias formas, com recurso a tabelas, gráficos, mapas e outros tipos de visualizações. Permite a criação de *dashboards* interativos e versáteis, com capacidade de filtrar dados na visualização e editar o *dashboard*. É primariamente utilizado para análise de mensagens de registo. O Kibana foi desenvolvido para funcionar apenas com Elasticsearch, um motor de busca e análise de dados, e como tal, não suporta outras fontes de dados [57].

2.4.6 Grafana

O Grafana é uma plataforma *open source* de visualização de dados baseado no Kibana que surgiu com o objetivo de dar suporte para métricas. Oferece uma fácil criação e edição de *dashboards*.

O Grafana foi desenvolvido para análise e visualização de métricas e portanto não permite querying de texto. No entanto, funciona com várias fontes de dados, ao contrário do Kibana.

O Grafana possui um sistema de controlo e autenticação que permite restringir o acesso aos dados e aos *dashboards*. Possui também um editor de queries para cada fonte de dados, com a sintaxe de cada editor a variar com a fonte de dados.

Os *dashboards* do Grafana são compostos por painéis, que o utilizador pode arranjar da maneira que pretender, e cada painel pode conter fontes de dados diferentes. O Grafana possui vários tipos de visualizações, como grafos, tabelas, mapas de calor, gráficos e muitos mais. Possui também um sistema de alerta que permite aos utilizadores definirem condições que, quando cumpridas, fazem disparar o alarme que gera uma notificação [57].

2.5 Benchmarks

Tendo abordado os prós e os contras de cada uma das plataformas, queremos agora comparar as plataformas, através de alguns requisitos que julgo necessários para a realização das plataformas de visualização de dados. Estes requisitos são os seguintes:

- Suporte ao tipo de base de dados utilizado. No nosso caso são utilizadas bases de dados PostgreSQL, e é necessário garantir que a plataforma utilizada seja compatível com este tipo de dados.
- Permitir a criação de visualizações utilizando as várias técnicas de visualização de dados estudadas.
- Permitir a utilização de tipos de visualização open source personalizados.
- Permitir a criação de *dashboards* interativos, que permitam uma visualização em maior detalhe dos dados apresentados.
- Permitir a integração das visualizações em websites, de modo a ser integrados numa plataforma web criada para englobar os dashboards com as visualizações criadas.
- Por ultimo mas não menos importante, é importante comparar os preços da cada um dos softwares, junto com os restantes requisitos, de modo a escolher a opção mais adequada e com menores custos. Verificar também a existência ou não de soluções grátis ou mais acessíveis para cada plataforma.

Tendo em conta estes requisitos, é apresentada de seguida uma tabela que apresenta as 6 soluções estudadas e os requisitos que cada uma cumpre, com base no que já foi estudado sobre cada uma, assim como a comparação realizada pela CSG PRO [58].

Tabela 2.1: Tabela dos requisitos necessários que apresenta quais os requisitos que cada solução cumpre.

Requisito	Power BI	Tableau	Qlik Sense	ThoughtSpot	Grafana	Kibana
Ambiente de Desenvolvimento	Desktop	Desktop	Web Browser	Implementação	Web Browser	Web Browser
Suporte de Base de Dados	✓	✓	✓	✓	✓	✗
Criação de Visualizações	✓	✓	✓	✓	✓	✓
Suporte de Visualizações open source	✓	✗	✓	✗	✓	✓
Criação de dashboards interativos	✓	✓	✓	✓	✓	✓
Integração das visualizações	✓	✓	✓	✓	✓	✓
Versão grátis com <i>full-features</i>	✗	Ferramenta separada	Ferramenta separada	✗	Open Source	Open Source
Preços	Pro (9.99\$/mês) Premium (20\$/mês)	Creator (70\$/mês) Explorer (35\$/mês) Viewer (15\$/mês)	Pago (Preços não revelados)	Pago com base na capacidade de dados, em incrementos de 250GB. Preços não revelados	Open Source	Open Source

DESENVOLVIMENTO DA PLATAFORMA

Este capítulo aborda o desenvolvimento da plataforma de visualização, desde a análise dos dados e escolha das métricas, até criação das visualizações, com recurso à plataforma escolhida, e implementação das visualizações na plataforma Web criada para apresentar estas visualizações sobre a forma de *dashboards* interativos.

3.1 Definição da Plataforma

Tendo em conta as plataformas mencionadas na secção 2.4, analisados os prós e contras de cada uma, e com o benchmark da Tabela 2.1, todas as plataformas apresentadas, à exceção do Kibana, que não suporta a base de dados utilizada, poderiam ser utilizadas para a criação dos *dashboards*. No entanto, a plataforma escolhida para o desenvolvimento das visualizações foi o Grafana. Tendo em conta que o Grafana é uma plataforma *open source*, que é versátil nas fontes de dados que aceita e é compatível com PostgreSQL, o formato da fonte de dados utilizada, é utilizado para visualização de métricas e tendo em conta a fácil criação e edição de diferentes tipos de visualizações, foi escolhida esta plataforma para a implementação dos *dashboards*.

Esta escolha teve em conta o *benchmark* da Tabela 2.1, assim como o estudo feito a cada uma das plataformas existentes, e reflete o facto de o Grafana possuir as funcionalidades necessárias para a criação dos *dashboards* para a plataforma, sendo elas a sua utilização ser simples e intuitiva, possuir suporte a dados temporais, permitir a criação de visualizações personalizadas e *plugins* de visualizações, e permitir a exportação das visualizações para fazer *embed* na plataforma web.

Embora as outras soluções tenham funcionalidades extra, quando comparadas com o Grafana, o Grafana é a única plataforma, dentro das plataformas estudadas, que é *open*

source e suporta a base de dados utilizada, que se traduz em menos custos no desenvolvimento e implementação.

3.2 Arquitetura Conceptual

A arquitetura conceptual da plataforma é apresentada na Figura 3.1. Temos os dados armazenados numa base de dados SQL, nos quais são feitas *queries* pela plataforma Grafana e *views - queries* que ficam armazenadas na base de dados e apresentam os resultados numa nova tabela virtual, que permitem associar dados, fazer *queries* mais complexas e possibilitam análises de dados de tabelas diferentes com o benefício de tornar as *queries* no lado da plataforma mais simples - criadas na base de dados. Estas alimentam a plataforma Grafana com dados, de modo a criar as visualizações que queremos.

Depois de criadas as visualizações, estas são exportadas para a plataforma Web criada para a visualização dos dados pelos utilizadores finais, ou seja, as empresas e os seus funcionários. Nesta plataforma, as visualizações estarão divididas em vários *dashboards* diferentes, de modo a que estas estejam organizadas de acordo com os processos de produção.

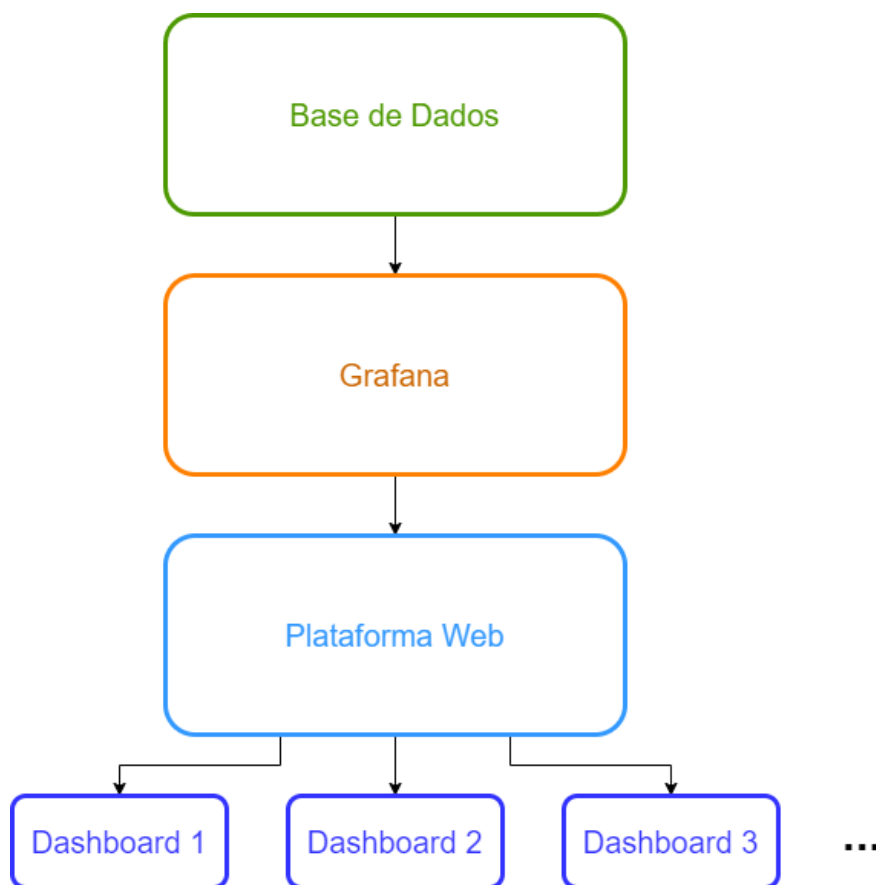


Figura 3.1: Arquitetura conceptual da plataforma de visualização interativa de dados.

3.3 Caso de Estudo Volkswagen Autoeuropa

Esta secção aborda o desenvolvimento da plataforma para o caso de estudo da Volkswagen Autoeuropa, começando pela análise dos dados disponíveis, seleção de métricas, criação das visualizações e implementação das visualizações na plataforma web do projeto Boost 4.0.

3.3.1 Análise de Dados e Seleção das Métricas

Após definida a plataforma a ser utilizada, foram analisados os dados do processo intra-logístico das baterias de forma a escolher as métricas relevantes para os objetivos do projeto.

Olhando para a forma como estão estruturados os dados, como mostra a figura 3.2, podemos retirar algumas informações. Temos a tabela *part*, que contém os dados do número e família das peças, e a tabela *part family*, que contém os dados sobre cada família de peças, o nome das famílias e as peças que pertencem a cada família. Estas duas tabelas, em conjunto, dão-nos todas as informações sobre cada peça individual utilizada na fábrica. A tabela *daily stock* contém os dados de stock em armazém de paletes de cada tipo de peça, por número de peça, sendo que os valores são amostrados de uma em uma hora. As tabelas *unload operation* e *transport* contêm os dados de transporte das peças dentro da fábrica. A tabela *order part* contém os dados de encomendas de peças, sendo que a tabela *order* contém o número e data da encomenda e a tabela *supplier* o número e nome dos vendedores das peças encomendadas.

Trabalhámos sobre duas versões da base de dados. A primeira apenas com os dados das baterias. Esta base de dados contém os dados de stock, recepção (*receiving*) e sequenciação das baterias utilizadas na fábrica da Volkswagen Autoeuropa. Esta base de dados possui dados desde o dia 1 de Janeiro de 2018 até 27 de Março de 2018 para a tabela do *stock*, e desde o dia 2 de Janeiro de 2018 até 21 de Junho de 2019, para a tabela de encomendas. A segunda versão da base de dados possui já os dados para todos os tipos de peças com as quais a fábrica trabalha. No entanto, a amostra é mais pequena, visto que possui apenas dados para o mês de Janeiro de 2020.

O objetivo do projeto Boost 4.0, para este caso de estudo, é aumentar a eficiência dos processos intra-logísticos na fábrica da Volkswagen Autoeuropa e, com isso, reduzir também os custos inerentes. Como tal, uma das métricas que é necessário estudar é o número de peças armazenadas nos armazéns da fábrica. Para garantir o normal funcionamento da fábrica, é necessário garantir que existam em armazém um número de peças aproximadamente 2.5 vezes o número médio de cada tipo de peças utilizadas num dia de funcionamento da fábrica, para que em caso de falha na entrega de uma encomenda, a fábrica não seja obrigada a parar a produção. No entanto, ter em armazém muito mais peças do que esse número é desnecessário e aumenta bastante os custos para a fábrica. Queremos então visualizar esta métrica para perceber o seu comportamento, com o objetivo de

normalizar o stock de peças em torno da margem de segurança sem que este número aumente muito acima da margem, sem, no entanto, descer abaixo da margem de segurança. Para fazer isso, é preciso identificar de forma estável os dados estruturados e como se pode fazer a correlação entre os dados de forma a extrair informação que pretendemos.

Como este *dataset* possui bastantes tabelas, cada uma com vários parâmetros, iremos nos focar apenas nas tabelas de *order*, *order_part*, *part*, *part_family*, *daily_stock* e *seq_task*, visto que estas tabelas possuem os dados que necessitamos de visualizar de forma a realizar os objetivos do projeto.

Cada uma das tabelas da base de dados é composta por várias colunas com os parâmetros, e as linhas correspondem a cada entrada dos dados, em formato CSV. A Figura 3.3 mostra um exemplo de uma destas tabelas, a tabela de *hourly stock*.

3.3.2 Criação do dashboard e visualizações

As visualizações foram criadas com a plataforma Grafana, sendo que os dados foram extraídos de uma base de dados PostgreSQL, que possui todos os dados disponibilizados, desde os dados das peças e stock das mesmas até encomendas e sequencing, com recurso a queries SQL à base de dados. A listagem 3.1, na página 56, ilustra um exemplo de query utilizada. Esta tabela possui 6 colunas, a primeira representa a data e hora dos dados, enquanto que as restantes 5 representam os 5 tipos de baterias utilizadas na produção de automóveis, na fábrica da Volkswagen Autoeuropa. Cada linha representa o número de paletes para cada um dos tipos de baterias a cada data e hora dadas pela coluna *start_datetime*, sendo que o número de baterias por palete é dada pelo parâmetro *quantity_per_europallet* da tabela *battery_metadata*: 54 para as baterias do tipo 1S0915105A, 48 para as baterias dos tipos 6R0915105B, 5TA915105B e 7P0915105 e 36 para as baterias do tipo 7P0915105A.

De forma a melhor organizar as visualizações, o *dashboard* foi dividido em 3 *dashboards* separados, estando cada um relacionado com um processo diferente na fábrica: o *dashboard Receiving* possui os dados da chegada das encomendas à fábrica, o *dashboard Warehousing* possui dados das peças em stock e o *dashboard Sequencing*, que apresenta os dados do processo de sequenciação das peças, antes de serem transportadas para o PoF. Para o *dashboard* do processo de *Warehousing*, foram criadas algumas visualizações que nos permitem ver o stock por famílias de peças e por número de peça, dentro de cada família, tornando a visualização bastante simples, sendo que basta introduzir numa caixa de pesquisa a família da peça e em outra caixa o número da peça para obter as visualizações apenas para essas peças.

3.3.2.1 Dashboard Receiving

Para o *dashboard* do processo de *receiving* (processo descrito no subcapítulo 1.2, e na Figura 1.1), o objetivo é visualizar e detetar os padrões nos dados das encomendas. Para tal, foram criadas algumas visualizações que ilustram os número das encomendas de

3.3. CASO DE ESTUDO VOLKSWAGEN AUTOEUROPA

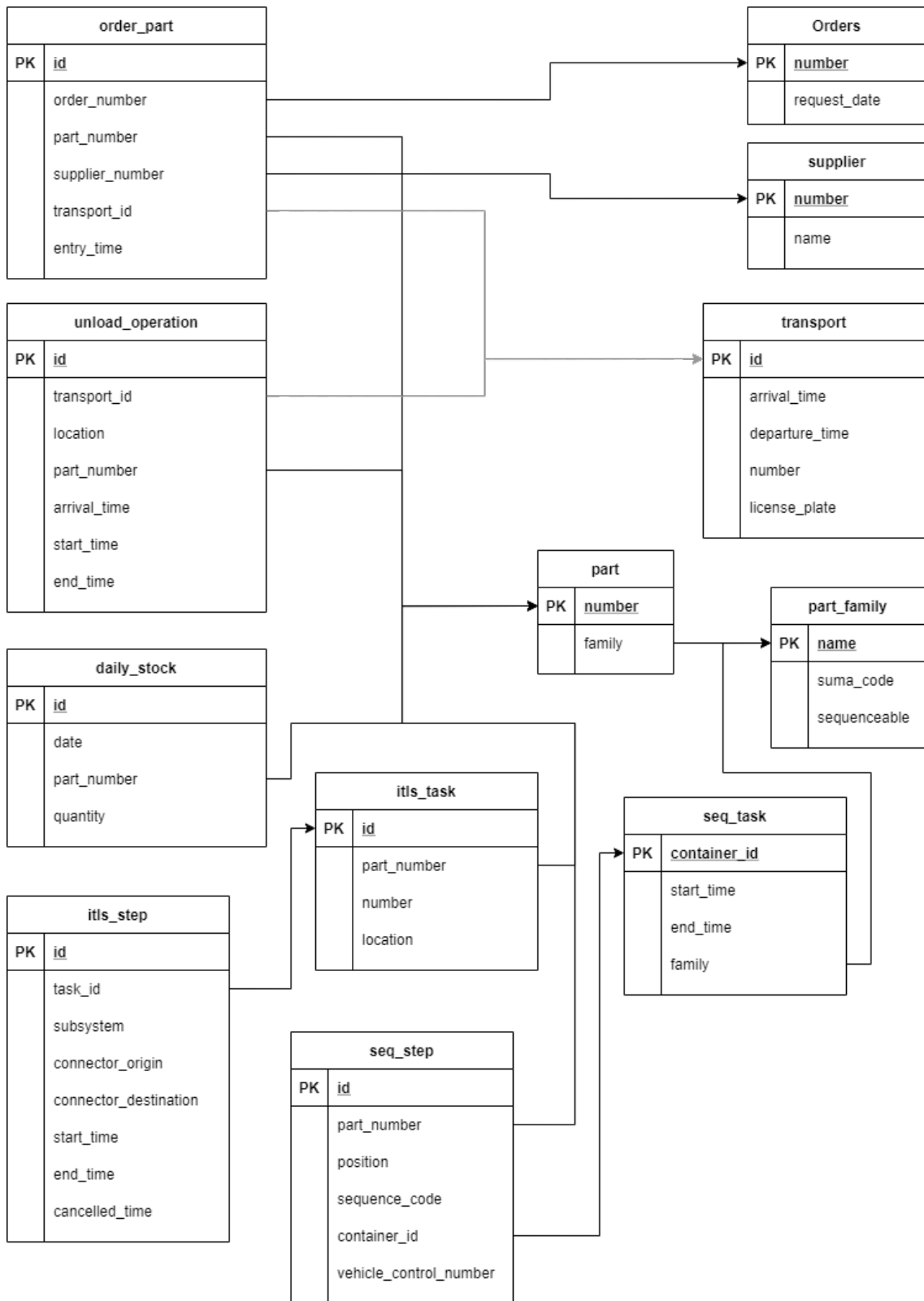


Figura 3.2: Diagrama de entidades e relações da base de dados da Volkswagen Autoeuropa.

