



# **Da Análise dos Dados à Comunicação de Resultados: Uma Abordagem com Data Storytelling**

Mestrado em Ciência de Dados

Fernando Cruz Correia

Leiria, março de 2026



# **Da Análise dos Dados à Comunicação de Resultados: Uma Abordagem com Data Storytelling**

Mestrado em Ciência de Dados

Fernando Cruz Correia

Trabalho de Projeto realizado sob a orientação da Professora Doutora Rosa Isabel Alves Cordeiro Matias e da Professora Doutora Maria Beatriz Guerra da Piedade

Leiria, março de 2026

# **Originalidade e Direitos de Autor**

O presente relatório de projeto é original, elaborado unicamente para este fim, tendo sido devidamente citados todos os autores cujos estudos e publicações contribuíram para o elaborar.

Reproduções parciais deste documento serão autorizadas na condição de que seja mencionado o Autor e feita referência ao ciclo de estudos no âmbito do qual o mesmo foi realizado, a saber, Curso de Mestrado em Ciência de Dados, no ano letivo 2025/2026, da Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria, Portugal, e, bem assim, à data das provas públicas que visaram a avaliação destes trabalhos.

# Agradecimentos

Em primeiro lugar, quero agradecer às minhas orientadoras, Professora Doutora Rosa Isabel Alves Cordeiro Matias e Professora Doutora Maria Beatriz Guerra da Piedade, pela orientação, disponibilidade e acompanhamento ao longo do desenvolvimento deste trabalho, cujos contributos foram fundamentais para a sua concretização.

Agradeço também à instituição de ensino e ao corpo docente do curso, pela formação académica adquirida ao longo do percurso, que forneceu as bases teóricas e práticas necessárias para a realização deste projeto.

Agradeço aos colegas e amigos, que contribuíram com apoio, incentivo e partilha de conhecimento durante este processo.

Por fim, agradeço à minha família pelo apoio constante, compreensão e motivação ao longo de todo o percurso académico, bem como à minha namorada, pelo incentivo e apoio ao longo deste trabalho.

# Resumo

O crescimento do comércio eletrônico tem levado a um aumento significativo do volume e da complexidade dos dados gerados pelas plataformas digitais, reforçando a necessidade de soluções de *Business Intelligence* (BI) capazes de transformar dados em informação relevante para suporte à tomada de decisão. Neste contexto, a integração entre processos de preparação, modelação e exploração de dados, visualização de informação e *data storytelling* assume um papel importante na interpretação estruturada do desempenho organizacional.

O presente trabalho propõe o desenvolvimento de uma solução de BI orientada ao contexto de *marketplaces* digitais, estruturada em etapas de integração e transformação de dados, definição de indicadores-chave de desempenho e construção de *dashboards* interativos. A abordagem seguida combina modelação analítica e visualização de informação, permitindo uma análise integrada do desempenho do negócio.

A metodologia é aplicada a um caso de estudo baseado no *marketplace* brasileiro Olist, permitindo validar a arquitetura da solução e os princípios de concepção definidos. A análise realizada evidencia padrões associados ao crescimento do negócio, à dinâmica da base de clientes e à relação entre eficiência logística e satisfação do consumidor. Para além do contexto específico analisado, a solução desenvolvida apresenta características adaptáveis a diferentes plataformas de *e-commerce*, demonstrando o potencial de generalização do modelo proposto.

Conclui-se que a combinação entre processos de BI e *data storytelling* constitui uma abordagem eficaz de apoio à decisão estratégica, tática e operacional em ambientes digitais, contribuindo para uma comunicação mais estruturada e orientada ao negócio.

**Palavras-chave:** *Business Intelligence, Data Storytelling, E-commerce, Visualização de Informação*

# Abstract

The growth of e-commerce has led to a significant increase in the volume and complexity of data generated by digital platforms, reinforcing the need for Business Intelligence (BI) solutions capable of transforming data into relevant information to support decision-making. In this context, the integration of data preparation, modeling and exploration processes, information visualization, and data storytelling plays an important role in the structured interpretation of organizational performance.

This work proposes the development of a BI solution tailored to the context of digital marketplaces, structured around stages of data integration and transformation, definition of key performance indicators, and the development of interactive dashboards. The followed approach combines analytical modeling and information visualization, enabling an integrated analysis of business performance.

The methodology is applied to a case study based on the Brazilian marketplace Olist, allowing the validation of the solution's architecture and the defined design principles. The analysis highlights patterns related to business growth, customer base dynamics, and the relationship between logistics efficiency and customer satisfaction. Beyond the specific context analyzed, the developed solution presents adaptable characteristics to different e-commerce platforms, demonstrating the potential for generalization of the proposed model.

It is concluded that the combination of BI processes and data storytelling constitutes an effective approach to supporting strategic, tactical, and operational decision-making in digital environments, contributing to more structured and business-oriented communication.

**Keywords:** Business Intelligence, Data Storytelling, Information Visualization, E-commerce

# Índice

|   |            |
|---|------------|
| <b>Originalidade e Direitos de Autor .....</b>                    | <b>iii</b> |
| <b>Agradecimentos.....</b>  | <b>iv</b>  |
| <b>Resumo .....</b>   | <b>v</b>   |
| <b>Abstract.....</b>  | <b>vi</b>  |
| <b>Lista de Figuras.....</b>                                      | <b>xi</b>  |
| <b>Lista de Tabelas .....</b>                                     | <b>xiv</b> |
| <b>Lista de siglas e acrónimos.....</b>                           | <b>xv</b>  |
| <b>1. Introdução .....</b>  | <b>1</b>   |
| <b>1.1. Problema.....</b>   | <b>1</b>   |
| <b>1.2. Objetivo .....</b>  | <b>2</b>   |
| <b>1.3. Organização do Documento .....</b>                        | <b>3</b>   |
| <b>2. Estudo prévio.....</b>                                      | <b>5</b>   |
| <b>2.1. Dos Dados à Sabedoria: A Pirâmide DIKW.....</b>           | <b>5</b>   |
| <b>2.2. Conceitos de Business Intelligence .....</b>              | <b>6</b>   |
| 2.2.1. Camada de Fonte de Dados.....                              | 8          |
| 2.2.2. Camada ETL .....   | 8          |
| 2.2.3. Camada de Armazenamento.....                               | 9          |
| 2.2.4. Camada de Apresentação .....                               | 11         |
| 2.2.5. Camada de Metadados .....                                  | 12         |
| <b>2.3. Business Intelligence na era do Big Data .....</b>        | <b>13</b>  |
| <b>2.4. Representação Visual da Informação .....</b>              | <b>15</b>  |
| 2.4.1. Conceito e Importância da Visualização de Informação ..... | 15         |
| 2.4.2. Fundamentos da Visualização de Informação.....             | 16         |
| 2.4.3. Elementos Visuais Essenciais .....                         | 18         |
| 2.4.4. Tipos de Gráficos e as suas Aplicações .....               | 20         |
| 2.4.5. Ferramentas de Visualização.....                           | 22         |
| <b>2.5. Narrativas Baseadas em Dados.....</b>                     | <b>23</b>  |
| 2.5.1. Definição e Enquadramento.....                             | 24         |
| 2.5.2. Estrutura e Elementos do Data Storytelling .....           | 25         |
| 2.5.3. Desafios e Limitações .....                                | 27         |