CAPÍTULO 3. DESENVOLVIMENTO DA PLATAFORMA

	start_datetime timestamp without time zone	7P0915105A integer	7P0915105 integer	6R0915105B integer	1S0915105A integer	5TA915105B integer
1	2018-01-01 23:00:00		4	2	10	52
2	2018-01-01 23:00:00		4	2	10	52
3	2018-01-02 00:00:00		4	2	10	52
4	2018-01-02 01:00:00		4	2	10	52
5	2018-01-02 02:00:00		4	2	10	52
6	2018-01-02 03:00:00		4	2	10	52
7	2018-01-02 04:00:00		4	2	10	52
8	2018-01-02 05:00:00		4	2	10	52
9	2018-01-02 06:00:00		4	2	10	52
10	2018-01-02 07:00:00		4	2	10	52
11	2018-01-02 08:00:00		4	2	10	52
12	2018-01-02 09:00:00		4	2	10	51
13	2018-01-02 10:00:00		4	2	9	51
14	2018-01-02 11:00:00		4	2	9	50
15	2018-01-02 12:00:00		4	2	9	50

Figura 3.3: Tabela de *hourly_stock*, em formato CSV. Cada coluna representa o tipo de bateria utilizada, enquanto que as linhas representam o número de paletes para cada tipo de bateria, para a data e hora dada na coluna *start_datetime*.

baterias feitas pela fábrica. A primeira visualização, Figura 3.4, apresenta o número de paletes de baterias encomendadas, para cada tipo de bateria utilizada pela fábrica.

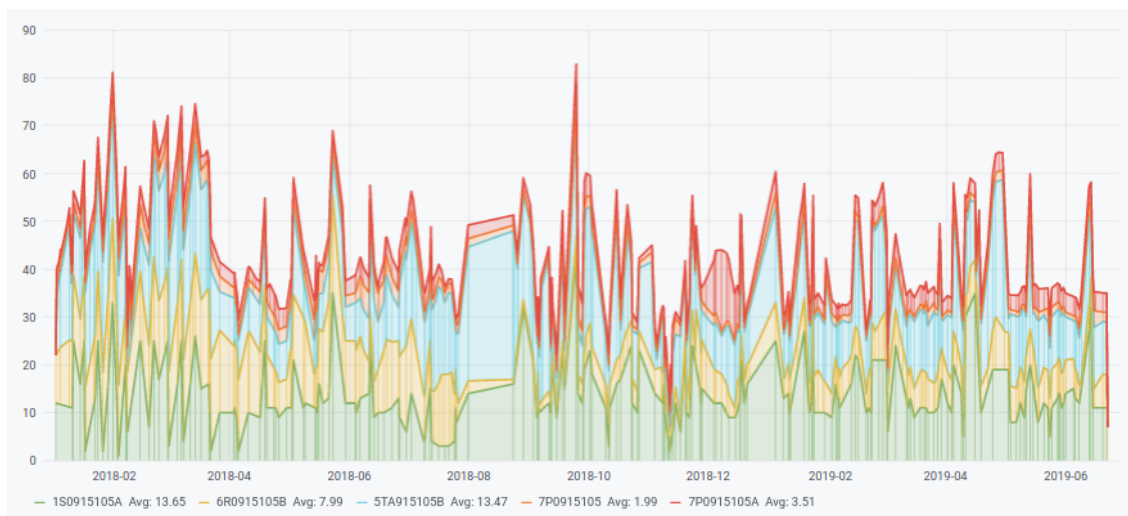


Figura 3.4: Número de paletes de baterias encomendadas, dividido por número de peça de cada tipo de bateria.

A visualização da Figura 3.5 ilustra o número total de baterias encomendadas em percentagem para a janela de tempo abrangida pelos dados, dando uma noção ainda mais clara de quais as baterias são mais requisitadas para a produção, que necessitam de um número mais elevado de encomendas, e quais são menos utilizadas, e portanto, a necessidade de encomendas destes tipos de bateria é menor.

A Figura 3.6 ilustra o gráfico do número de paletes recebidas de cada fornecedor de baterias, ao longo da janela de tempo para a qual os dados estão disponíveis (Janeiro de 2018 a Junho de 2019). Este gráfico mostra quais as empresas forneceram baterias à fábrica, e em que quantidades o fizeram, ao longo da janela de tempo dos dados.

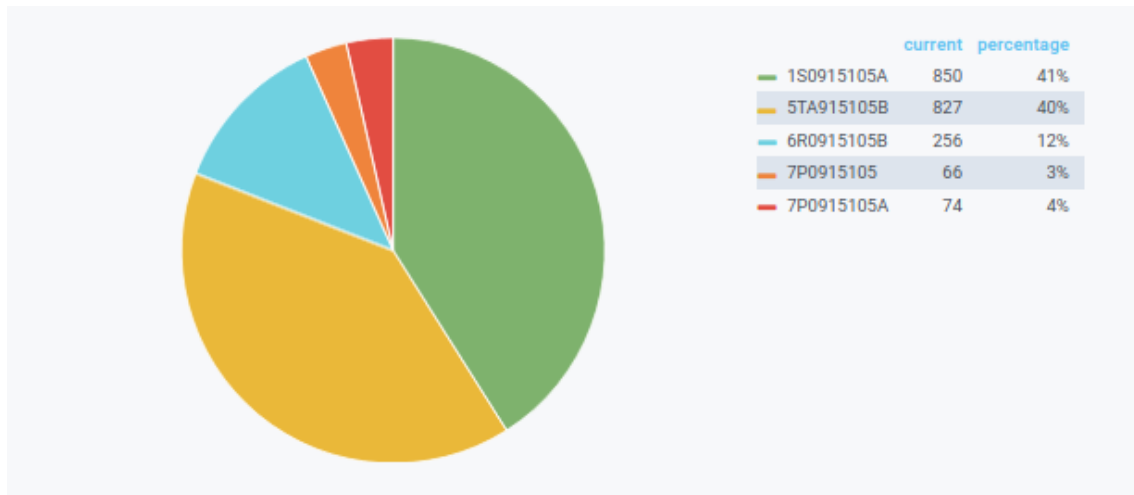


Figura 3.5: Número total de encomendas de paletes de baterias em percentagem para cada tipo de bateria utilizada na fábrica.

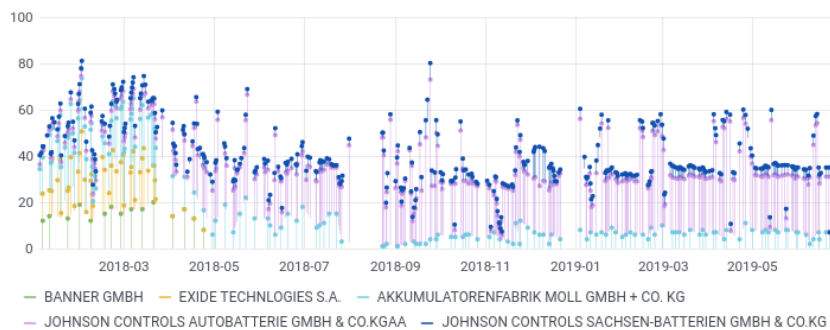


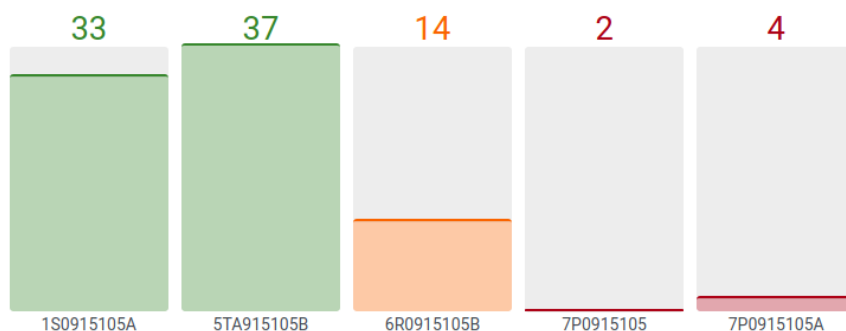
Figura 3.6: Gráfico do número de baterias encomendadas a cada fornecedor, ao longo da janela de tempo dos dados.

3.3.2.2 Dashboard Warehousing

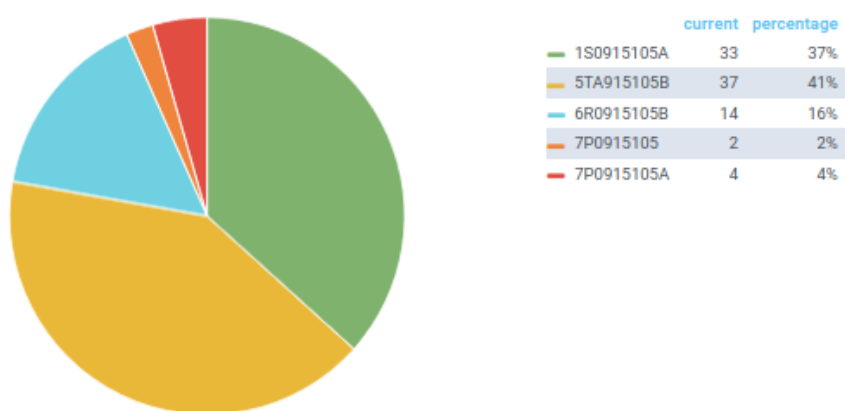
Para o *dashboard* do processo de *warehousing* (processo descrito no subcapítulo 1.2, e na Figura 1.1), foram desenvolvidas visualizações que ilustram o stock de peças no armazém, em específico o stock das baterias, ao longo de um período de tempo, assim como o stock atual, e as operações de transporte das baterias do armazém para os pontos onde serão utilizadas.

Começando por algumas visualizações simples, foram feitas para o *dashboard Warehousing* duas visualizações para o stock das baterias, no *dashboard Warehousing*, que ilustram o stock atual de paletes no armazém. As figuras ilustram o stock absoluto, através de um gráfico de barras (Figura 3.7(a)) e em percentagem com recurso a um *pie chart* (Figura 3.7(b)).

Foi também criada uma visualização que ilustra o estado do stock de cada tipo de bateria, em que cada linha representa o stock das baterias ao longo do tempo, utilizando um código de cores para as quantidades, sendo que o vermelho representa os valores



(a) Gráfico de barras



(b) Pie chart com percentagem

Figura 3.7: Gráficos para o stock atual de contêntores de baterias, para cada tipo de bateria.

mais baixos de stock e o verde os valores mais altos, com a cor amarela a representar o valor intermédio. A escala destes valores está representada no canto inferior esquerdo da Figura 3.8.

Outra visualização possível para o stock das baterias é apresentada na Figura 3.9, sob a forma de um gráfico de linha, em que cada uma das linhas representa um tipo de bateria. No eixo x temos a linha temporal dos dados do stock, enquanto de o eixo y representa o número de paletes de baterias. É possível identificar quais as baterias mais e menos utilizadas pelo stock de cada tipo de bateria. As baterias que apresentam os números mais baixos, de números de peça 7P0915105 e 7P0915105A, são as baterias *low runner*, que são utilizadas em carros com um volume de produção mais baixo. As baterias que apresentam os números mais elevados, de números de peça 1S0915105A, 5TA915105B e 6R0915105B, com números mais elevados, são consideradas baterias *high runners*, pois são utilizadas em carros com grandes volumes de produção, o que se reflete no stock destas baterias. A utilização das baterias nos carros construídos na Volkswagen Autoeuropa depende do modelo do carro - T-Roc tem um volume de produção mais elevado que os outros dois modelos produzidos na fábrica (Alhambra e Sharan) -, da potência do motor, do tipo de

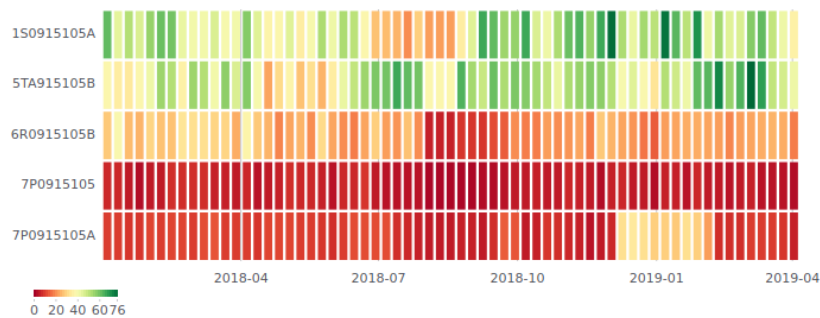


Figura 3.8: Gráfico do estado do stock de cada tipo de bateria. Este gráfico, no qual cada linha representa o stock de uma peça ao longo do tempo, utiliza um código de cor para identificar o número de paletes em stock para cada tipo de peça.

caixa de velocidades e de outros extras.

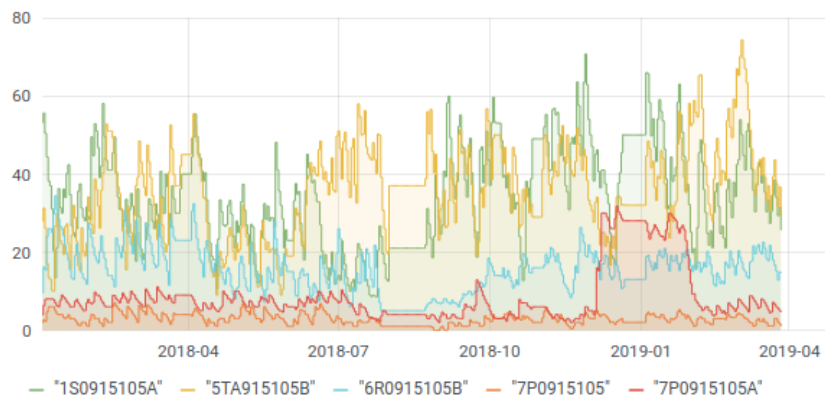


Figura 3.9: Gráfico do stock de paletes de baterias, para cada tipo de bateria.

Visto que um dos objetivos é otimizar o stock das peças, tal como mencionado na subsecção 3.3.1, e tendo a estrutura da base de dados em conta, queremos então visualizar os dados das tabelas de *daily stock* (tabela com os dados diários de stock para cada número de peça) e *order part* (tabela com os dados das encomendas realizadas), e correlacionar os dados destas duas para identificar possíveis padrões de interesse nos dados.

Como tal, é feita uma *query SQL*, tal como ilustrada na listagem 3.1, que nos devolve o número de paletes em stock no armazém, para o número de peça escolhido, com recurso a uma caixa de texto com função *autocomplete*, um script *open source* desenvolvido em JavaScript, e disponibilizado na página da jQuery [59], que utiliza o *input* do utilizador numa caixa de texto para devolver apenas os resultados da pesquisa que contêm a *string* de caracteres introduzida na caixa de texto. O resultado desta *query* é depois utilizado para construir uma visualização do stock das peças. Tendo em consideração que são obtidas duas colunas de dados, a primeira com a data e a segunda com os contentores em stock, a melhor visualização é um gráfico de linha, com a data no eixo *x* e o stock de contentores no eixo *y* (Figura 3.10).