|  |           |
|--|-----------|
| <b>2.6. E-commerce: Conceito e Relevância Analítica .....</b>                              | <b>28</b> |
| 2.6.1. Conceito e Evolução do E-commerce .....   | 28        |
| 2.6.2. Modelos de E-commerce .....   | 30        |
| 2.6.3. Benefícios e Desafios do E-commerce .....   | 31        |
| <b>2.7. Integração de Conceitos no Contexto Analítico do E-commerce .....</b>              | <b>31</b> |
| 2.7.1. Visualização de Dados e Business Intelligence no E-commerce .....                   | 32        |
| 2.7.2. Integração do Data Storytelling no E-commerce .....                                 | 34        |
| 2.7.3. KPIs no E-commerce: Operacionais, Estratégicos, Táticos e Financeiros .....         | 36        |
| 2.7.4. Classificação dos Dashboards: Operacional, Tático e Estratégico no E-commerce ..... | 40        |
| <b>3. Metodologia do Caso de Estudo .....</b>  | <b>43</b> |
| <b>3.1. Enquadramento Geral do Caso de Estudo .....</b>                                    | <b>43</b> |
| <b>3.2. Caracterização dos Dados Fonte .....</b>   | <b>44</b> |
| <b>3.3. Ferramentas e Tecnologias Utilizadas .....</b>                                     | <b>45</b> |
| <b>4. Desenho do Sistema .....</b>   | <b>47</b> |
| <b>4.1. Análise Exploratória de Dados .....</b>  | <b>47</b> |
| 4.1.1. Identificação e Descrição das Variáveis do Conjunto de Dados .....                  | 48        |
| 4.1.2. Resumo dos Datasets .....   | 48        |
| 4.1.3. Análise de Valores Nulos .....  | 49        |
| 4.1.4. Análise de Integridade Referencial .....  | 50        |
| 4.1.5. Análise Descritiva dos Campos Numéricos .....                                       | 51        |
| 4.1.6. Análise de Correlações Entre Variáveis Numéricas .....                              | 52        |
| 4.1.7. Análise Geral de Pedidos e Pagamentos .....   | 55        |
| 4.1.8. Análise Temporal das Vendas e Entregas .....  | 56        |
| 4.1.9. Análise das Categorias de Produto e da Satisfação dos Clientes .....                | 57        |
| 4.1.10. Análise de Clientes e Localização .....  | 58        |
| 4.1.11. Análise de Sentimentos e Palavras-Chave nas Avaliações dos Clientes .....          | 60        |
| <b>4.2. Requisitos do Sistema .....</b>  | <b>60</b> |
| 4.2.1. Perguntas de Negócio .....  | 61        |
| 4.2.2. User Stories .....  | 62        |
| 4.2.3. Requisitos Funcionais e Não Funcionais .....  | 63        |
| 4.2.4. KPIS e Métricas Relevantes .....  | 65        |
| <b>4.3. Definição da Narrativa Analítica .....</b>   | <b>67</b> |
| <b>4.4. Modelo Dimensional .....</b>   | <b>68</b> |
| <b>4.5. Arquitetura do Sistema de BI .....</b>   | <b>71</b> |
| <b>5. Processo de Preparação, Limpeza e Transformação dos Dados .....</b>                  | <b>73</b> |

|  |            |
|--|------------|
| <b>5.1. Preparação de Dados de Origem .....</b>                        | <b>73</b>  |
| <b>5.2. Limpeza e Transformação de Dados .....</b>                     | <b>74</b>  |
| 5.2.1. Procedimentos Gerais .....                                      | 74         |
| 5.2.2. Tratamento de Valores por Preencher .....                       | 75         |
| 5.2.3. Normalização de Dados .....                                     | 75         |
| 5.2.4. Desenvolvimento da Dimensão Calendário e Dimensão Hora .....    | 76         |
| <b>5.3. Importação de Dados no Power BI .....</b>                      | <b>76</b>  |
| <b>6. Desenvolvimento dos Dashboards e Análise de Resultados .....</b> | <b>77</b>  |
| <b>6.1. Conceção dos Dashboards: Mockups .....</b>                     | <b>77</b>  |
| <b>6.2. Implementação dos Dashboards .....</b>                         | <b>78</b>  |
| <b>6.3. Funcionalidades Transversais.....</b>                          | <b>83</b>  |
| 6.3.1. Navegação e Acesso às Páginas .....                             | 84         |
| 6.3.2. Filtros e Segmentadores Globais .....                           | 84         |
| 6.3.3. Interações entre visualizações .....                            | 84         |
| 6.3.4. Controlo Temporal .....   | 86         |
| 6.3.5. Apoio à Exploração da Informação .....                          | 86         |
| 6.3.6. Tooltips e Informação de Apoio.....                             | 87         |
| <b>6.4. Conjunto de Medidas e Indicadores .....</b>                    | <b>89</b>  |
| <b>6.5. Estruturas de Apoio à Análise .....</b>                        | <b>90</b>  |
| <b>6.6. Interatividade e Uso de Bookmarks.....</b>                     | <b>92</b>  |
| <b>6.7. Análise de Resultados.....</b>                                 | <b>93</b>  |
| 6.7.1. Visão Global do Negócio.....                                    | 93         |
| 6.7.2. Desempenho dos Produtos .....                                   | 94         |
| 6.7.3. Desempenho Logístico e Cumprimento das Entregas .....           | 97         |
| 6.7.4. Análise do Comportamento dos Clientes.....                      | 100        |
| 6.7.5. Análise dos Vendedores.....                                     | 103        |
| 6.7.6. Avaliações dos Clientes e Qualidade do Serviço .....            | 105        |
| 6.7.7. Síntese da Análise Global .....                                 | 106        |
| <b>7. Conclusão e Trabalho Futuro .....</b>                            | <b>109</b> |
| <b>Bibliografia.....</b>   | <b>110</b> |
| <b>Anexos .....</b>  | <b>117</b> |
| <b>Anexo A – Estrutura dos Conjuntos de Dados .....</b>                | <b>117</b> |
| <b>Anexo B – Análise Gráfica das Variáveis Numéricas .....</b>         | <b>121</b> |

|  |            |
|--|------------|
| <b>Anexo C – Modelo Dimensional.....</b>                       | <b>123</b> |
| <b>Anexo D – Modelo de Dados Implementado no Power BI.....</b> | <b>124</b> |
| <b>Anexo E – Mockups dos Dashboards.....</b>                   | <b>125</b> |