Listagem 3.1: Query SQL para o stock diário de peças

```

1 SELECT
2     $__timeGroupAlias(date, $__interval),
3     avg("$part_number") AS "$partNumber"
4 FROM daily_stock
5 WHERE $__timeFilter(date)
6 GROUP BY 1
7 ORDER BY 1

```

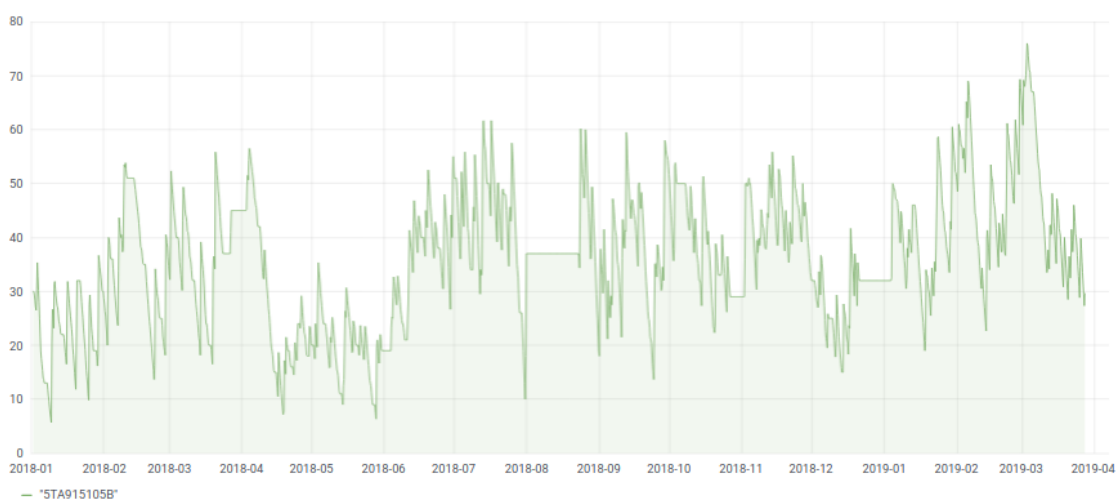


Figura 3.10: Gráfico do *stock* diário de contentores em armazém com o número de peça "5TA915105B", da família das baterias.

Com base neste gráfico, foi criada uma visualização que compara o stock das peças em armazém por hora, em comparação com a média do número de peças utilizadas por dia e com a margem de segurança de 2.5 vezes a média do número de peças utilizadas. Esta margem de segurança é dada pelo facto de que as entregas de peças ocorrem a cada 2 dias úteis, à Segunda-feira, Quarta-feira e Sexta-feira de cada semana. Para garantir que não aconteça uma situação de escassez de peças em caso de atrasos nas entregas de peças, foi definida esta margem de segurança para que a fábrica não seja forçada a parar a produção. Como o número médio de carros produzidos é de cerca de 900 carros, foi utilizado o parâmetro *production_avg*, parâmetro da tabela *battery_metadata* que representa a percentagem média de todas as baterias utilizadas nos carros fabricados, para calcular o número médio de utilização de cada tipo de bateria num dia de produção. A tabela *battery_metadata* existe numa versão mais antiga da base de dados apenas com os dados das baterias, sobre a qual foram desenvolvidas algumas das visualizações, incluindo a da figura 3.11 e a da figura 3.12 visto que apenas foi disponibilizada *metadata* para as baterias e não para as outras peças. Podemos observar pela figura 3.11 que em grande parte do tempo, a fábrica tem em stock um número de peças muito acima do necessário,

e tem poucos dias em que o stock cai para valores abaixo da margem de segurança.

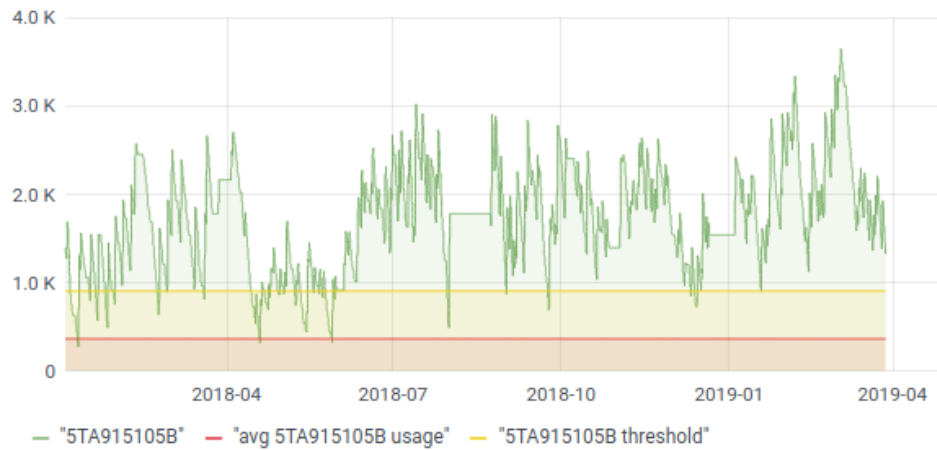


Figura 3.11: Gráfico do *stock* diário de baterias, com thresholds de utilização média (vermelho) e margem de segurança (amarelo) para a peça "5TA915105B", da família das baterias.

Tendo estes dados em conta, foi feita uma outra visualização que compara o número de peças em stock com os dados das encomendas (Figura 3.12). É possível observar através deste gráfico que são feitas um grande número de encomendas, de grandes volumes de peças, em dias nos quais o stock está acima da margem de segurança, calculada com recurso ao parâmetro *production_avg* anteriormente mencionado e ao número médio de carros produzidos diariamente, aumentando ainda mais o stock para além desta margem. Conseguimos também ver a correlação entre os dados de stock e encomendas, visto que após cada encomenda, marcadas a amarelo no gráfico, existem picos no stock correspondentes ao número de peças encomendadas.

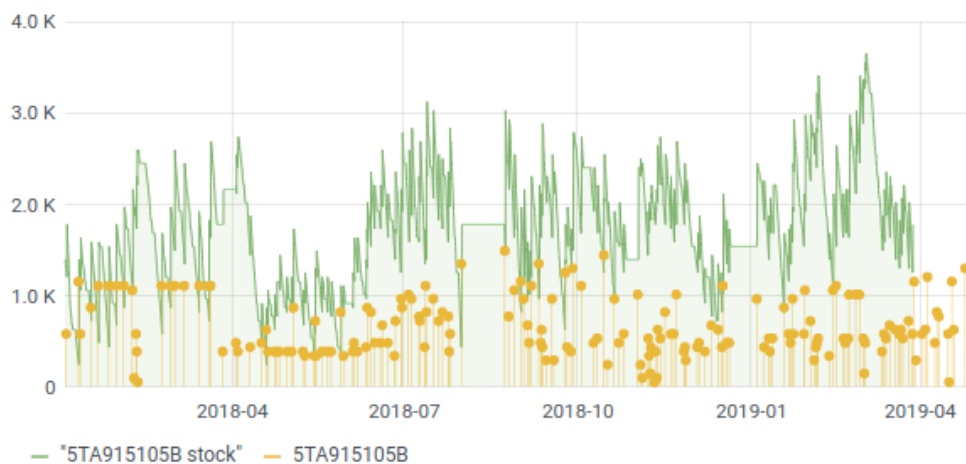
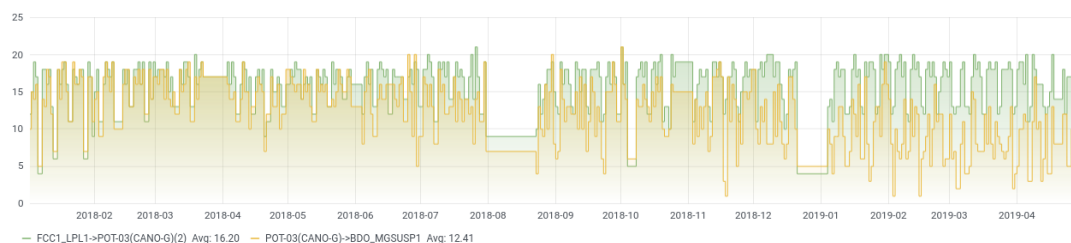
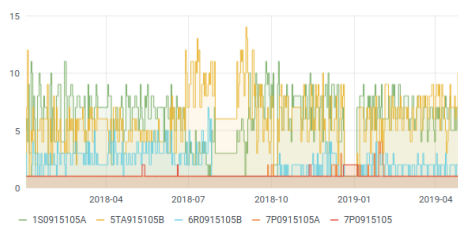


Figura 3.12: Gráfico do stock diário e encomendas de baterias para a peça "5TA915105B"

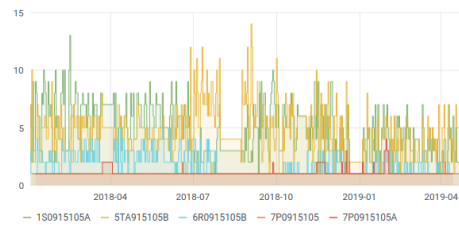
Para além do stock, foram também desenvolvidas algumas visualizações para o transporte das peças desde o armazém até vários pontos da fábrica onde as baterias serão utilizadas. O primeiro gráfico, Figura 3.13(a) detalha o número de paletes de baterias transportadas por tipo de movimento. Os dois gráficos abaixo representam o número de paletes de baterias transportadas, para cada tipo de bateria, em dois movimentos diferentes - LOZ-Banhof, do armazém para o comboio de transporte de peças, e Bahnhof-SUMA, do comboio para o local de sequenciação), nas Figuras 3.13(b) e 3.13(c) respetivamente. Tal como seria de esperar, é possível identificar nestes gráficos quais as baterias com um número de transportes mais baixo, que correspondem às baterias com o stock também mais baixo. O mesmo acontece para as baterias com stock mais elevado. Também é possível ver alguma correspondência entre estas figuras e a Figura 3.9. Significa isto que provavelmente existe correlação entre os valores de stock e de transporte de baterias.



(a) Número de paletes de baterias transportadas, por tipo de movimento



(b) Número de paletes transportadas, por tipo de bateria, no movimento LOZ-Banhof



(c) Número de paletes transportadas, por tipo de bateria, no movimento Bahnhof-SUMA

Figura 3.13: Gráficos para as operações de transporte de peças do armazém para outros pontos da fábrica.

3.3.2.3 Dashboard Sequencing

O processo de sequenciação existe com o propósito de agilizar a montagem das peças no *point of fit*. Isto é feito através da organização das peças na mesma ordem dos carros na linha de montagem, nos quais elas vão ser montadas.

Este processo possui um protocolo a ser seguido pelos operadores, no qual as peças são colocadas num carrinho de transporte seguindo a sequência dos veículos na linha de produção. Depois da peça ser sequenciada, é registada numa base de dados os dados da peça e a data e hora a que foi feita a sequenciação da mesma. Este processo é repetido

para todas as peças recebidas no SUMA - a área de sequenciação - e assim deve ser realizado para que os dados recolhidos neste processo sejam representantes da realidade do processo. Depois de concluído o processo de sequenciação, os carrinhos são transportados para o *point of fit*.

A necessidade de seguir este protocolo prende-se com o facto de a Volkswagen Autoeuropa necessitar dos dados do processo, nomeadamente os tempos de sequenciação, de modo a que seja possível fazer uma comparação entre o tempo que um operador demora a realizar o processo e o tempo que demora o mesmo processo automatizado. O incumprimento do protocolo impossibilita também a medição do tempos de sequenciação de cada peça individual.

Para o *dashboard* do processo de *sequencing*, queremos visualizar não só o número de baterias que foram sequenciadas, mas também o tempo das sequenciações que seguiram corretamente o protocolo de sequenciação da Volkswagen Autoeuropa. Primeiro que tudo, queremos visualizar o número de baterias sequenciadas. Nesse sentido, foi criada a visualização da Figura 3.14, que ilustra o número contentores de baterias sequenciadas, divididas pelo tipo de bateria em cada contentor.

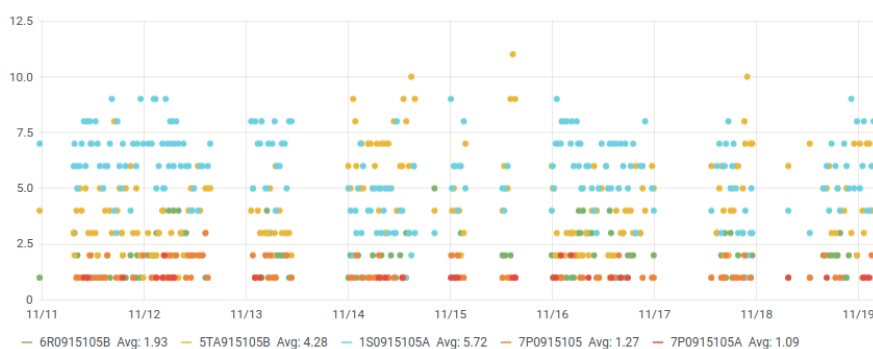


Figura 3.14: Número contentores de baterias sequenciadas, por tipo de bateria, ao longo da janela de tempo dos dados.

Visto que queremos então uma visualização que ilustre as baterias que foram corretamente sequenciadas, existe um parâmetro nos dados que indica se a bateria foi bem sequenciada, ou seja, se esta bateria foi sequenciada seguindo o protocolo correto, então o valor do parâmetro é falso. Se o contentor não foi sequenciado utilizando o protocolo correto, então o valor do parâmetro é verdadeiro. Utilizando este parâmetro, foram separados na visualização os valores da sequenciação feita seguindo o protocolo correto, com os primeiros representados com a cor verde, e os valores que não seguem o protocolo representados a vermelho.

A visualização resultante, na Figura 3.15, permite-nos ver que existe um limite para o qual, abaixo deste, todos os valores correspondem à não utilização do protocolo, com o valor aproximado de 1 minuto e 40 segundos. Traçando este limite, obtemos o gráfico da Figura 3.16.

Eliminando estes valores a vermelho, resultantes do incumprimento do protocolo de

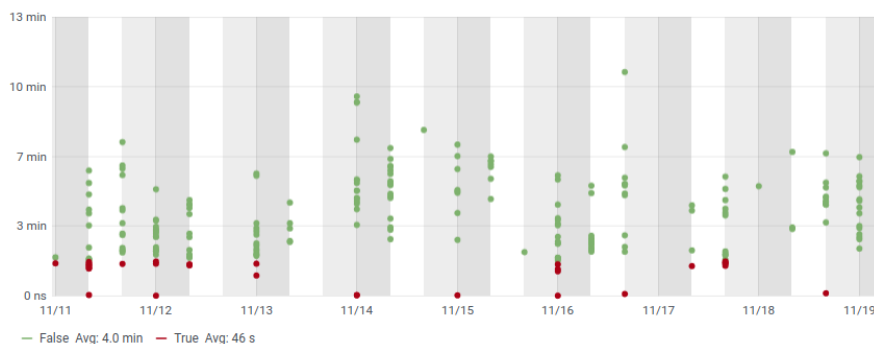


Figura 3.15: Tempos de sequenciação por turno, com os valores da sequenciação correta a verde e outliers a vermelho.

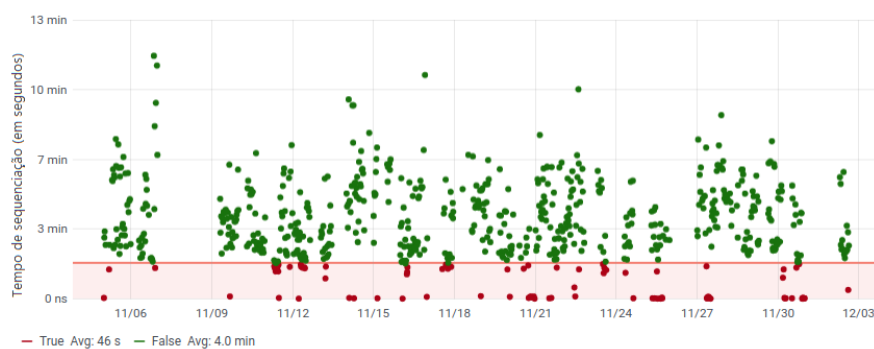


Figura 3.16: Tempos de sequenciação, com os valores da sequenciação correta a verde e com os valores de uma sequenciação errada a vermelho, e um *threshold*, marcado com uma linha horizontal a vermelho, que separa estas sequenciações erradas dos restantes valores.

sequenciação, e que não representam os tempos que se esperariam de uma sequenciação seguindo o protocolo, obtemos o gráfico seguinte, da Figura 3.17, que apresenta os tempos de sequenciação por turno.

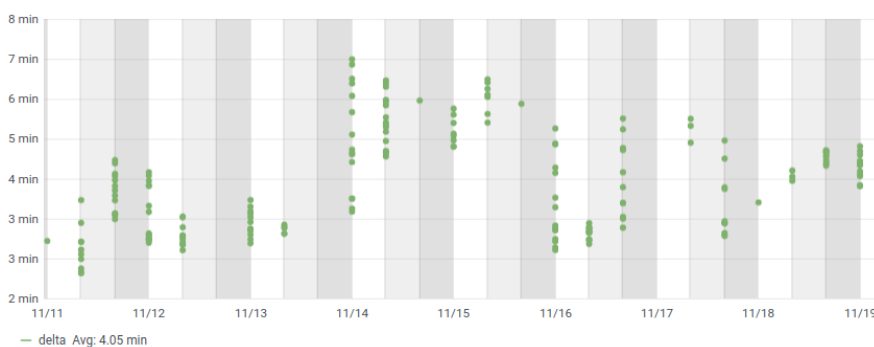


Figura 3.17: Tempos de sequenciação por turno, apenas com os valores da sequenciação seguindo o protocolo correto.

3.3.3 Implementação na plataforma Web

Com as visualizações desenvolvidas, o próximo passo é implementar as visualizações na página web previamente criada para as albergar. Para tal, é utilizada uma `iframe` html criada pelo Grafana para integrar os gráficos em *websites*, de modo a embutir as visualizações na plataforma do projeto Boost 4.0.

Para implementar a seleção da peça para a qual queremos visualizar os dados, foi utilizado um algoritmo open source *autocomplete*, desenvolvido pela jQuery e disponibilizado no seu website [59]. É feita uma query à base de dados que devolve um json com a lista de todas as famílias e número de peças, e este json é utilizado como data source nesta query de *autocomplete* para completar o input do utilizador numa caixa de texto. Quando selecionadas a família e o número da peça, a página é atualizada com as visualizações para a peça escolhida. Para que isto seja possível, plataforma web possui uma aplicação *backend*, desenvolvida em python, que faz as queries à base de dados, e devolve o json com a lista das famílias e números de peça, que é depois enviada para o algoritmo de Autocomplete. A figura 3.18 ilustra a versão final da caixa de texto criada para a seleção das peças.



Figura 3.18: Caixas de texto criadas para a seleção das peças a visualizar nos gráficos do *dashboard*.

Para além das caixas de texto foi também criada uma *sliding bar* que permite modificar a janela de tempo dos dados, para que seja possível visualizar com uma melhor resolução e com mais detalhes os dados para um período específico, como é possível ver na Figuras 3.20, 3.21 e 3.22.

O resultado desta implementação é uma página web, que só pode ser acedida por utilizadores com credenciais válidas, e que incorpora, para já, dois *dashboards*, divididos pelos processos associados: *Receiving* e *Warehousing*, com o *dashboard* de *Sequencing* ainda por implementar na plataforma. São apresentadas de seguida algumas imagens da versão mais recente da plataforma (Figuras 3.19, 3.20, 3.21 e 3.22).

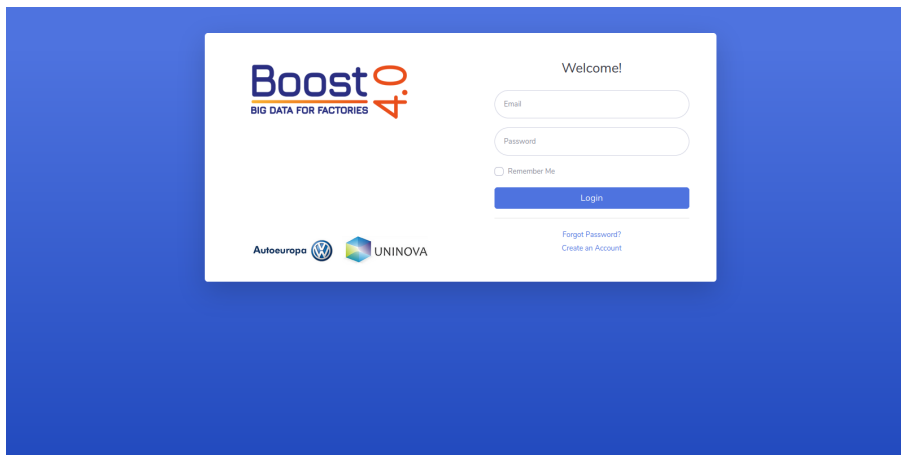


Figura 3.19: Página de login da plataforma de visualização do projeto Boost 4.0.

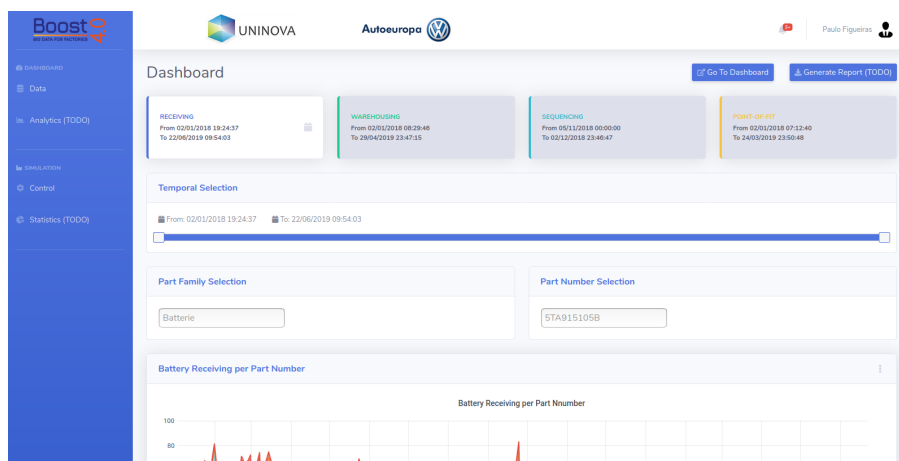


Figura 3.20: Página inicial da plataforma de visualização do projeto Boost 4.0, após login.

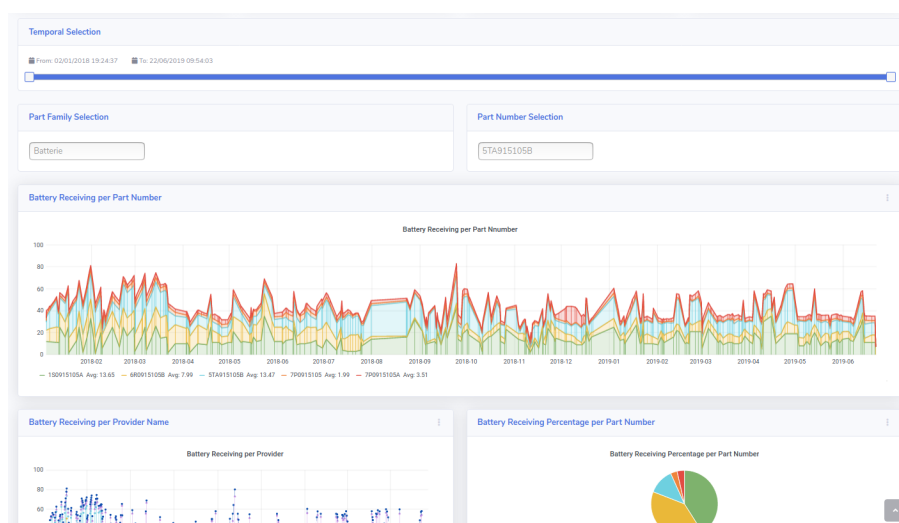


Figura 3.21: Dashboard do processo de *Receiving* na plataforma de visualização do projeto Boost 4.0.

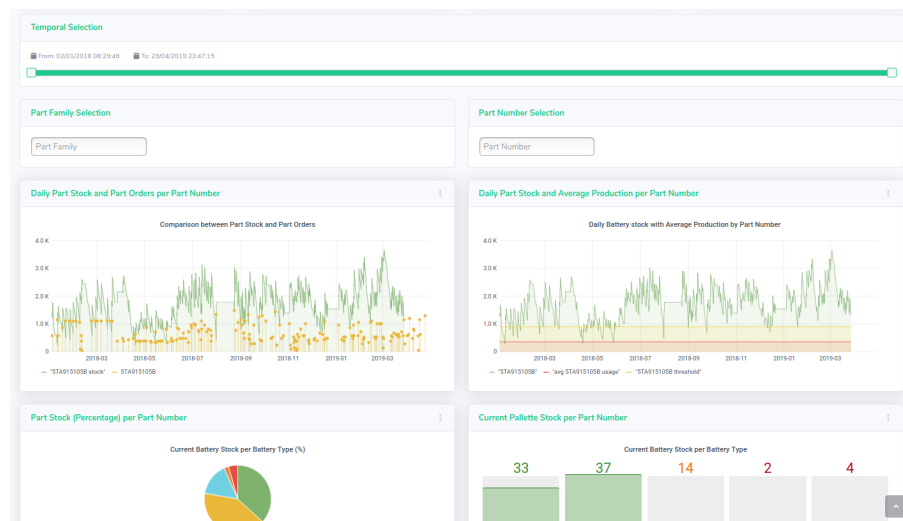


Figura 3.22: *Dashboard* do processo de *Warehousing* na plataforma de visualização do projeto Boost 4.0.

3.4 Caso de Estudo Ria Stone

Nesta secção, é abordado o caso de estudo da Ria Stone, uma empresa do ramo da indústria cerâmica. Neste caso de estudo, temos dois processos distintos, a produção e a avaliação de qualidade, e em ambos os processos são recolhidos dados que iremos analisar, com o objetivo de reduzir o erro na produção e deste modo, diminuir o número de produtos com defeitos e rejeitados, e como consequência, aumentar a percentagem de qualidade dos produtos. Para tal, iremos analisar os dados de qualidade e os dados do processo de *Glazing*, e tentar encontrar uma correlação entre os dados de forma a identificar o que provoca os defeitos nos produtos.

O processo de *glazing* ou processo de vidragem é o processo que aplica uma camada de revestimento vítreo aos produtos cerâmicos, de modo a proteger e impermeabilizar o produto. No caso da Ria Stone, este processo consiste na aplicação, através de pulverização, de um vidrado líquido, que é depois cozido num forno, a temperaturas elevadas, durante cerca de 9 horas.

3.4.1 Análise dos Dados e Seleção das Métricas

O objetivo do projeto Boost 4.0, para este caso de estudo, é fornecer uma plataforma de visualização de dados para suportar a visualização dos dados recolhidos pelos sensores recentemente implementados na linha de produção da Ria Stone.

Os dados recolhidos pela Ria Stone estão divididos em 2 tabelas. A primeira são os dados de qualidade, recolhidos pelos trabalhadores que fazem a avaliação da qualidade dos produtos, e possui dados como a percentagem de qualidade do lote, os defeitos que possam existir nos produtos, e *metadata* sobre o processo de avaliação, como a data da avaliação, o turno e a mesa em que esta foi feita. Esta tabela está representada na Figura 3.23.

A segunda tabela contém os dados do processo de *glazing* ou vidragem, a última fase da produção, e contém os vários parâmetros da linha de produção, como a temperatura de cozadura, densidade do vidro fluido, velocidade da rotação dos pratos e outros parâmetros recolhidos ao longo do processo.

Com base nos dados fornecidos, temos duas métricas importantes, a qualidade do produto, com base nos dados recolhidos no processo de análise de qualidade realizado pela Ria Stone, e os parâmetros do processo de *Glazing* dos produtos, recolhidos na linha de produção. Com estes dados, queremos abordar as seguintes análises:

- Análise dos dados de qualidade, com base nos dados do processo de qualidade, de acordo com os turnos e mesas de análise de qualidade;
- Análise dos dados de *glazing*, com base nos dados de *glazing*, de modo a identificar possíveis anomalias na produção;
- Correlação e análise entre os dados de qualidade e os dados do processo de *glazing*, de forma a tentar encontrar alguma relação entre anomalias na produção e os defeitos nos produtos.

3.4.2 Criação das visualizações

As visualizações para o caso de estudo da Ria Stone foram criadas na plataforma Grafana, tal como no caso da Volkswagen Autoeuropa (capítulo 3.3), sendo que a fonte dos dados é também uma base de dados PostgreSQL, que contém as tabelas com os dados de qualidade e do processo de *glazing*. Tendo em conta que temos dois processos distintos, as visualizações foram divididas em dois *dashboards*. O primeiro *dashboard* contém as visualizações do processo de avaliação de qualidade do produto, e o segundo contém as visualizações do processo de *glazing*, assim como a correlação destes com os dados de qualidade.

3.4.2.1 Dashboard do Processo de Avaliação de Qualidade

O *dashboard* do processo de avaliação de qualidade utiliza algumas visualizações desenvolvidas a partir de análises previamente realizadas em relatórios de qualidade da Ria Stone, combinadas com algumas visualizações novas que fornecem mais alguma informação sobre os dados de qualidade.

Começamos então por fazer uma visualização da qualidade, em percentagem, dos produtos por cada turno. Visto que queremos visualizar percentagens, utiliza-se um *pie chart* que possui três divisões: *first class* (primeira classe) que representam os pratos de qualidade mais elevada, *second class* (segunda classe) que representam os pratos de qualidade mais baixa que são aprovados, e por últimos, *rejected* (rejeitados) que, tal como o nome implica, foram rejeitados devido aos seus defeitos. Para visualizar estes dados por turno, foi utilizado o parâmetro *shift* (turno) da tabela do processo de avaliação de

3.4. CASO DE ESTUDO RIA STONE

	id [PK] bigint	date timestamp without time zone	shift character varying (2)	hour integer	prod_id bigint	color_id bigint	first_class integer	second_class integer	rejected integer	
1	1061	2019-10-28 00:00:00	T2		1	2	5	360	30	12
2	1062	2019-10-28 00:00:00	T2		3	2	5	240	50	12
3	1063	2019-10-28 00:00:00	T2		4	2	5	60	60	29
4	1064	2019-10-28 00:00:00	T2		5	6	5	480	30	9
5	1065	2019-10-28 00:00:00	T2		6	6	5	720	10	4
6	1066	2019-10-28 00:00:00	T2		2	13	3	864	20	8
7	1067	2019-10-28 00:00:00	T2		1	8	5	720	20	3
8	1068	2019-10-28 00:00:00	T2		2	8	5	696	40	4
9	1069	2019-10-28 00:00:00	T2		1	13	5	880	20	4
10	1070	2019-10-28 00:00:00	T2		1	12	1	528	0	16
11	1071	2019-10-28 00:00:00	T2		2	12	1	528	0	48
12	1072	2019-10-28 00:00:00	T2		3	12	1	720	0	7
13	1073	2019-10-28 00:00:00	T2		4	12	1	816	0	23
14	1074	2019-10-28 00:00:00	T2		5	11	1	408	0	12
15	1075	2019-10-28 00:00:00	T2		6	11	1	864	0	3
16	1076	2019-10-28 00:00:00	T2		2	13	6	480	20	9
17	1077	2019-10-28 00:00:00	T2		3	13	6	384	0	7
18	1078	2019-10-28 00:00:00	T2		1	8	6	912	40	15
19	1079	2019-10-28 00:00:00	T3		8	6	5	840	10	28
20	1080	2019-10-28 00:00:00	T3		5	13	3	792	30	12

(a) Estas colunas representam, da esquerda para a direita, o id do lote, a data de avaliação, o turno em que foi avaliada, a hora do turno, o id de produto, o id da cor, o número de produtos de primeira classe (a qualidade mais alta), o número de produtos da segunda classe e o número de produtos rejeitado.

total integer	checkpoint_nr integer	quality_percentage double precision	defects character varying (100)	stop_time integer	stop_cause character varying (20)
402	1	0.89552241563797	frete nicado/picos/pingos	5	51.0
302	1	0.794701993465424	picos/pingos	10	51.0
149	1	0.402684569358826	picos/pingos	25	56.0
519	1	0.924855470657349	picos/pingos	20	51.0
734	1	0.98092645406723		15	51.0
892	3	0.968609869480133		10	51.0
743	4	0.969044387340546		10	51.0
740	4	0.940540552139282	picos/frete nicado	10	51.0
904	5	0.973451316356659		5	51.0
544	6	0.970588207244873		25	51.0
576	6	0.916666686534882		25	51.0
727	6	0.990371406078339		5	51.0
839	6	0.972586393356323		5	51.0
420	7	0.971428573131561		5	
867	7	0.996539771556854		10	
509	7	0.943025529384613	frete vidrado zB e zC/picos	30	51;55
391	7	0.982097208499908		20	51
967	8	0.943123042583466	residuos/pingos/picos/f. beira	10	51.0
878	1	0.956719815731049		15	51.0
834	3	0.949640274047852	picos a; pingos a	8	51.0

(b) Estas colunas representam, da esquerda para a direita, o total de produtos, o número do *checkpoint*, a percentagem de qualidade do lote, os defeitos registados, o tempo de paragem e o código da causa de paragem.

Figura 3.23: Tabela do processo de avaliação de qualidade.

qualidade como métrica para selecionar apenas os pratos que foram produzidos no turno selecionado. O turno é selecionado com recurso a uma variavel de seleção que permite escolher entre os três turnos de produção da fábrica da Ria Stone. A Figura 3.24 ilustra o *pie chart* desenvolvido, para o primeiro turno de produção.