# Lista de Figuras

|   |    |
|---|----|
| Figura 1. A Pirâmide DIKW (extraído de Rowley (2007)).....  | 6  |
| Figura 2. Arquitetura proposta de <i>Business Intelligence</i> (extraído de Ong et al. (2011)).....   | 8  |
| Figura 3. Diagramas de Técnica e de Dados criados no BiDaML para o exemplo de análise de tráfego (extraído de Khalajzadeh et al. (2020))..... | 12 |
| Figura 4. Evolução das arquiteturas analíticas: DW, DL e <i>Lakehouse</i> (extraído de Armbrust et al. (2021))..                              | 15 |
| Figura 5. <i>Pipeline</i> de visualização de informação (extraído de Qin et al. (2020)).....  | 19 |
| Figura 6. Magic Quadrant for Analytics and Business Intelligence Platforms (extraído de Ganeshan et al. (2025)).....                          | 23 |
| Figura 7. Exemplo com aplicação dos cinco princípios de <i>data storytelling</i> (A–E) (extraído de Echeverria et al. (2018)) .....           | 27 |
| Figura 8. <i>Dashboard</i> de <i>e-commerce</i> em Power BI com filtros interativos (extraído de Bhargava et al. (2023)) .....                | 33 |
| Figura 9. Coffee Circle: <i>data storytelling</i> vs. tradicional (extraído de Karampournioti & Wiedmann (2021))                              | 35 |
| Figura 10. Correlação entre o preço do produto e o valor dos portes de envio.....   | 52 |
| Figura 11. Correlação entre Dimensões Físicas dos Produtos .....  | 53 |
| Figura 12. Correlação entre o valor do pagamento e o número de prestações.....  | 54 |
| Figura 13. Correlação entre a pontuação das avaliações e o tempo de entrega .....   | 54 |
| Figura 14. Distribuição dos estados dos pedidos.....  | 55 |
| Figura 15. Distribuição dos métodos de pagamento .....  | 56 |
| Figura 16. Evolução mensal do número de encomendas entre 2016 e 2018 .....  | 56 |
| Figura 17. Tempo médio de entrega mensal (em dias) .....  | 57 |
| Figura 18. Top 10 Categorias de Produto por Vendas e Satisfação .....   | 58 |
| Figura 19. Distribuição dos 10 estados com maior número de clientes registados .....  | 59 |
| Figura 20. Distribuição dos 10 estados com maior número de encomendas realizadas .....  | 59 |
| Figura 21. Nuvem de palavras das avaliações positivas e negativas dos clientes .....  | 60 |
| Figura 22. Modelo relacional original do <i>dataset</i> Olist.....  | 69 |
| Figura 23. Arquitetura geral do sistema de BI da Olist.....   | 72 |
| Figura 24. Excerto do modelo de dados desenvolvido no Power BI .....  | 72 |
| Figura 25. Normalização dos nomes de cidades.....   | 75 |

|   |     |
|---|-----|
| Figura 26. <i>Dashboard</i> Sumário .....   | 79  |
| Figura 27. <i>Dashboard</i> Produtos.....   | 80  |
| Figura 28. <i>Dashboard</i> Logística .....   | 81  |
| Figura 29. <i>Dashboard</i> Clientes.....   | 82  |
| Figura 30. <i>Dashboard</i> Vendedores .....  | 82  |
| Figura 31. <i>Dashboard</i> Avaliações .....  | 83  |
| Figura 32. Filtro global de ano .....   | 84  |
| Figura 33. Segmentador de categorias .....  | 84  |
| Figura 34. Interação entre visualizações e filtros no painel Produtos.....                                | 85  |
| Figura 35. Alternância entre modos de visualização (gráfico e mapa).....                                  | 86  |
| Figura 36. Exemplo de controlo temporal num gráfico.....  | 86  |
| Figura 37. Paginação dinâmica aplicada ao gráfico de lucro por categoria .....                            | 87  |
| Figura 38. Filtro textual de subcategorias no painel Produtos .....                                       | 87  |
| Figura 39. <i>Tooltip</i> personalizada com informação adicional de análise temporal.....                 | 88  |
| Figura 40. <i>Tooltip</i> personalizada com informação detalhada por categoria/subcategoria .....         | 88  |
| Figura 41. <i>Tooltip</i> padrão na distribuição de encomendas por método de pagamento .....              | 89  |
| Figura 42. Recorrência de compras por produto.....  | 91  |
| Figura 43. Distribuição de produtos por faixa de preço .....  | 92  |
| Figura 44. Alternância de perspetivas de análise no gráfico temporal com recurso a <i>bookmarks</i> ..... | 92  |
| Figura 45. Receita mensal e pico em novembro de 2017 .....  | 94  |
| Figura 46. Evolução mensal das unidades vendidas, com pico em novembro de 2017.....                       | 95  |
| Figura 47. Distribuição do lucro total por categoria de produto .....                                     | 96  |
| Figura 48. Recorrência de compras por produto.....  | 96  |
| Figura 49. Distribuição das unidades vendidas por faixa de preço .....                                    | 97  |
| Figura 50. Evolução temporal da taxa de atraso nas entregas.....  | 98  |
| Figura 51. Taxa de atraso nas entregas por estado, com destaque para São Paulo e Alagoas.....             | 99  |
| Figura 52. Taxa de atraso e tempo médio de entrega por categoria .....                                    | 99  |
| Figura 53. Perfil etário dos clientes e respetiva receita média .....                                     | 101 |
| Figura 54. Distribuição da média de encomendas por dia do mês .....                                       | 101 |
| Figura 55. Volume de encomendas por dia da semana e período do dia .....                                  | 102 |

|   |     |
|---|-----|
| Figura 56. Participação dos métodos de pagamento no total de encomendas .....                         | 102 |
| Figura 57. Top 15 empresas por lucro .....  | 104 |
| Figura 58. Evolução temporal do lucro dos principais vendedores .....                                 | 105 |
| Figura 59. Relação entre o cumprimento do prazo de entrega e a avaliação atribuída pelo cliente ..... | 106 |
| Figura 60. Esquema do processo analítico desenvolvido no projeto.....                                 | 108 |