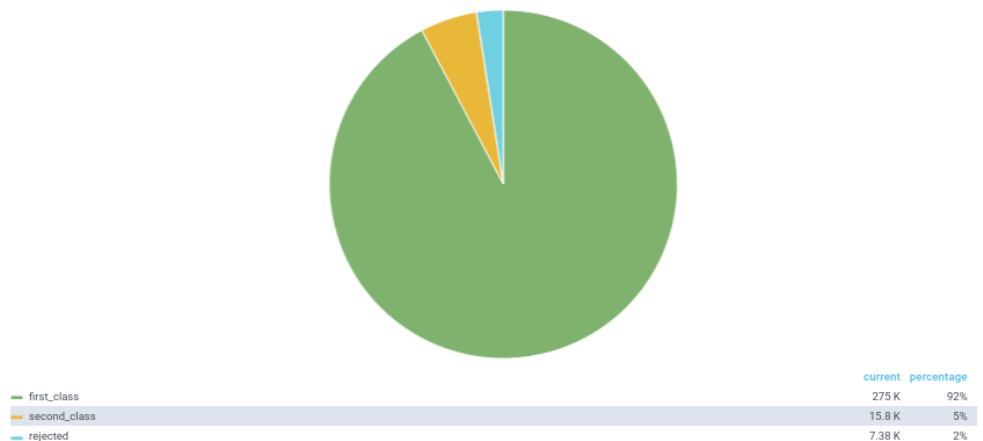


Figura 3.24: *Pie chart* que representa a qualidade dos produtos, em percentagem, por turno. A verde estão representados os produtos de qualidade mais alta, a amarelo os produtos de qualidade mais baixa que foram aprovados, e a azul os produtos que foram rejeitados.

Um dado interessante que também queremos ver é a qualidade média do produto por cada turno. Tendo em conta que temos três turnos de produção e queremos visualizar a média de qualidade de cada turno, utilizamos um gráfico de barras para esta visualização, em que cada barra corresponde á qualidade média, em valor absoluto, de 0 a 1, para cada um dos três turnos de produção. A Figura 3.25 ilustra a visualização em questão.

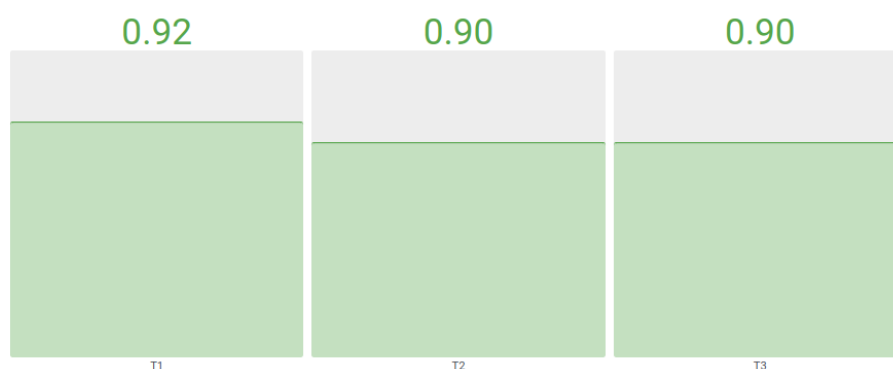


Figura 3.25: Gráfico de barras que ilustra a qualidade média dos produtos por turno, no qual T1 representa o primeiro turno, T2 o segundo turno e T3 o terceiro turno de produção.

Foi criada também uma visualização, como se pode ver na Figura 3.26 que serve como auxílio ao gráfico da Figura 3.25, que ilustra o número total de pratos produzidos em cada turno de produção. Este gráfico de barras, combinado com o gráfico da qualidade média

dos produtos por turno, permite-nos ter uma noção de quantos pratos foram aprovados em cada turno.

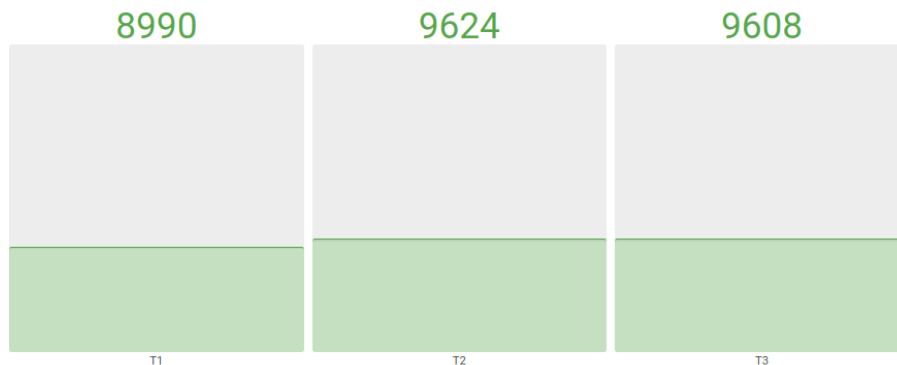


Figura 3.26: Gráfico de barras que ilustra o número total de pratos produzidos por turno, no qual T1 representa o primeiro turno, T2 o segundo turno e T3 o terceiro turno de produção.

As diferenças entre a qualidade de produto por cada mesa de avaliação de qualidade é também um dado interessante de visualizar. Como tal, foi criada uma visualização, também na forma de um gráfico de barras, em que cada barra corresponde a uma mesa de avaliação, com um total de 8 mesas e, portanto, o mesmo número de barras. Na Figura 3.27, correspondente a esta visualização, temos a verde as mesas com uma percentagem de qualidade superior a 90%, enquanto que a vermelho temos as mesas com uma percentagem de qualidade igual ou inferior a 90%.

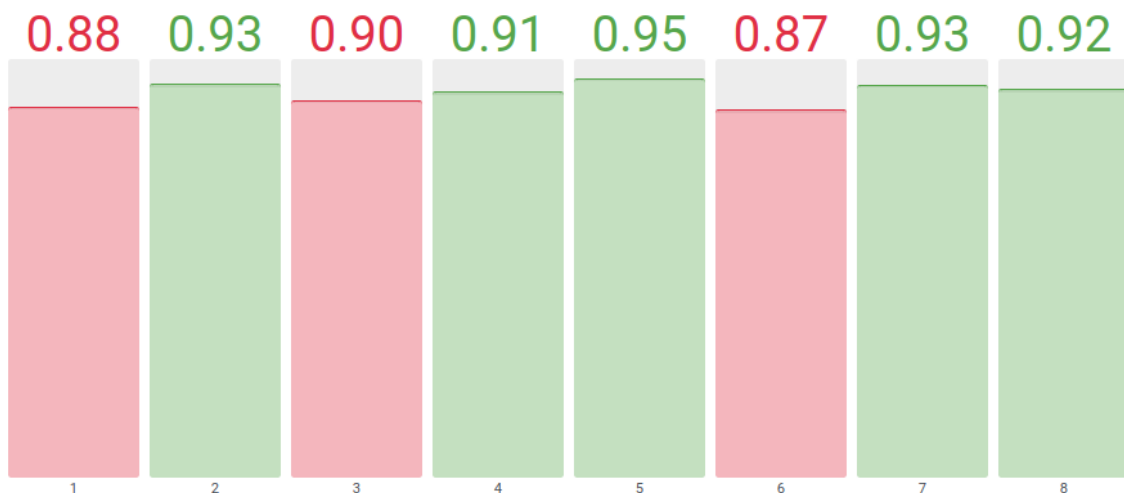


Figura 3.27: Gráfico de barras que ilustra a qualidade média dos produtos por mesa de avaliação de qualidade, no qual os números de 1 a 8 correspondem ao número de cada mesa.

Semelhantes ao gráfico de barras da Figura 3.27, temos ainda dois gráficos de barras que ilustram, respetivamente, a qualidade do produto em relação à cor (Figura 3.28) e ao tipo de produto (Figura 3.29). Tal como no gráfico da Figura 3.27, as barras a vermelho

representam os valores de qualidade abaixo dos 90% e as barras a verde representam os valores acima de 90%.

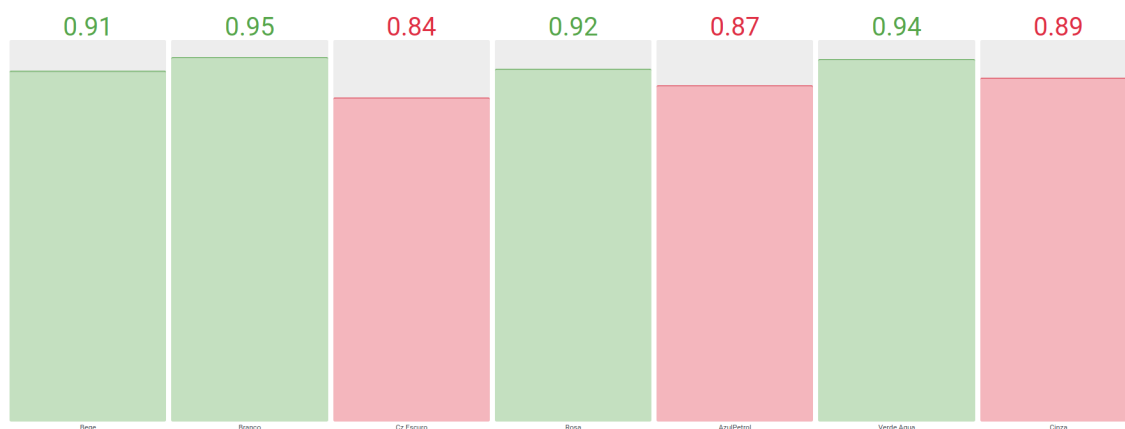


Figura 3.28: Gráfico de barras que ilustra a qualidade média dos produtos por cada cor de produto. Da esquerda para a direita temos: Bege, Branco, Cinza Escuro, Rosa, Azul Petróleo, Verde Água e Cinza

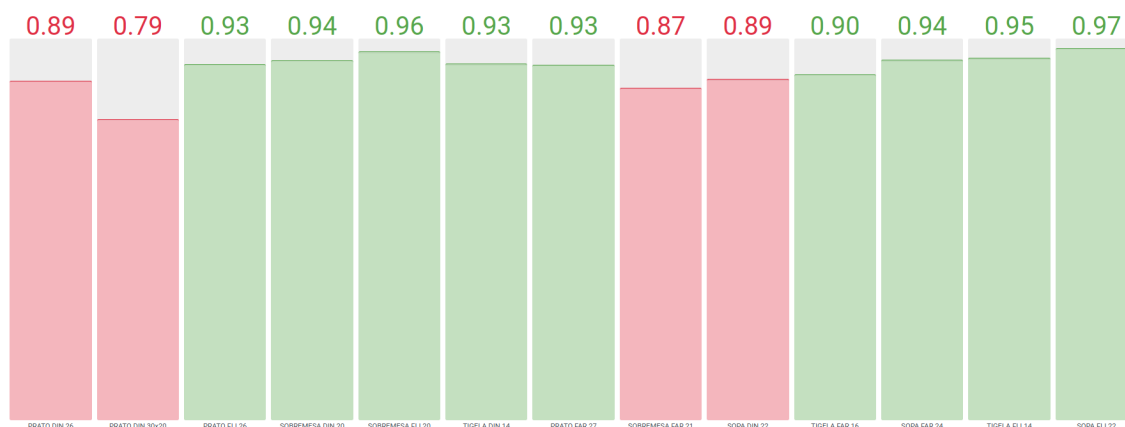


Figura 3.29: Gráfico de barras que ilustra a qualidade média dos produtos por tipo de produto. Os tipos de pratos representados no gráfico são, da esquerda para a direita: Prato DIN 26, Prato DIN 30x20, Prato FLI 26, Prato de sobremesa DIN 20, Prato de sobremesa FLI 20, Tigela DIN 14, Prato FAR 27, Prato de sobremesa FAR 21, Prato de sopa DIN 22, Tigela FAR 16, Prato de sopa FAR 24, Tigela FLI 14 e Prato de sopa FLI 22.

3.4.2.2 Dashboard do Processo de Glazing

Para o *dashboard* do processo de *Glazing*, queremos visualizar os parâmetros do processo e correlacionar estes dados com os dados de qualidade com o objetivo de encontrar correlações entre estes parâmetros e os defeitos reportados no processo de análise de qualidade. O primeiro passo é, portanto, introduzir algumas visualizações com os parâmetros do processo de *glazing*. Os parâmetros apresentados em visualizações serão aqueles que são considerados os mais influentes no surgimento de possíveis defeitos nos pratos.

Tendo em conta que queremos visualizar as alterações nestes parâmetros ao longo de um período de tempo, neste caso ao longo do mês de Setembro, as visualizações serão maioritariamente baseadas em gráficos de linha, com o eixo x a corresponder ao eixo temporal e o eixo y a representar o valor numérico dos dados. Começamos pelos parâmetros de velocidade da linha de produção e velocidade do torno, que estão ilustradas nas Figuras 3.30 e 3.31 respetivamente.

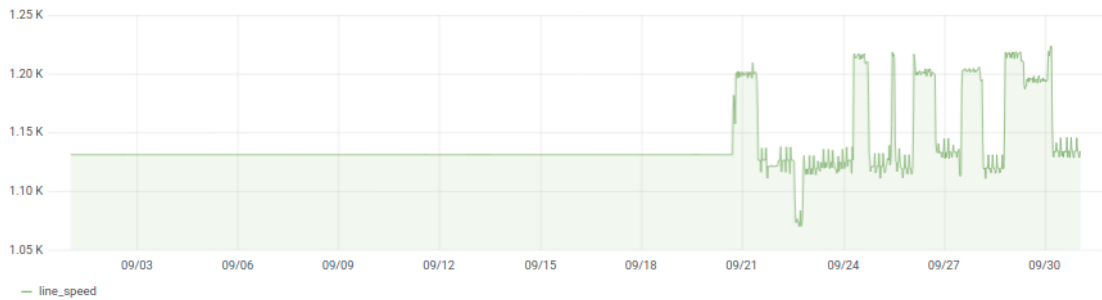


Figura 3.30: Gráfico de linha que representa visualmente a velocidade média da linha de produção para os pratos produzidos ao longo de um mês.

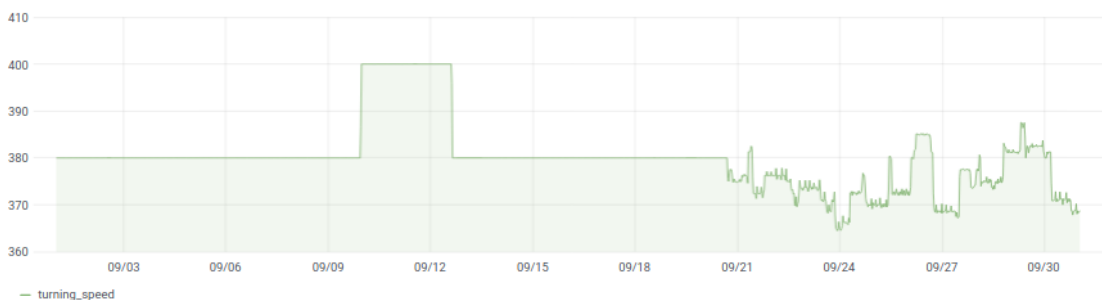


Figura 3.31: Gráfico de linha que representa visualmente a velocidade média do torno para os pratos produzidos ao longo de um mês.

Temos também os parâmetros de viscosidade do vidro líquido utilizado no processo de *glazing*, dados pela Figura 3.32, na qual a verde temos a viscosidade do vidro aplicado nos pratos, e a amarelo temos a viscosidade do vidro dentro do reservatório da fabrica, antes de ser aplicado.

Outro parâmetro para o qual foi criada uma visualização é a densidade do vidro. A Figura 3.33 apresenta a densidade do vidro utilizado no processo de *glazing* ao longo de um mês. É possível observar na figura que os valores de densidade do vidro caem bastante a partir por volta do dia 20.

Por outro lado, podemos observar o problema inverso na Figura 3.34, que ilustra o gráfico da temperatura do vidro no decorrer do mês. Neste caso, podemos observar que os valores da temperatura aumentam bastante, para valores a rondar os 24 mil graus Celsius, por volta do dia 20.

Para além destas visualizações, foram criadas algumas visualizações de suporte a estas,

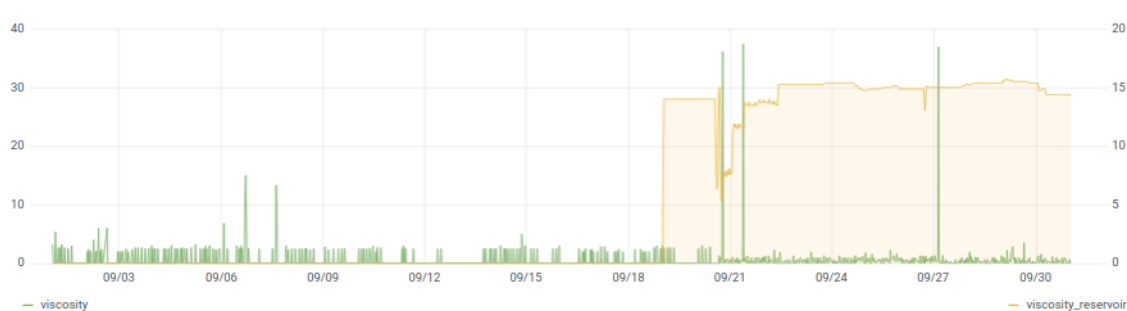


Figura 3.32: Gráfico de linha que representa visualmente a viscosidade média do vidro aplicado no processo de *glazing*, ao longo de um mês. A verde temos a viscosidade do vidro aplicado nos pratos, e a amarelo a viscosidade do vidro dentro do reservatório.

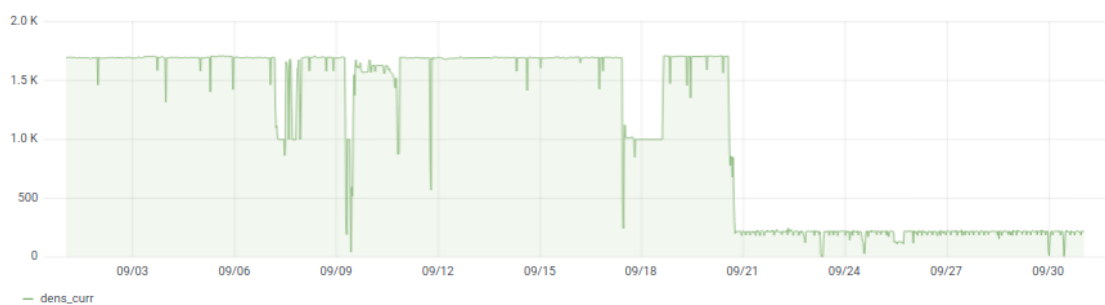


Figura 3.33: Gráfico de linha que representa visualmente a densidade média do vidro utilizado no processo de *glazing*, ao longo de um mês.

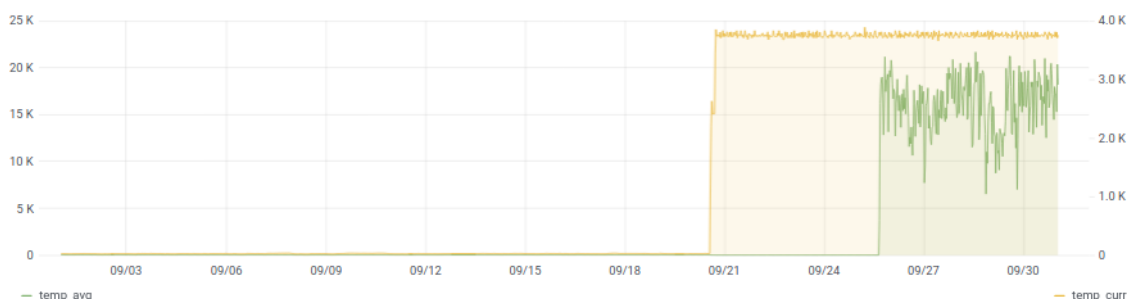


Figura 3.34: Gráfico de linha que representa visualmente a temperatura média do vidro utilizado no processo de *glazing*, ao longo de um mês. A verde temos a temperatura atual para cada prato vidrado no processo, enquanto que a amarelo temos a temperatura média do vidro utilizado no processo.

para os parâmetros de temperatura e densidade do líquido de *glazing*, por turno, dadas pelas Figuras 3.35 e 3.36, e por hora do dia, dadas pelas Figuras 3.37 e 3.38.

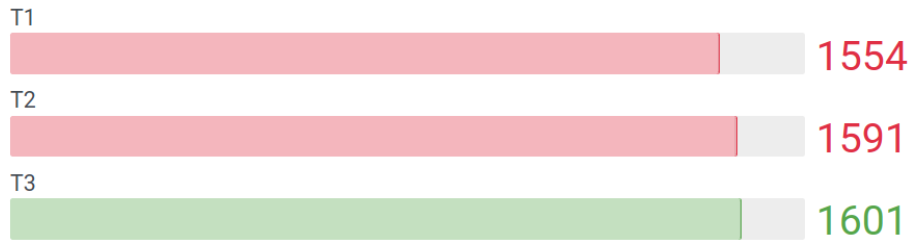


Figura 3.35: Densidade do líquido de *glazing* por cada turno de produção.

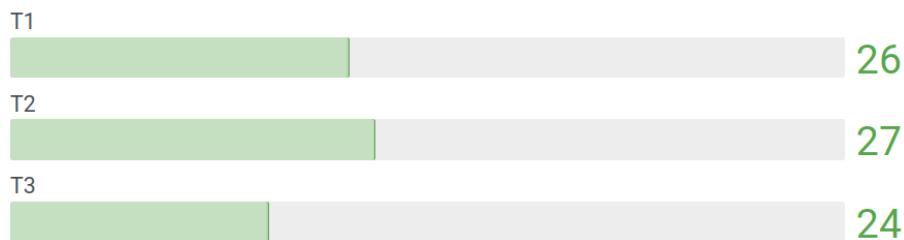


Figura 3.36: Temperatura do líquido de *glazing* por cada turno de produção.

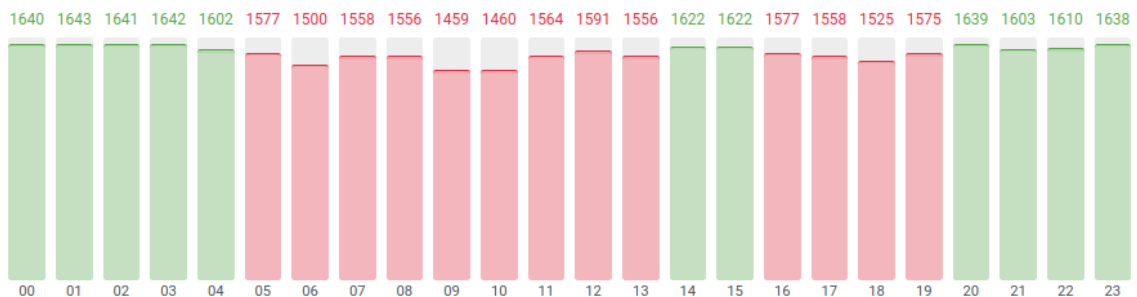


Figura 3.37: Densidade do líquido de *glazing* por hora do dia.

Com estas visualizações construídas, queremos agora visualizar a correlação destes parâmetros com a percentagem de qualidade, obtida no processo de avaliação de qualidade. Para tal, foi feita uma análise das correlações entre estes parâmetros e o parâmetro de percentagem de qualidade, dos dados do processo de avaliação de qualidade. Para que isso seja possível, foi criada uma *view* SQL que juntou os parâmetros que queremos correlacionar. Visto que o processo de avaliação de qualidade é realizado, segundo a Ria Stone, cerca de 8 a 9 horas depois de terminado o processo de *glazing*, foi utilizada a aproximação de 9 horas para alinhar os dados de qualidade com os parâmetros do processo de *glazing*. Disto resulta que não existe uma correspondência exata entre os dados de *glazing* e os dados de qualidade. Desta *view*, foi exportada para um CSV a tabela criada com os parâmetros a correlacionar.

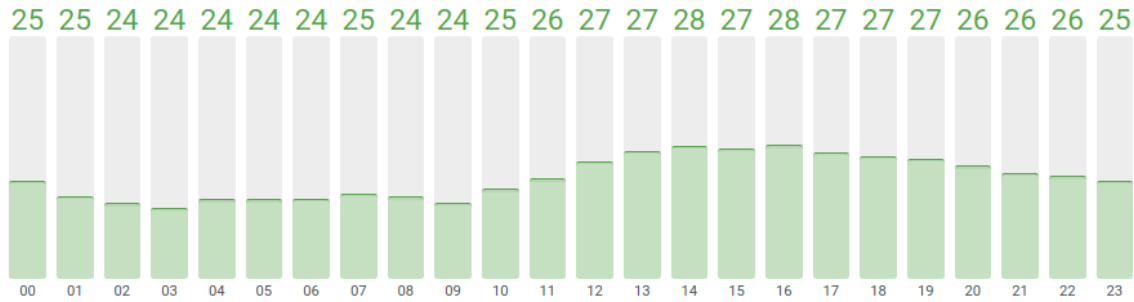


Figura 3.38: Temperatura do liquido de *glazing* por hora do dia.

Utilizando estes dados, e com recurso às funções de correlação de *dataframes* da biblioteca Pandas, do Python, foi criada uma matrix de correlação com os 5 parâmetros da tabela do processo de *glazing* e o parâmetro de percentagem de qualidade da tabela do processo de avaliação de qualidade. Para realizar esta correlação, foi utilizado o método de correlação de Pearson, dado pela expressão 3.1, no qual x e y são as colunas, o numerador dá-nos a covariância e o denominador dá-nos o desvio padrão. O resultado é a matriz da Figura 3.39, que ilustra a força das correlações, que utiliza uma escala de cores na qual o verde representa correlações mais fortes e vermelho representa as correlações mais fracas.

$$r = \frac{(n(\sum xy) - (\sum x)(\sum y))}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (3.1)$$

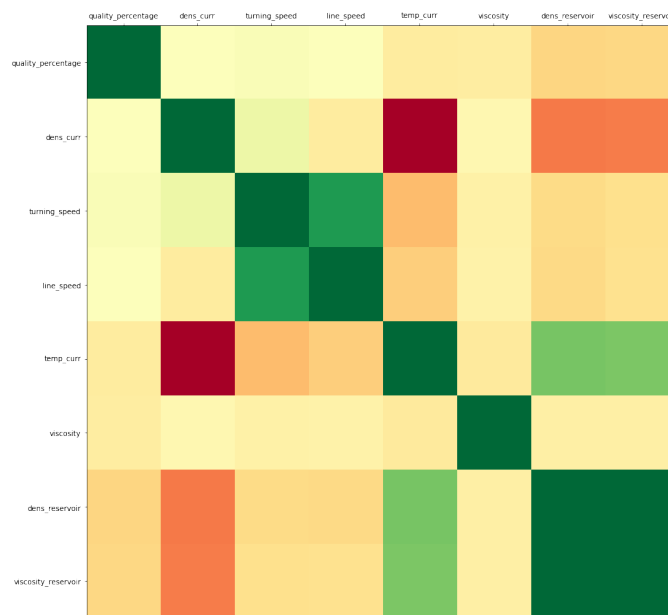


Figura 3.39: Visualização gráfica da matriz de correlação dos parâmetros do processo de *glazing* com a percentagem de qualidade.

Os valores da correlação são dados pela seguinte tabela:

Tabela 3.1: Análise de correlação dos parâmetros do processo de *glazing* com o parâmetro de percentagem de qualidade dos pratos, do processo de avaliação de qualidade.

Parâmetro	Correlação com Qualidade
Densidade	0.085308
Velocidade do Torno	0.097231
Velocidade da Linha	0.082432
Viscosidade	-0.028042
Temperatura	-0.037926
Viscosidade do Reservatório	-0.139207
Densidade do Reservatório	-0.145384

Como é possível observar pela Figura 3.39 e pela Tabela 3.1, os valores da correlação são bastante baixos e a correlação é quase inexistente entre a qualidade e estes parâmetros do processo de *glazing*. A isto deve-se em parte ao facto de que, de acordo com a Ria Stone, a partir de cerca do dia 19 do mês de Setembro, os dados retirados do processo de *glazing* são dados defeituosos, por força de erros de leitura por parte dos sensores recentemente instalados na fabrica para a recolha dos dados, e como tal, não correspondem aos reais valores do processo. Como tal, decidimos descartar estes dados, e focar-nos apenas e só nos dados anteriores ao dia 19.

Foi então criada uma *view* semelhante à anterior, em que temos apenas os dados desde o dia 1 até ao dia 19, da qual foi exportada para CSV a tabela resultante. Com este CSV foi realizada uma nova análise de correlação. Desta análise obtemos a seguinte matriz de correlação, da Figura 3.40.

Um outro motivo para a baixa correlação é a baixa ou quase inexistente integração dos dados de qualidade e do processo de *glazing*. Nos dados fornecidos, não existe nenhum parâmetro único na tabela do processo de *glazing* que identifique cada uma das peças, ou seja, não existe uma *primary key* que permita identificar da peça e fazer a associação à tabela de qualidade. Como tal, foi feita uma aproximação utilizando uma estimativa do intervalo de tempo entre a saída da peça do processo de *glazing* e a sua passagem no processo de avaliação de qualidade. Esta aproximação dá-nos resultados menos precisos, o que, por sua vez, baixa os valores da correlação obtidos.

Na Tabela 3.2, podemos ver os novos valores de correlação dos parâmetros do processo de *glazing* com a qualidade dos pratos. Algumas tabelas não têm valores visto que o desvio padrão das mesmas é zero, e portanto matematicamente, o resultado é infinito, ou seja, não é um número finito. Podemos verificar que existe uma grande subida na correlação da densidade do vidro com a qualidade dos pratos, cerca de 40%, quando comparado com a correlação inicial. No entanto, a correlação da temperatura desceu ligeiramente em relação à primeira correlação. De notar que estes valores são possivelmente mais baixos devido às aproximações realizadas para fazer a relação entre a qualidade e os parâmetros

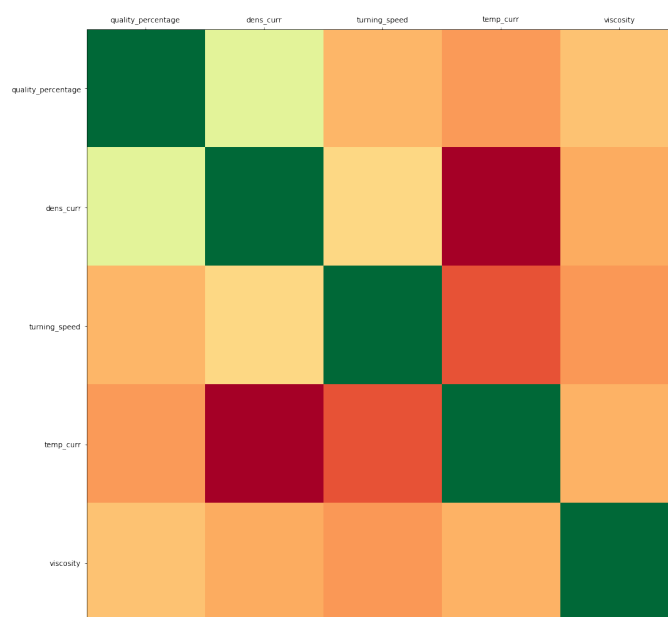


Figura 3.40: Visualização gráfica da matriz de correlação dos parâmetros do processo de *glazing* com a percentagem de qualidade, para os dados anteriores ao dia 19.

do processo de *glazing*.

Tabela 3.2: Comparação das análises de correlação dos parâmetros do processo de *glazing* com o parâmetro de percentagem de qualidade dos pratos, do processo de avaliação de qualidade, para o mês completo e para os dados entre o dia 1 e o dia 19.

Parâmetro	Correlação dados mês completo	Correlação dados dia 1-19
Densidade	0.085308	0.394152
Velocidade do Torno	0.097231	0.030667
Velocidade da Linha	0.082432	—
Viscosidade	-0.028042	0.067288
Temperatura	-0.037926	-0.034954
Viscosidade do Reservatório	-0.139207	—
Densidade do Reservatório	-0.145384	—

Estas diferenças de valores são possíveis observar nas Figuras 3.41 e 3.42. Na Figura 3.41 podemos observar que não existe uma relação óbvia entre os valores de temperatura e os valores de qualidade na maior parte dos pontos do gráfico. Na Figura 3.42, no entanto, já é possível observar uma maior correlação entre os dados da densidade e os dados da qualidade, com 3 "picos" na linha da densidade, a verde, que correspondem a "picos" na mesma altura na linha da qualidade.

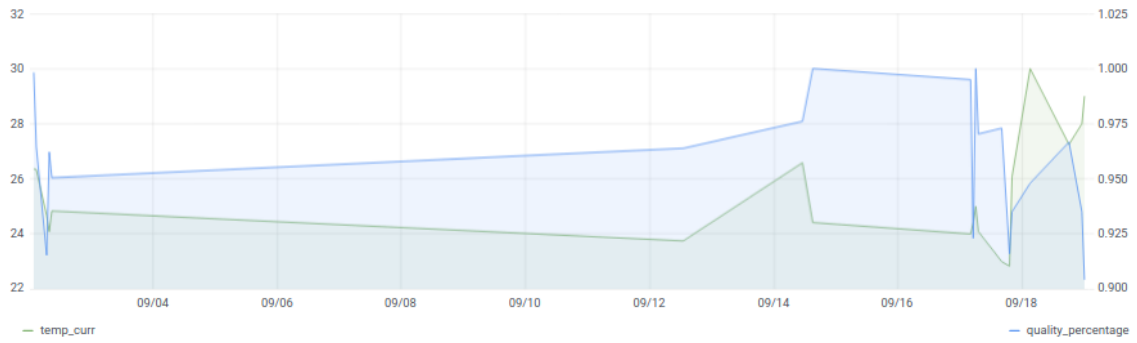


Figura 3.41: Gráfico da correlação entre os valores de temperatura do vidro (a verde) e os valores de percentagem de qualidade (a azul).