# Lista de Tabelas

|  |    |
|--|----|
| Tabela 1. Tipos de gráficos e respetivas características (adaptado de Tucakov (2025)). | 21 |
| Tabela 2. Tipos de <i>Dashboard</i> . (adaptado de Hayward (2024))                     | 42 |
| Tabela 3. Resumo estrutural dos ficheiros do <i>dataset</i> Olist.                     | 48 |
| Tabela 4. Distribuição de valores nulos por coluna                                     | 49 |
| Tabela 5. Validação de integridade referencial.  | 50 |
| Tabela 6. Estatísticas descritivas das variáveis numéricas relevantes                  | 51 |
| Tabela 7. Perguntas de negócio que suportam a definição de KPIs e <i>dashboards</i> .  | 62 |
| Tabela 8. Definição de <i>User Stories</i> para o Desenvolvimento do Sistema de BI.    | 62 |
| Tabela 9. Requisitos Funcionais  | 64 |
| Tabela 10. Requisitos Não Funcionais   | 65 |
| Tabela 11. KPIs e métricas relevantes para o sistema de BI da Olist                    | 65 |

## Lista de siglas e acrónimos

|         |  |
|---------|--|
| API     | Application Programming Interface              |
| AIPO    | Average Items per Order                        |
| AOV     | Average Order Value                            |
| AR      | Realidade Aumentada                            |
| ARPC    | Average Revenue per Customer                   |
| ARPV    | Average Revenue per Visitor                    |
| B2A     | Business-to-Administration                     |
| B2B     | Business-to-Business                           |
| B2C     | Business-to-Consumer                           |
| BI      | Business Intelligence                          |
| BiDaML  | Big Data Analytics Modeling Languages          |
| BIP-Map | Business Intelligence Product and Metadata Map |
| BPMN    | Business Process Modeling Notation             |
| C2A     | Consumer-to-Administration                     |
| C2B     | Consumer-to-Business                           |
| C2C     | Consumer-to-Consumer                           |
| CAC     | Customer Acquisition Cost                      |
| CLV     | Customer Lifetime Value                        |
| CM      | Contribution Margin                            |
| CNN     | Convolutional Neural Network                   |
| COGS    | Cost of Goods Sold                             |
| CPA     | Cost per Acquisition                           |
| CPC     | Cost per Click                                 |
| CR      | Conversion Rate                                |
| CRM     | Customer Relationship Management               |
| CTR     | Click-through Rate                             |
| DIKW    | Data, Information, Knowledge, Wisdom           |
| DL      | Data Lake                                      |
| DW      | Data Warehouse                                 |

|        |  |
|--------|--|
| EDA    | Exploratory Data Analysis                |
| ERP    | Enterprise Resource Planning             |
| ESTG   | Escola Superior de Tecnologia e Gestão   |
| ETL    | Extraction, Transform, Load              |
| IA     | Inteligência Artificial                  |
| IP-Map | Information Product Map                  |
| IQR    | Interquartile Range                      |
| KPI    | Key Performance Indicator                |
| ML     | Machine Learning                         |
| NCM    | Net Customer Margin                      |
| NLP    | Natural Language Processing              |
| NP     | Net Profit                               |
| OLAP   | Online Analytical Processing             |
| OLTP   | Online Transaction Processing            |
| OTIF   | On-Time In-Full                          |
| PYTD   | Previous Year-To-Date                    |
| ROAS   | Return on Advertising                    |
| ROI    | Return of Investment                     |
| RV     | Realidade Virtual                        |
| SPSN   | Situation, Problem, Solution, Next Steps |
| TRI    | Teoria de Resposta ao Item               |
| YTD    | Year-To-Date                             |

# 1. Introdução

O aumento da quantidade de dados disponíveis nas organizações tem vindo a reforçar a importância de abordagens eficazes para a sua análise e interpretação. Neste contexto, o BI assume um papel central ao permitir integrar, analisar e visualizar dados com o objetivo de apoiar a tomada de decisão (Chaudhuri et al., 2011; Rangam, 2009). Contudo, a simples apresentação de indicadores e *dashboards* nem sempre garante uma compreensão clara dos resultados por parte dos decisores, sobretudo quando não existe uma estrutura narrativa que oriente a interpretação da informação (Aparicio et al., 2015; Unwin, 2020).

O *data storytelling* surge como uma abordagem complementar, ao combinar dados, visualização e narrativa para comunicar *insights* de forma estruturada e orientada para objetivos, facilitando a compreensão e a retenção da informação (Echeverria et al., 2018; Ren et al., 2023). A sua aplicação no contexto do *e-commerce* revela-se particularmente relevante, dada a complexidade e diversidade dos dados associados a vendas, logística, comportamento dos clientes e qualidade do serviço (Jain et al., 2021).

## 1.1. Problema

Apesar da utilização crescente de ferramentas de BI, muitas organizações enfrentam dificuldades na comunicação eficaz dos resultados analíticos. A apresentação isolada de gráficos e métricas pode dificultar a identificação de padrões, tendências e relações relevantes, reduzindo o impacto da análise no processo de decisão (Dykes, 2019; Herschel & Clements, 2017).

No contexto do *e-commerce*, onde o desempenho do negócio resulta da interação entre múltiplos fatores, torna-se necessário assumir uma abordagem integrada que permita não apenas analisar os dados, mas também estruturar e comunicar os resultados de forma clara e orientada à decisão (Boldsova & Luoto, 2020; Pan et al., 2021).

O presente trabalho tem como base um caso de estudo no contexto do *e-commerce*, recorrendo a um conjunto de dados reais utilizado em projetos académicos e de investigação na área de BI. O *dataset* utilizado corresponde ao *e-commerce* brasileiro Olist, que agrega

informação sobre encomendas, produtos, clientes, vendedores, pagamentos, logística e avaliações dos clientes. Este conjunto de dados permite analisar diferentes dimensões do negócio e constitui um suporte para o desenvolvimento de uma solução de análise e comunicação de resultados baseada em *data storytelling* (Sionek, 2018).

Contudo, o *dataset* apresenta diversas limitações estruturais que justificam a necessidade de processos de preparação e modelação de dados. Entre as principais lacunas identificadas destacam-se:

- a ausência de uma dimensão calendário estruturada, que permita análises temporais;
- a inexistência de uma dimensão hora que suporte análises mais granulares do comportamento das encomendas;
- a necessidade de normalização e reorganização de dados distribuídos por múltiplos ficheiros relacionais;
- a presença de valores em falta e inconsistências que exigem tratamento prévio.

Estas limitações evidenciam que a simples disponibilização de dados não é suficiente para suportar uma análise estruturada, reforçando a importância de processos de integração, transformação e modelação no contexto de soluções de BI.

## 1.2. Objetivo

O objetivo geral deste trabalho consiste na conceção e implementação de uma solução de BI aplicada ao contexto do *e-commerce*, estruturada segundo princípios de modelação analítica, visualização de informação e *data storytelling*, de forma a apoiar a tomada de decisão organizacional.