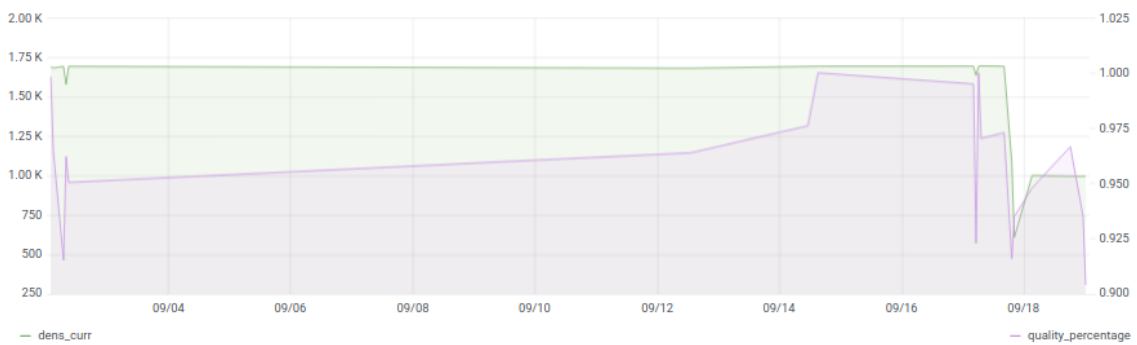


Figura 3.42: Gráfico da correlação entre os valores da densidade do vidro (a verde) e os valores de percentagem de qualidade (a roxo).

DISCUSSÃO DE RESULTADOS

Este capítulo aborda os resultados obtidos no capítulo anterior, e será dividido em duas partes, correspondentes a cada um dos casos de estudo apresentados, sendo estes: (i) o caso de estudo da Volkswagen Autoeuropa e, (ii) o caso de estudo da Ria Stone.

O objetivo do capítulo é discutir os resultados obtidos das visualizações, o conhecimento obtido com estas visualizações, assim como os problemas e complicações com os quais nos deparamos.

4.1 Caso de Estudo Volkswagen Autoeuropa

Com recurso às visualizações desenvolvidas na secção 3.3, foi possível retirar novos conhecimentos que nos permitem obter algumas conclusões e atuar sobre os processos logísticos, de modo a otimizá-los.

Com base nas visualizações do dashboard do processo de *receiving*, das figuras 3.4 e 3.5, da página 52, e as figuras 3.7, 3.8, da página 55, e 3.9, da página 55, é possível visualizar quais os tipos de baterias que são consideradas as *high runners* e quais são as *low runners*. As baterias do tipo 7P0915105 e 7P0915105A têm números de stock muito baixos, quando comparados com os valores de stock das baterias do tipo 1S0915105A, 5TA915105B e 6R091515B. Como tal, podemos considerar as primeiras como *low runners*, e são utilizadas em modelos, ou variantes de modelos, com menor volume de produção, e podemos considerar os outros três tipos de baterias como *high runners*, tendo em conta os mais elevados valores de stock. Para além deste conhecimentos, ainda no processo de *receiving*, obtemos também, com a Figura 3.6, da página 53, a distribuição das encomendas a cada fornecedor, para cada dia no qual existem encomendas registadas na base de dados.

Para o processo de *sequencing*, as visualizações permitem-nos obter o número de baterias sequenciadas por tipo de bateria (3.14 da página 59), assim como os tempos de

sequenciação por turnos (3.15, página 60, e 3.17, página 60). Os tempos de sequenciação das baterias permitem-nos ter uma imagem do tempo médio que um operador humano demora a realizar o processo de sequenciação, e comparar com o tempo que demora o mesmo processo automatizado. Após realizada a análise das visualizações, é possível observar algumas discrepâncias entre os tempos, com os mais baixos na ordem dos 2 a 3 minutos, e os mais altos superiores a 10 minutos, com a grande maioria dos tempos a concentrarem-se na faixa entre 1.5 e os 7 minutos. Isto resulta não só do facto de os operadores realizarem a sequenciação a velocidades diferentes seguindo o protocolo de sequenciação, mas também do facto de nem todos os operadores respeitarem o protocolo de sequenciação. Como tal, é difícil retirar uma conclusão precisa dos tempos médios de sequenciação dos operadores humanos.

Para o processo do processo de *warehousing*, tendo definida uma margem de segurança de duas vezes e meia a média de utilização diária de cada tipo de bateria, queríamos visualizar quais eram os valores de stock ao longo dos meses do ano, visualizar padrões e tendências nestes dados, e verificar a relação dos mesmos com a margem de segurança. Olhando para o gráfico da figura 3.9 da página 55, é possível observar alguns períodos no qual não existe variação de stock, em Abril, que corresponde à Páscoa, em Agosto, altura em que a fábrica fecha para férias de verão, e final de Dezembro, na época de Natal. Podemos também observar que antes da paragem para as férias de verão, de 1 a 23 de Agosto, o stock de baterias tem um "pico" negativo, que advém de um número mais baixo de baterias encomendadas antes desta paragem. Observando o gráfico da figura 3.12 da página 57, é possível observar que é isso mesmo que acontece, ou seja, a quantidade de baterias encomendadas diminui ao longo do mês de Julho, e aumenta outra vez com a última encomenda antes da paragem. Durante esta paragem, e também durante as outras paragens, é possível ver que não existiu qualquer atividade na fábrica pela ausência de encomendas e pela constância do número de baterias em stock. Na figura 3.11 da página 57, é possível observar a comparação do stock ao longo da nossa janela temporal, de Janeiro de 2018 a Abril de 2019, com duas métricas: o valor médio aproximado de utilização deste tipo de bateria, por dia de produção, que é apresentado pela linha a vermelho, e o valor da margem de segurança necessária para que não existam quebras de stock que potenciem uma paragem da produção, sendo que este valor foi definido como duas vezes e meia o valor médio de utilização diária de cada peça. É possível observar que os valores de stock têm vários "picos" com valores bastante superiores aos valores da margem de segurança, alguns deles mais de 2 e 3 vezes superiores à margem. De modo a otimizar o stock e reduzir custos associados ao processo intra-logístico da fábrica, o ideal seria que os valores de stock estabilizem em valores ligeiramente acima da margem de segurança.

De um modo geral, os resultados obtidos foram positivos, e permitiram extrair novos conhecimentos, tal como pretendido. Com as visualizações criadas, conseguimos quantificar o volume de stock e comparar o mesmo com o consumo e com as encomendas realizadas, de uma forma simples e intuitiva, o que poderá ajudar na otimização da gestão do

stock. No entanto, existem ainda algumas visualizações que podem ainda ser realizadas de modo a aprofundar este conhecimento, as quais irão ser abordadas no capítulo 5, de conclusões e trabalho futuro. Um problema com o qual me deparei foi a diferença da janela temporal entre os dados disponíveis para as baterias e as restantes peças, visto que os dados das baterias foram recolhidos desde o início de Janeiro até Abril de 2019, enquanto que apenas possuímos dados das restantes peças para o mês de Janeiro de 2020, o que tornou qualquer análise para as restantes peças quase impossível, dada a pequena janela temporal e a ausência de outros dados nesta janela temporal. É no entanto objetivo da Volkswagen Autoeuropa estender estas análises às restantes peças, sendo matéria para trabalho futuro.

4.2 Caso de Estudo Ria Stone

No caso de estudo da Ria Stone, os dados estão divididos em 2 processos distintos. Para o *dashboard* do processo de qualidade, foram reaproveitadas algumas visualizações já criadas previamente para o relatório do processo de qualidade, disponibilizado pela Ria Stone. São exemplos destas visualizações as Figuras 3.24 e 3.25, da página 66, e as Figuras 3.26 e 3.27, da página 67. Para além destas, foram ainda criadas outras visualizações que não existiam previamente, que ilustram a percentagem de qualidade para cada cor utilizada (Figura 3.28, página 68) e para o tipo de produto (Figura 3.29, página 68). Estas visualizações permitem-nos identificar quais os turnos, que tipos de produtos e quais as cores onde a qualidade é mais baixa, o que permite à empresa identificar mais facilmente onde estão os problemas que resultam numa baixa de qualidade dos produtos. No entanto, apenas os dados de qualidade não são suficientes para identificar a possível causa dos defeitos. Como tal, queremos também visualizar os dados dos outros processos de produção, de forma a melhor identificar e corrigir os fatores que resultam nestes defeitos.

Tendo em conta a natureza dos defeitos (pingos, fissuras, escorrimentos, vidro picado, entre outros), visualizamos os parâmetros do processo de *glazing* que podem estar na base destes defeitos. Como tal, foram criadas visualizações para vários parâmetros individuais do processo de *glazing*, de modo a identificar possíveis irregularidades nos valores. Os resultados destas visualizações são os gráficos das Figuras 3.30 e 3.31, da página 69, e os gráficos das Figuras 3.32, 3.33 e 3.34, da página 70. É possível observar pelas visualizações que a partir cerca do dia 20/09, os valores são bastante irregulares, em alguns casos muitos acima ou muito abaixo dos valores dos 19 dias anteriores. A Ria Stone foi abordada acerca deste problema, e foi concluído que estes valores, após o dia 20/09, são defeituosos e devem-se a falhas no funcionamento dos sensores recentemente instalados na linha de produção, devido ao facto de ser uma tecnologia com a qual a empresa não tem experiência.

Com o objetivo de identificar os parâmetros que influenciam a qualidade e a variação que estes parâmetros têm nos defeitos, foi feita uma *view SQL*, com recurso a um *INNER JOIN* entre a tabela do processo de qualidade e a tabela de *glazing* que devolve apenas os

produtos cuja cor, tipo de produto e hora de saída do processo de *glazing*. No entanto a tabela do processo de qualidade não possui uma coluna com a hora de saída do processo de *glazing*, apenas a hora em que é feita a avaliação. Não existe também uma *foreign key*, uma coluna com valores únicos que identifique cada um dos produtos, em nenhuma das tabelas, o que torna difícil ligar as duas tabelas. Como tal, foi feita uma aproximação do tempo que demora desde sair do processo de *glazing* até à entrada no processo de avaliação de qualidade, de 9h.

Sobre esta *view* foi feita uma análise de correlação entre a qualidade e os parâmetros do processo de *glazing*. Os resultados desta análise, dados pela Tabela 3.1 da página 73, permitem-nos concluir que a correlação entre a qualidade e os parâmetros do processo de *glazing* é bastante baixa, quase inexistente, consequência das condições em que foi feita a correlação entre as tabelas, mas também devido aos valores defeituosos dos dados após o dia 20 de Setembro. Como tal foi feita uma outra correlação entre as duas tabelas, utilizando apenas os dados anteriores ao dia 20 de Setembro. Os resultados da nova correlação são dados pela tabela 3.2 da página 74, que compara os valores da nova correlação com a primeira. Podemos observar que o valor da correlação entre a densidade e a qualidade subiu consideravelmente após eliminarmos um dos fatores de erro, nomeadamente, os dados posteriores ao dia 19 de Setembro. Visto que nos dados anteriores ao dia 20 de Setembro é visível alguns pontos em que as subidas e descidas de qualidade encontram correspondência nas subidas e descidas do valor da densidade do vidro, enquanto que essas correspondências não existem nos dados posteriores ao dia 19 de Setembro, No entanto, o valor continua a ser baixo, cerca de 40%. Isto deve-se em parte às aproximações feitas para associar as duas tabelas, que não garantem que o parâmetro de qualidade associado a um produto seja o verdadeiro valor de qualidade desse produto.

O resultado destas correlações são as Figuras 3.41 e 3.42, da página 75, que correspondem aos dois parâmetros com a correlação mais alta com a qualidade. Conseguem-se observar nestas figuras alguns "picos" na qualidade que correspondem a "picos" nos dois parâmetros, mais evidente no gráfico da densidade.

Com as visualizações desenvolvidas, centralizámos as análises dos dados de qualidade, que já existiam, numa plataforma de visualização com tecnologia mais adequada para o efeito, assim como pudemos retirar alguns novos conhecimentos, nomeadamente sobre o processo de *glazing*. No entanto, os resultados obtidos nas visualizações que envolvem os dados do processo de *glazing* foram condicionados pelos dados disponíveis e pela estrutura destes dados, que não permitiu retirar todas as conclusões pretendidas, nomeadamente, qual a relação entre os parâmetros do processo de *glazing* e a qualidade dos pratos. Devido à forma como os dados foram obtidos dos processos de avaliação de qualidade e de *glazing*, dois processos separados dentro da fábrica da Ria Stone, não existe qualquer rastreabilidade dos pratos que permita uma relação sólida dos dados dos dois processos. Existem 3 fornos que são utilizados no processo de *glazing*, e 8 mesas de avaliação utilizadas no processo de avaliação de qualidade, pelo que é impossível utilizar o método *first in, first out* para relacionar as duas tabelas. Não existe também qualquer

ID único que identifique cada peça. Como tal, foram necessárias realizar aproximações rudimentares de forma a poder obter alguma correlação entre os dados dos dois processos. Isto também impossibilitou outras análises, como a associação dos tipos de defeitos aos parâmetros do processo de *glazing*, de forma a identificar qual a influencia dos parâmetros nos defeitos que são encontrados em alguns produtos.

Para além destes problemas, existe também o problema da falta de dados de qualidade no *dataset* da Ria Stone. Esta falta de dados deve-se a dois fatores, sendo eles a instalação recente dos sensores na linha do processo de *glazing* da fábrica, e falhas na leitura dos parâmetros. Estas falhas na leitura dos parâmetros devem-se ao facto de os sensores serem descalibrados devido à condições do forno de *glazing*, como a temperatura e humidade.

Estes dois problemas levam a que exista uma amostra de dados muito baixa, o que dificulta ainda mais a realização das análises pretendidas.

CONCLUSÃO E TRABALHO FUTURO

Este capítulo encontra-se dividido em duas secções. A primeira, secção 5.1, apresenta as conclusões sobre os resultados obtidos nos capítulos 3 e 4. A secção 5.2 aborda uma sugestão de trabalho futuro para melhoramento e continuação do trabalho desenvolvido no âmbito desta dissertação.

5.1 Conclusões

A pergunta de investigação principal, e a respetiva hipótese, apresentadas na secção 1.5, propõe que seria possível descobrir padrões nos dados, correlações entre os dados disponíveis e, existindo estes padrões e correlações, será então possível correlacionar os dados da produção e detetar problemas de optimização da produção, para o caso de estudo da Volkswagen Autoeuropa, ou de qualidade do produto, para o caso de estudo da Ria Stone, a partir dos dados disponibilizados. A outra questão levantada foi como seria desenvolvida a plataforma de visualização de dados e quais as tecnologias que existem para visualização de grandes volumes de dados, para aplicar no contexto da indústria. A hipótese dada para esta pergunta foi o desenvolvimento de uma plataforma web que integra tecnologias de *Big Data* e visualização interativa de dados.

Tendo em conta a primeira pergunta, o objetivo desta dissertação consistiu em analisar os dados disponibilizados, nos dois casos de estudo, de modo a retirar novos conhecimentos através da correlação dos dados da produção. No caso de estudo da Volkswagen Autoeuropa, queremos implementar uma plataforma de visualização com visualizações que ajudem a otimizar os processos logísticos da fábrica, de modo a reduzir custos e aumentar a eficiência dos processos. No caso de estudo da Ria Stone, o objetivo era implementar uma plataforma com visualizações que permitam identificar as causas para os defeitos encontrados nos produtos com falhas de qualidade, nomeadamente quais os

parâmetros do processo de vidragem (*glazing*) dos produtos que influenciam a qualidade dos produtos, e qual a relação entre os parâmetros e o tipo de defeitos.

Para responder à segunda pergunta, foram estudadas algumas técnicas de visualização que nos permitem responder a estes problemas, assim como as plataformas existentes no mercado que nos permitem visualizar estes volumes de dados, plataformas de visualização de dados ou de *Business Intelligence*. Algumas das técnicas de visualização de dados estudadas foram implementadas nas visualizações criadas, tais como os tipos de visualização mais apropriados para os dados estudados e as escalas de cores utilizadas para algumas das visualizações criadas. A plataforma selecionada para desenvolver as visualizações foi o Grafana, e para além desta plataforma, as tecnologias utilizadas incluem PostgreSQL, para fazer a gestão da base de dados e a criação de *views*, e algumas linguagens de programação utilizadas para desenvolver o *back end* da plataforma web, como Python e JavaScript, assim como linguagens como html e CSS para o *front end* da plataforma.

Tendo em conta estes objetivos, dados pela pergunta de investigação, e com base nos resultados obtidos nos capítulos 3 e 4, podemos concluir que o objetivo foi parcialmente cumprido, visto que conseguimos alcançar uma parte dos objetivos traçados. Para o caso de estudo da Volkswagen Autoeuropa, os principais objetivos, de visualizar métricas que permitam otimizar os processos logísticos da fábrica, assim como o de implementar a plataforma web, foram concluídos com sucesso. Para o caso de estudo da Ria Stone, foram implementadas visualizações do processo de qualidade, e algumas visualizações do processo de vidragem. No entanto, existiram complicações com as correlações entre dados de qualidade e os dados do processo de vidragem, em grande parte devido à qualidade dos dados, que não permitiram obter bons resultados. Outro objetivo que não foi cumprido foi a implementação das visualizações numa plataforma web, tal como foi feito para o caso de estudo da Volkswagen Autoeuropa.