Mais concretamente, o trabalho combina fundamentos teóricos de BI com a construção de uma arquitetura analítica que integra processos de integração, limpeza, normalização e modelação dimensional dos dados, assegurando a sua organização numa perspetiva orientada à análise multidimensional. Paralelamente, procede-se à definição de indicadores-chave de desempenho alinhados com diferentes níveis de decisão e ao desenvolvimento de *dashboards* interativos concebidos com base em princípios de hierarquia visual e coerência narrativa.

Pretende-se, assim, estruturar a análise descritiva do desempenho do negócio, evidenciando relações entre variáveis comerciais, operacionais e relacionais. O trabalho procura ainda ultrapassar as limitações identificadas ao nível da estrutura e organização dos dados, promovendo uma análise descritiva mais eficiente com uma narração orientada dos resultados.

Embora o *dataset* analisado seja utilizado como caso de estudo, pretende-se que a abordagem desenvolvida sirva como referência para projetos semelhantes em *marketplaces* digitais, demonstrando a sua aplicabilidade a outros contextos organizacionais com características semelhantes.

### **1.3. Organização do Documento**

O presente documento encontra-se estruturado em sete capítulos, organizados de forma sequencial desde o enquadramento teórico até à implementação e análise dos resultados da solução desenvolvida.

O Capítulo 2 apresenta o estudo prévio, estabelecendo o enquadramento conceptual do trabalho através da abordagem à pirâmide DIKW, aos fundamentos de BI, à visualização de informação, ao *data storytelling* e ao contexto analítico do *e-commerce*.

O Capítulo 3 descreve a metodologia seguida no caso de estudo, incluindo o enquadramento geral, a caracterização dos dados de origem e as ferramentas e tecnologias utilizadas.

O Capítulo 4 dedica-se ao desenho do sistema, iniciando com a análise exploratória dos dados e prosseguindo com a definição de requisitos, *Key Performance Indicators* (KPIs), especificação das narrativas analíticas, modelação dimensional e definição da arquitetura do sistema de BI.

O Capítulo 5 apresenta o processo de preparação, limpeza e transformação dos dados, detalhando os procedimentos gerais, o tratamento de valores em falta e os processos de normalização implementados para garantir consistência e qualidade da informação.

O Capítulo 6 descreve o desenvolvimento dos *dashboards*, as funcionalidades transversais implementadas e a análise dos resultados segundo uma abordagem de *data storytelling*.

Por fim, o Capítulo 7 apresenta as conclusões do trabalho e propõe possíveis linhas de investigação futura.

## 2. Estudo prévio

Este capítulo tem como objetivo realizar um estudo sobre conceitos associados ao projeto. Entre outros, são abordados conceitos de BI, de Visualização de Dados, de *Data Storytelling* e *E-commerce*. Posteriormente, estes conceitos são integrados, evidenciando de que forma se complementam e contribuem para o desenvolvimento de soluções de análise de dados de suporte à tomada de decisão em contextos digitais.

### 2.1. Dos Dados à Sabedoria: A Pirâmide DIKW

As organizações são orientadas por dados, pelo que, torna-se essencial compreender o percurso que os transforma em conhecimento útil e em sabedoria. Esta progressão é geralmente representada através do modelo hierárquico DIKW (*Data, Information, Knowledge, Wisdom*), reconhecido na literatura da ciência da informação e gestão do conhecimento como um modelo de referência para descrever as transformações sucessivas dos dados para níveis superiores de valor agregado e cognitivo (Rowley, 2007).

Segundo Ackoff (1989), o modelo DIKW organiza-se como uma hierarquia cognitiva. Os dados são definidos como símbolos isolados, que por si só ainda não têm significado. A informação surge quando esses dados são processados para responder a perguntas como "quem?", "o quê?" ou "quando?". O conhecimento aparece quando essa informação, após reflexão, é aplicada de forma prática e em contextos específicos. No topo da hierarquia está a sabedoria, entendida como a capacidade de agir com eficácia, guiada pela experiência (Ackoff, 1989).

Fricke (2019) complementa esta perspetiva ao afirmar que a informação só adquire valor quando os dados são selecionados com base na sua relevância e enriquecidos por processos analíticos. O conhecimento surge da aplicação contextualizada dessa informação, enquanto a sabedoria, situada no topo da hierarquia, requer análise crítica princípios éticos e capacidade de adaptação a diferentes contextos, características que a distinguem qualitativamente das etapas anteriores (Fricke, 2019).

A Figura 1 ilustra a pirâmide DIKW, evidenciando que, à medida que se sobe na hierarquia, aumenta o nível de abstração, complexidade e valor da informação, passando de dados brutos para representações progressivamente mais agregadas e interpretativas. Embora este modelo

apresente uma visão linear da passagem dos dados até à sabedoria, na prática, esse processo é dinâmico e envolve fatores adicionais, como o contexto, a experiência e a capacidade de análise crítica.

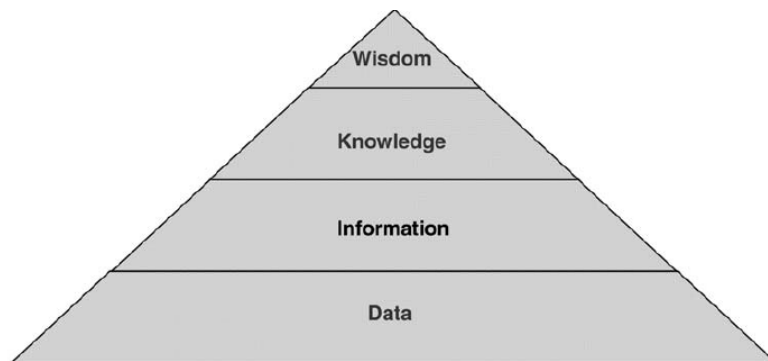


Figura 1. A Pirâmide DIKW (extraído de Rowley (2007))

Para além da dimensão estrutural e cognitiva, é fundamental considerar também a dimensão social e interpretativa do processo de transformação em sabedoria. A existência de sabedoria só promove valor real aos dados quando há transmissão efetiva do seu significado e conteúdo. Neste contexto, McDowell (2021) propõe uma abordagem narrativa à hierarquia DIKW, argumentando que a sabedoria não resulta apenas da acumulação analítica, mas também da troca de experiências, da interpretação em grupo e da criação de significados partilhados.

Por isso, entender esse percurso é importante para perceber como os sistemas de informação e os processos organizacionais podem ajudar na tomada de decisões mais conscientes e éticas (McDowell, 2021).

## 2.2. Conceitos de Business Intelligence

*Business Intelligence* (BI) é um conjunto de métodos e tecnologias que ajudam as empresas a recolher, armazenar e analisar dados para apoiar a tomada de decisão, melhorando a competitividade (Chaudhuri et al., 2011; Ranjan, 2009). O termo BI foi introduzido por Howard Dresner, do Gartner Group, em 1989, mas as suas bases já existiam em sistemas anteriores de apoio à decisão. (Negash & Gray, 2008).

Com o tempo, O BI evoluiu com tecnologias como *Data Warehousing*, *Online Analytical Processing* (OLAP) e *Data Mining*, permitindo transformar informação em conhecimento útil (Negash & Gray, 2008). Mais recentemente, a integração com Inteligência Artificial (IA)

e *Machine Learning* (ML) tem contribuído para tornar a análise de dados mais eficiente, melhorando a tomada de decisões (Praful Bharadiya, 2023).

O BI é essencial para otimizar operações e prever tendências, ajudando as empresas a adaptarem-se rapidamente às mudanças do mercado (Ranjan, 2009).

A arquitetura de BI é composta por diferentes camadas que estruturam o fluxo de dados, permitindo a sua extração, transformação e disponibilização para análise (Ong et al., 2011). Estas camadas organizam a informação desde a origem até à apresentação, garantindo qualidade e integridade (Zimmer et al., 2012). Desta forma, os dados são processados de forma eficiente e estruturada, permitindo a obtenção de *insights* para suporte estratégico das empresas (Lopes et al., 2020).

A estrutura da arquitetura de BI assenta em diversas camadas fundamentais sendo sucintamente (Ong et al., 2011):

- Camada de Fonte de Dados;
- Camada ETL (Extract, Transform, Load);
- Camada de Apresentação;
- Camada de Metadados.

A Figura 2 ilustra a arquitetura proposta de BI, representando as diferentes camadas funcionais do sistema, e evidenciando os fluxos de dados e metadados entre os seus diversos componentes.

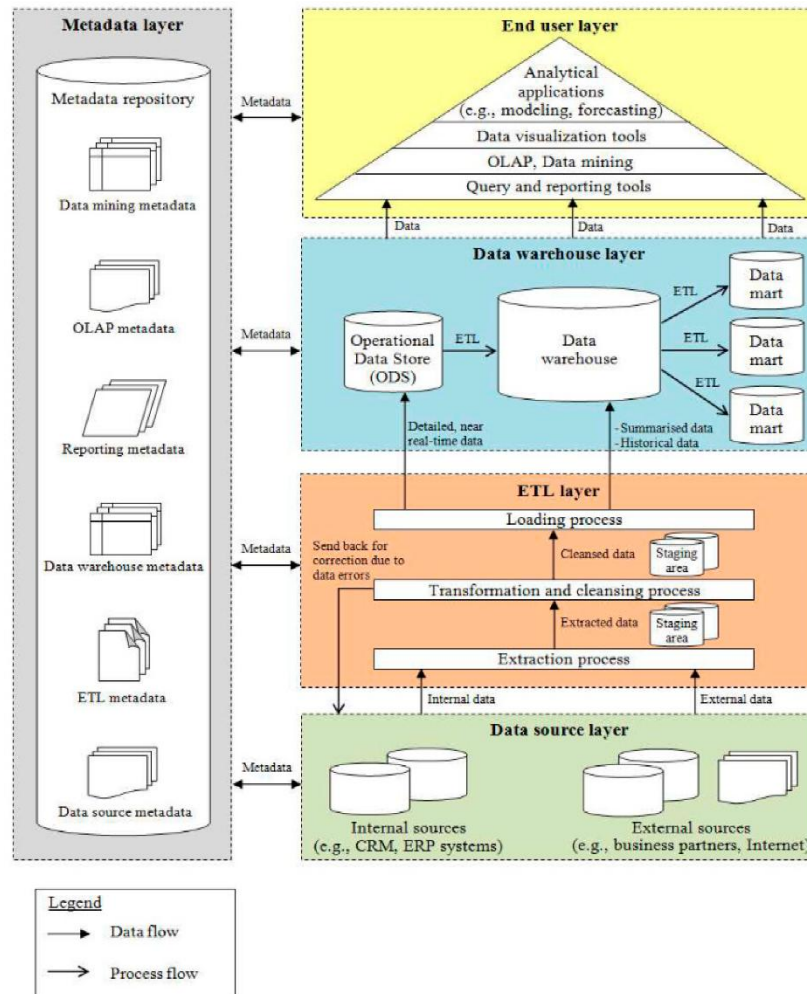


Figura 2. Arquitetura proposta de *Business Intelligence* (extraído de Ong et al. (2011))

De seguida, cada uma dessas camadas será descrita em maior detalhe, permitindo compreender o papel específico que cada uma desempenha no processo global de gestão e transformação de dados dentro de um sistema de BI.

### 2.2.1. Camada de Fonte de Dados

Esta camada representa as diversas fontes, como base de dados transacionais, sistemas *Enterprise Resource Planning* (ERP) e *Customer Relationship Management* (CRM), e fontes externas como redes sociais e *Application Programming Interface* (APIs). Identificar corretamente as fontes é essencial para evitar inconsistências e redundâncias (Ong et al., 2011).

### 2.2.2. Camada ETL

A camada ETL desempenha um papel importante na arquitetura de BI, sendo responsável por garantir a movimentação eficiente e estruturada dos dados desde as fontes originais até

ao armazenamento final. Segundo Ong et al. (2011), esta camada envolve três processos principais: extração, transformação e carregamento. Cada uma dessas etapas cumpre uma função específica, que visa assegurar a qualidade, consistência e disponibilidade dos dados para análise posterior (Ong et al., 2011). A seguir, descrevem-se em detalhe as atividades envolvidas em cada uma dessas fases.

**Extração:** A etapa de extração consiste na recolha de dados de diversas fontes, sendo um dos principais desafios garantir a compatibilidade entre diferentes formatos e estruturas (El-Sappagh et al., 2011).

As fontes de dados podem incluir bases de dados relacionais, ficheiros de texto e logs (El-Sappagh et al., 2011).

**Transformação:** A etapa de transformação é responsável por converter os dados extraídos em formatos compatíveis com o modelo do *Data Warehouse* (DW), assegurando a sua limpeza, padronização e integração (El-Sappagh et al., 2011).

Este processo inclui operações como conversões de tipos de dados, junções, agregações, tratamentos de valores nulos e aplicação de regras de negócio (A. Sharma, 2010).

A transformação garante que os dados estejam completos, consistentes e prontos para análise, sendo considerada uma das fases mais críticas do ETL (El-Sappagh et al., 2011).