Concluindo, relativamente à pergunta de investigação, esta dissertação demonstrou que é possível utilizar tecnologias de visualização e análise de dados para descobrir padrões e correlações nos dados, se estes existirem. No entanto, para que estas possam ser aplicadas, é necessário garantir a qualidade dos dados que queremos analisar e visualizar, pois, como se pode ver, pelo caso de estudo da Ria Stone, foi necessário adaptar um parâmetro de uma das tabelas, da hora de entrada no processo de avaliação de qualidade, com base em aproximações pouco precisas dadas por responsáveis da empresa ligados ao projeto, para corresponder à hora de saída do processo de vidragem (*glazing*), de modo a permitir a correlação dos dados dos dois processos. Como esta aproximação não é muito precisa, torna a correlação dos dados muito difícil, pois baixa os valores de correlação obtidos, e torna impossível saber os verdadeiros valores de correlação entre a qualidade e os parâmetros do processo de vidragem.

5.2 Trabalho Futuro

Tendo em conta os problemas que foram encontrados no desenvolvimento do projeto, ficaram por realizar alguns objetivos, assim como existem alguns pontos a melhorar. Como tal, esta secção aborda alguns pontos para possível trabalho futuro, de forma a desenvolver o trabalho realizado, aumentar o conhecimento adquirido e melhorar a plataforma de visualização de dados.

Os pontos de trabalho futuro referente ao caso de estudo da Volkswagen Autoeuropa são os seguintes:

- Estudar outras métricas de interesse que permitam obter novos conhecimentos que ajudem na tomada de decisão e na otimização do processo logístico.
- Criar novas visualizações para as métricas já estudadas, que sejam intuitivas e que confirmem uma visão dos dados de forma mensal ou trimestral.
- Desenvolver mais visualização para os processos de *Receiving* e *Sequencing*.
- Estender as visualizações atuais do processo de *Warehousing* para as restantes peças utilizadas na construção dos automóveis.

Em relação ao caso de estudo da Ria Stone, os pontos de trabalho futuro são os seguintes:

- Realizar uma nova correlação entre os dados de qualidade e os dados de vidragem, dos processos do caso de estudo da Ria Stone, utilizando novos dados de qualidade, com o objetivo de obter melhores resultados.
- Integrar as visualizações na plataforma web do projeto Boost 4.0.

Para além destes pontos, a Ria Stone está também envolvida no projeto i4Q. O projeto i4Q visa "melhorar a qualidade dos produtos de modo a obter manufatura *ZERO-DEFECT* (zero defeitos)", utilizando uma Framework que garante dados de confiança, com funções agrupadas nas seguintes categorias [60]:

- Sensing;
- Comunicação;
- Infraestruturas de computação;
- Armazenamento;
- Análise e otimização.

O objetivo do projecto é fornecer a pequenas e médias empresas Europeias o serviço IoT-based Reliable Industrial Data Services (RIDS) de forma a permitir gerir as grandes quantidades de dados produzidos no ambiente industrial, com o objetivo de ajudar estas empresas a entrar na Quarta Revolução Industrial [61].

Este projeto irá continuar o trabalho desenvolvido nesta dissertação utilizando e melhorando os resultados obtidos.

BIBLIOGRAFIA

- [1] H. Lasi, P. Fettke, H. G. Kemper, T. Feld e M. Hoffmann. “Industry 4.0”. Em: *Business and Information Systems Engineering* 6.4 (2014), pp. 239–242. ISSN: 18670202. DOI: 10.1007/s12599-014-0334-4.
- [2] Volkswagen Autoeuropa, Uninova e INNOVALIA. *Boost 4.0 Chapter 2*. URL: <https://boost40.eu/deliverables/>.
- [3] Volkswagen Autoeuropa. "Acedido: 22/03/2021 às 14:07". URL: <https://www.volkswagenautoeuropa.pt/media/>.
- [4] Volkswagen Autoeuropa, Uninova e INNOVALIA. *Boost 4.0 Chapter 1*. URL: <https://boost40.eu/deliverables/>.
- [5] Ria Stone. "Acedido: 22/03/2021 às 12:30". URL: <https://grupovisabeira.com/pt/industria/ria-stone>.
- [6] M. Khan e S. Shah Khan. *Data and Information Visualization Methods, and Interactive Mechanisms: A Survey Readability from Screen View project News Linking and Preservation View project Data and Information Visualization Methods, and Interactive Mechanisms: A Survey*. Rel. téc. 1. 2011, pp. 975–8887. URL: <https://www.researchgate.net/publication/264623537>.
- [7] L. Chittaro. “Visualizing Information on Mobile Devices”. Em: *IEEE Computer Society* (2006).
- [8] S. Murray. “Interactive Data Visualization for the Web”. Em: (2017), pp. 2–3. URL: <https://www.oreilly.com/library/view/interactive-data-visualization/9781491921296/>.
- [9] S. Silva, B. Sousa Santos e J. Madeira. “Using color in visualization: A survey”. Em: *Computers and Graphics (Pergamon)* 35.2 (2011), pp. 320–333. ISSN: 00978493. DOI: 10.1016/j.cag.2010.11.015. URL: <http://dx.doi.org/10.1016/j.cag.2010.11.015>.
- [10] R. C. Gonzalez, R. E. Woods e S. L. Eddins. *Digital Image Processing Using Matlab - Gonzalez Woods & Eddins.pdf*. 2004. DOI: 10.1117/1.3115362.

- [11] E. Landa e M. Fairchild. "Charting Color from the Eye of the Beholder". Em: *American Scientist* 93.5 (2005), p. 436. ISSN: 0003-0996. DOI: 10.1511/2005.55.436. URL: https://www.researchgate.net/publication/343142798_Charting_Color_from_the_Eye_of_the_Beholder.
- [12] P. L. Rheingans. "Task-based Color Scale Design". Em: *28th AIPR Workshop: 3D Visualization for Data Exploration and Decision Making* 3905 (2000), pp. 35–43. DOI: 10.1117/12.384882.
- [13] H. Levkowitz e G. T. Herman. "Scales for Image Data". Em: *IEEE Computer Graphics and Applications* 12.1 (1992), pp. 72–80. DOI: 10.1109/38.135886.
- [14] Ocean Bottom Magnetology Laboratory. *Primer on Marine Magnetism*. URL: <http://deeptow.who.edu/primer.html>.
- [15] C. Brewer, M. Harrower e Pennsylvania State University. *ColorBrewer 2.0*. 2010. URL: <http://colorbrewer2.org/>.
- [16] W. H. Tedford, S. L. Bergquist e W. E. Flynn. "The size-color illusion." Em: *The Journal of general psychology* 97.1 st Half (1977), pp. 145–149. ISSN: 00221309. DOI: 10.1080/00221309.1977.9918511.
- [17] W. S. Cleveland e R. McGill. "A color-caused optical illusion on a statistical graph". Em: *American Statistician* 37.2 (1983), pp. 101–105. ISSN: 15372731. DOI: 10.1080/00031305.1983.10482720.
- [18] B. E. Rogowitz e L. A. Treinish. "How not to lie with visualization". Em: *Computers in physics* 10.3 (1996), pp. 268–273. ISSN: 08941866. DOI: 10.1063/1.4822401.
- [19] B. E. Rogowitz e L. A. Treinish. "Why Should Engineers and Scientists Be Worried About Color?" Em: *Lloydia Cincinnati* (2009), pp. 1–19. URL: <http://www.research.ibm.com/people/l/lloyd/t/color/color.HTM>.
- [20] L. Wang, J. Giesen, K. T. McDonnell, P. Zolliker e K. Mueller. "Color design for illustrative visualization". Em: *IEEE Transactions on Visualization and Computer Graphics* 14.6 (2008), pp. 1739–1746. ISSN: 10772626. DOI: 10.1109/TVCG.2008.118.
- [21] L. W. MacDonald. "Using color effectively in computer graphics". Em: *IEEE Computer Graphics and Applications* 19.4 (1999), pp. 20–35. ISSN: 02721716. DOI: 10.1109/38.773961.
- [22] Transport for London. *Tube map*. URL: <https://tfl.gov.uk/tfl/syndication/widgets/tubemap/default-search.html>.
- [23] "Intelligent Data Analysis". Em: ed. por M. Berthold e D. J. Hand. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003. Cap. 11. ISBN: 978-3-540-43060-5. DOI: 10.1007/978-3-540-48625-1. URL: <http://link.springer.com/10.1007/978-3-540-48625-1>.

- [24] D Sridevi, A Kumaravel e S Gunasekaran. “A Review on Big Data Visualization Tools”. Em: 3.7 (2020), pp. 45–49.
- [25] M. C. Ferreira de Oliveira e H. Levkowitz. “From visual data exploration to visual data mining: A survey”. Em: *IEEE Transactions on Visualization and Computer Graphics* 9.3 (2003), pp. 378–394. ISSN: 10772626. DOI: 10.1109/TVCG.2003.1207445.
- [26] P. E. Hoffman e G. G. Grinstein. “Table Visualizations: A Formal Model and Its Applications”. AAI9950455. Tese de doutoramento. 2000. ISBN: 0599547960. URL: <https://www.proquest.com/openview/4021c553875ca2ecc18952f3330d645e/1?pq-origsite=gscholar&cbl=18750&diss=y>.
- [27] I. Spence. “No Humble Pie : The Origins and Usage of a Statistical Chart”. Em: *Journal of Educational and Behavioral Statistics* 30.4 (2005), pp. 353–368. URL: <https://journals.sagepub.com/doi/abs/10.3102/10769986030004353>.
- [28] Adobe. *Adobe Flex 3 Advanced Data Visualization Developer Guide*. 2008. URL: <https://www.oreilly.com/library/view/adobe-flex-3/00120090006SI/>.
- [29] J. Clark. *Multi-level Pie Charts*. 2006. URL: www.neoformix.com/2006/MultiLevelPieChart.html.
- [30] S. Few. “Save the Pies for Dessert”. Em: *Visual Business Intelligence Newsletter* (2007), pp. 1–14. URL: https://www.perceptualedge.com/articles/visual_business_intelligence/save_the_pies_for_dessert.pdf.
- [31] L. Po, F. Desimoni, N. Bikakis e G. Papastefanatos. “Linked Data Visualization Techniques, Tools, and Big Data Synthesis Lectures on Data, Semantics, and Knowledge”. Em: (2020), pp. 23–38. URL: <https://doi.org/10.2200/S00967ED1V01Y201911WBE019>.
- [32] K. Pearson. “Contributions to the mathematical theory of evolution.—II. Skew variation in homogeneous material”. Em: 1895. Cap. 10, pp. 343–414. URL: <https://doi.org/10.1098/rsta.1895.0010>.
- [33] J. M. Utts. “Seeing Through Statistics”. Em: *Technometrics* (2005), pp. 166–167. ISSN: 0040-1706. DOI: 10.1198/tech.2005.s336.
- [34] S. Hirve e C. H. Pradeep Reddy. *A Survey on Visualization Techniques Used for Big Data Analytics*. Vol. 924. Springer Singapore, 2019, pp. 447–459. DOI: 10.1007/978-981-13-6861-5_39.
- [35] *A Complete Guide to Scatter Plots*. URL: <https://chartio.com/learn/charts/what-is-a-scatter-plot/>.
- [36] Q. V. Nguyen, N. Miller, D. Arness, W. Huang, M. L. Huang e S. Simoff. “Evaluation on interactive visualization data with scatterplots”. Em: *Visual Informatics* 4.4 (2020), pp. 1–10. ISSN: 2468502X. DOI: 10.1016/j.visinf.2020.09.004. URL: <https://doi.org/10.1016/j.visinf.2020.09.004>.

- [37] S. Bai, X. Zhou, Y. Lyu, J. Wang e C. Pan. “Data Visualization Model Methods and Techniques”. Em: *IOP Conference Series: Earth and Environmental Science* 252.5 (2019). ISSN: 17551315. DOI: [10.1088/1755-1315/252/5/052063](https://doi.org/10.1088/1755-1315/252/5/052063).
- [38] N. J. Salkind. “Statistics for people who (think they) hate statistics: the Excel edition”. Em: *Choice Reviews Online* (2007), p. 109. ISSN: 0009-4978. DOI: [10.5860/choice.44-3333](https://doi.org/10.5860/choice.44-3333).
- [39] S. J. Fong e R. C. Millham. *Bio-inspired Algorithms for Data Streaming and Visualization, Big Data Management, and Fog Computing*. 2021, p. 196. ISBN: 978-981-15-6694-3. URL: <https://link.springer.com/book/10.1007/978-981-15-6695-0>.
- [40] G. Chawla, S. Bamal e R. Khatana. “Big Data Analytics for Data Visualization: Review of Techniques”. Em: *International Journal of Computer Applications* 182.21 (2018), pp. 37–40. DOI: [10.5120/ijca2018917977](https://doi.org/10.5120/ijca2018917977).
- [41] D. A. Keim, H. P. Kriegel e M. Ankerst. “Recursive pattern: a technique for visualizing very large amounts of data”. Em: *Proceedings of the IEEE Visualization Conference* (1995), pp. 279–286. DOI: [10.1109/visual.1995.485140](https://doi.org/10.1109/visual.1995.485140).
- [42] M. Ankerst, D. Keim e H. Kriegel. “‘Circle Segments’: A Technique for Visually Exploring Large Multidimensional Data Sets”. Em: *Proc. IEEE Visualization '96, Hot Topic Session* (1996), pp. 5–8. URL: https://www.researchgate.net/publication/30013012_Circle_Segments_A_Technique_for_Visually_Exploring_Large_Multidimensional_Data_Sets.
- [43] J. LeBlanc e M. Ward. “Exploring N-dimensional databases”. Em: *Proc. Visualization '90* (1990), pp. 230–239.
- [44] D. Asimov. “The Grand Tour: A Tool for Viewing Multidimensional Data”. Em: *SIAM Journal on Scientific and Statistical Computing* (1985). ISSN: 0196-5204. DOI: [10.1137/0906011](https://doi.org/10.1137/0906011).
- [45] E. A. Bier, M. C. Stone, K. Pier, K. Fishkin, T. Baudelf, M. Conway, W. Buxton e T. DeRose. “Toolglass and magic lenses: The see-through interface”. Em: *Conference on Human Factors in Computing Systems - Proceedings 1994-April*.Mmm (1994), pp. 445–446. DOI: [10.1145/259963.260447](https://doi.org/10.1145/259963.260447).
- [46] K. Fishkin e M. C. Stone. “Enhanced dynamic queries via movable filters”. Em: *Conference on Human Factors in Computing Systems - Proceedings 1* (1995), pp. 415–420. DOI: [10.1145/223904.223960](https://doi.org/10.1145/223904.223960).
- [47] R. Rao e S. K. Card. “Table lens: merging graphical and symbolic representations in an interactive focus + context visualization for tabular information”. Em: *Conference on Human Factors in Computing Systems - Proceedings* (1994), pp. 318–322.
- [48] G. W. Furnas. “Generalized fisheye views”. Em: *ACM SIGCHI Bulletin* 17.4 (1986), pp. 16–23. ISSN: 07366906. DOI: [10.1145/22339.22342](https://doi.org/10.1145/22339.22342).

-
- [49] M. Sarkar e M. H. Brown. "Graphical Fisheye Views". Em: *Communications of the ACM* 37.12 (1994), pp. 73–83. ISSN: 15577317. DOI: 10.1145/198366.198384.
- [50] M. por dados. *O que é um dashboard?* URL: <http://marketingpardados.com/analise-de-dados/o-que-e-dashboard>.
- [51] BetterBuys. *Microsoft Power BI Review*. URL: <https://www.betterbuys.com/bi/reviews/microsoft-power-bi/>.
- [52] BetterBuys. *ThoughtSpot Review*. URL: <https://www.betterbuys.com/bi/reviews/thoughtspot/>.
- [53] BetterBuys. *Tableau Review*. URL: <https://www.betterbuys.com/bi/reviews/tableau-business-intelligence/>.
- [54] BetterBuys. *Qlikview BI Review*. URL: <https://www.betterbuys.com/bi/reviews/qlikview-business-intelligence/>.
- [55] GetApp. *Qlikview BI Features*. URL: <https://www.getapp.com/business-intelligence-analytics-software/a/qlikview-9/features/>.
- [56] BetterBuys. *Qlik Sense Review*. "Acedido: 19/04/2021 às 16:53". URL: <https://www.betterbuys.com/bi/reviews/qlik-sense/>.
- [57] Logz.io. *Grafana vs Kibana*. URL: <https://logz.io/blog/grafana-vs-kibana/>.
- [58] C. PRO. *Business Intelligence Tools Comparison Chart*. "Acedido: 19/04/2021 às 17:24". URL: <https://www.csopro.com/blog/business-intelligence-tools-comparison-chart/>.
- [59] *jQuery UI Autocomplete Widget*. "Acedido: 22/03/2021 às 14:03". URL: <https://jqueryui.com/autocomplete/>.
- [60] *i4Q Project*. "Acedido: 22/03/2021 às 14:04". URL: <https://www.i4q-project.eu/>.
- [61] *i4Q Project Introduction and Objectives*. "Acedido: 22/03/2021 às 14:35". URL: <https://www.i4q-project.eu/project>.