**Carregamento:** A fase de carregamento representa o passo final do processo ETL, no qual os dados extraídos e transformados são inseridos nas estruturas multidimensionais do DW, tornando-os acessíveis para os utilizadores e sistemas de análise (El-Sappagh et al., 2011).

Esta etapa pode incluir a inserção de dados em tabelas de factos e dimensões e exige que os dados estejam num formato compatível com o modelo dimensional (A. Sharma, 2010).

Adicionalmente, o processo ETL pode funcionar de forma total (*full load*) ou incremental, dependendo dos requisitos de atualização e sincronização do sistema (El-Sappagh et al., 2011).

### 2.2.3. Camada de Armazenamento

A camada de armazenamento é responsável por manter os dados após a sua transformação, disponibilizando-os para análise e apoio à decisão (Garani et al., 2019).

Em inúmeras situações são usados DWs para armazenar dados estruturados, integrados e organizados por assunto (Ong et al., 2011). Estes dados são otimizados para consultas analíticas e suportam aplicações como OLAP e *Data Mining* (Garani et al., 2019).

De acordo com Garani et al. (2019), os DWs podem ser entendidos como bases de dados multidimensionais que organizam a informação em tabelas de factos, que registam os eventos relevantes para o negócio, e tabelas de dimensão, que descrevem e contextualizam esses factos. Esta estrutura é frequentemente implementada através de esquemas como o esquema em estrela ou o floco de neve, que permitem análises eficientes e operações típicas de OLAP, como agregações e comparações temporais.

Ainda segundo os autores, a utilização deste tipo de modelação é fundamental para permitir consultas analíticas eficientes, explorando operações típicas de OLAP, como agregações temporais e comparações entre diferentes perspetivas do negócio.

Importa ainda salientar que existe uma distinção fundamental entre os sistemas de suporte a operações, denominados OLTP (*Online Transaction Processing*), e os sistemas de suporte à decisão, denominados OLAP (*Online Analytical Processing*). Os primeiros são concebidos para gerir transações em tempo real, com estruturas normalizadas que garantem a integridade e a consistência dos dados, estando otimizados para o processamento de um elevado número de operações de inserção, atualização e eliminação que envolvem um pequeno número de registos. Nestes sistemas, os dados históricos são pouco relevantes, uma vez que o objetivo é assegurar a execução rápida e fiável de processos operacionais. Em contrapartida, os sistemas OLAP são concebidos para consolidar grandes volumes de dados, incluindo históricos, permitindo análises complexas e imprevisíveis que recorrem frequentemente a agregações e comparações multidimensionais. Assim, enquanto o OLTP responde às necessidades transacionais do dia a dia, o OLAP constitui a base para a análise estratégica, suportando a extração de conhecimento e a tomada de decisão (M Kirmani, 2017).

Neste contexto, destaca-se o papel do *Star Schema*, considerado a forma mais utilizada de modelação dimensional para suportar sistemas OLAP. Este modelo caracteriza-se por uma tabela central de factos, que contém as métricas de negócio e as chaves que estabelecem a ligação às dimensões, rodeada por um conjunto de tabelas de dimensão que descrevem e contextualizam os factos. As dimensões são frequentemente desnormalizadas, simplificando o modelo e otimizando a execução de consultas, o que facilita a compreensão da estrutura por parte dos utilizadores. Para além disso, o *Star Schema* apresenta vantagens significativas

em termos de flexibilidade e extensibilidade, uma vez que a adição de novas dimensões ou factos pode ser feita sem comprometer a estrutura já existente, garantindo a consistência dos dados e a continuidade das aplicações (M Kirmani, 2017).

#### **2.2.4. Camada de Apresentação**

A camada de apresentação é responsável por disponibilizar a informação aos utilizadores finais de forma acessível e orientada à tomada de decisão. É nesta camada que os dados transformados e armazenados ganham valor estratégico, através de recursos como *dashboards* (Ong et al., 2011).

De acordo com o modelo proposto por Ong et al. (2011), esta camada adapta-se às diferentes necessidades dos utilizadores, oferecendo uma visualização personalizada consoante o nível de responsabilidade, seja operacional, tático ou estratégico (Ong et al., 2011). Estudos mais recentes reforçam que a eficácia do BI depende fortemente do envolvimento dos utilizadores finais e da usabilidade das ferramentas (Jorgensen et al., 2021).

A adoção destas ferramentas está associada a fatores como a facilidade de uso, utilidade e suporte organizacional, conforme analisado por Tejero -Dakay et al. (2023).

Neste sentido, têm ganho relevância soluções orientadas ao utilizador que recorrem a ambientes visuais para apoiar a especificação e o desenho de problemas complexos de análise de dados. Estas abordagens são particularmente adequadas a sistemas que integram análise descritiva, preditiva e prescritiva, como os ecossistemas de BI combinados com Data Analytics, *Big Data* e ML (Khalajzadeh et al., 2020).

De acordo com Khalajzadeh et al. (2020), o desenvolvimento de soluções de análise de dados envolve equipas multidisciplinares e múltiplos níveis de abstração, desde conceitos de negócio de alto nível até detalhes técnicos de implementação. Neste contexto, o BiDaML (*Big Data Analytics Modeling Languages*) é proposto como um conjunto de linguagens visuais específicas de domínio que permite especificar, modelar e documentar soluções analíticas de forma integrada, cobrindo desde a definição de objetivos e requisitos de negócio até às fases de desenvolvimento, validação e *deployment*.

O BiDaML estrutura o processo analítico através de diferentes tipos de diagramas interligados, cada um correspondente a um nível de abstração distinto, permitindo representar de forma explícita os objetivos analíticos, os processos envolvidos, os dados utilizados, as técnicas aplicadas e os artefactos produzidos ao longo do ciclo de vida da solução, como ilustrado na Figura 3 (Khalajzadeh et al., 2020). Esta abordagem facilita a comunicação e a colaboração entre *stakeholders*<sup>1</sup> com diferentes perfis, como especialistas de domínio, analistas de negócio e engenheiros de software, promovendo uma linguagem comum para a definição e validação das soluções analíticas.

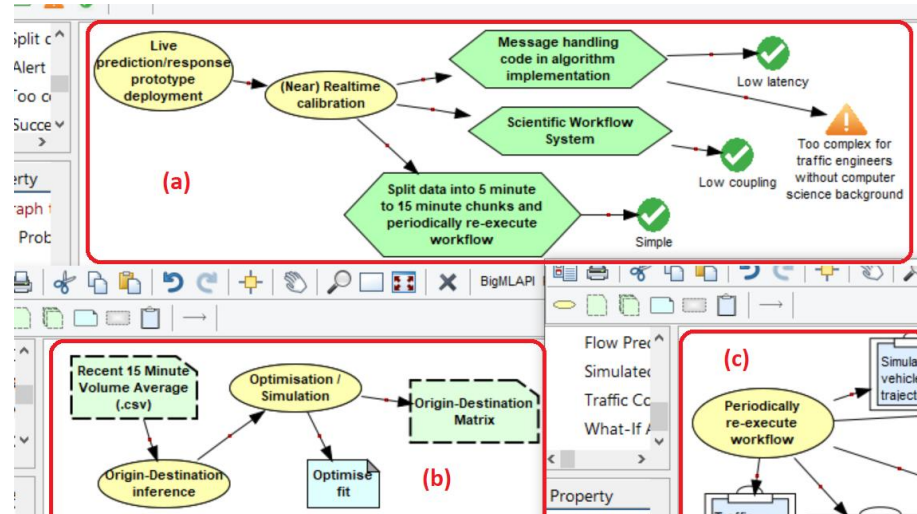


Figura 3. Diagramas de Técnica e de Dados criados no BiDaML para o exemplo de análise de tráfego (extraído de Khalajzadeh et al. (2020))

### 2.2.5. Camada de Metadados

A camada de metadados desempenha um papel fundamental numa arquitetura de BI, ao fornecer informação estruturada sobre a origem, significado, formato e transformações dos dados ao longo do sistema, assegurando a sua coerência e consistência no processo analítico (Ong et al., 2011). Esta camada suporta tanto utilizadores técnicos como de negócio na correta interpretação da informação disponível, promovendo decisões mais informadas e reduzindo o risco de erros de análise (Foshay et al., 2014). Assim, a utilização de metadados de definição, qualidade, linhagem e navegação é essencial para assegurar a coerência e a confiança na utilização das ferramentas de BI (Foshay et al., 2014).

Nesse contexto, o *framework Business Intelligence Product and Metadata Map* (BIP-Map) propõe uma estrutura integrada de metadados composta por três camadas: processos de

<sup>1</sup> *Stakeholders*: partes interessadas ou intervenientes com interesse direto ou indireto no problema ou decisão em análise.

negócio, de informação e componentes técnicos. Este *framework* facilita a compreensão e o acompanhamento dos produtos de BI, ao reunir diferentes tipos de metadados por meio da combinação de técnicas como *Business Process Model and Notation* (BPMN), *Information Product Map* (IP-Map) e gestão de metadados. A camada de processos de negócio modela os fluxos organizacionais e os participantes envolvidos. A camada de produção de informação detalha como os dados são capturados, processados e armazenados. Já a camada de metadados oferece informações técnicas e semânticas sobre os dados, como regras de validação, transformações e estruturas de armazenamento (Chee et al., 2011).

A centralização dos metadados num repositório específico facilita a governação dos dados, promovendo consistência e eficiência em todo o sistema de BI, garantindo um fluxo de dados claro e estruturado. Como resultado, os utilizadores finais podem confiar nos relatórios gerados e compreender a lógica subjacente à sua construção, promovendo a autonomia na análise da informação (Chee et al., 2011).

O BIP-Map, portanto, não só documenta os produtos de BI de uma organização, como também os liga de forma explícita aos metadados que lhes dão suporte, permitindo uma visão holística da cadeia de valor da informação.

### 2.3. Business Intelligence na era do Big Data

No contexto do *Big Data*, os dados caracterizam-se por um conjunto de propriedades que aumentam a complexidade da sua gestão e análise. De acordo com Ishwarappa & Anuradha (2015), estas propriedades são habitualmente descritas através dos 5Vs do *Big Data*: *Volume*, *Velocity*, *Variety*, *Veracity* e *Value*.

- **Volume (*Volume*):** refere-se à quantidade de dados gerados continuamente, frequentemente em escalas de *terabytes* ou *petabytes*, o que coloca desafios significativos ao armazenamento e processamento tradicional;
- **Velocidade (*Velocity*):** diz respeito à velocidade a que os dados são produzidos, transmitidos e necessitam de ser processados, exigindo, em muitos casos, capacidades de análise quase em tempo real;
- **Variedade (*Variety*):** reflete a diversidade de formatos e fontes de dados, incluindo dados estruturados, semiestruturados e não estruturados, provenientes de sistemas transacionais, sensores, redes sociais, entre outros;

- **Veracidade (*Veracity*):** está associada à qualidade, fiabilidade e consistência da informação, sendo influenciada por ruído, dados incompletos ou inconsistentes, o que pode comprometer a confiança nas análises realizadas;
- **Valor (*Value*):** corresponde à capacidade de extrair conhecimento útil e relevante a partir dos dados disponíveis, transformando grandes volumes de informação em suporte à tomada de decisão.

Estas características introduzem desafios aos sistemas tradicionais de BI, sobretudo no que diz respeito à integração, armazenamento e análise eficiente de grandes volumes de dados provenientes de fontes variadas.

Os DWs continuam a desempenhar um papel relevante na análise histórica e estruturada dos dados. No entanto, no contexto do Big Data, apresentam limitações significativas, sobretudo devido à dependência de esquemas rígidos (*schema-on-write*) e de processos tradicionais de ETL, que reduzem a flexibilidade e dificultam a ingestão de dados não estruturados ou provenientes de fontes em constante evolução (Armbrust et al., 2021).

Para responder a estas limitações, surgiram os *Data Lakes* (DLs), que permitem armazenar grandes volumes de dados no seu formato original, suportando *schema-on-read* e a ingestão de dados estruturados, semiestruturados e não estruturados. Contudo, apesar da sua flexibilidade e escalabilidade, os DLs apresentam fragilidades ao nível do controlo, da governação e do desempenho analítico (Armbrust et al., 2021).

Neste contexto, surgiu a arquitetura *Lakehouse*, que integra a flexibilidade e a capacidade de armazenamento dos DLs com os mecanismos de controlo, consistência e desempenho analítico característicos dos DWs. Os *Lakehouses* permitem suportar, numa única plataforma, cargas de trabalho de BI, *Data Science* e ML, assegurando simultaneamente mecanismos de governação e gestão de metadados essenciais para a qualidade da informação (Mazumdar et al., 2023).

A Figura 4 apresenta, da esquerda para a direita, a evolução das arquiteturas analíticas: DW tradicional, arquitetura combinada DL + DW e, por fim, *Lakehouse*.

Na primeira abordagem, os dados estruturados são processados por ETL e armazenados num DW, suportando essencialmente BI e reporting. Trata-se de uma arquitetura orientada a dados estruturados e análises históricas.

Na segunda abordagem, surge o DL, permitindo armazenar dados estruturados, semiestruturados e não estruturados no seu formato original. Nesta arquitetura, o DL e o DW funcionam em conjunto, suportando diferentes tipos de cargas analíticas, mas mantendo infraestruturas separadas.

Por último, a arquitetura *Lakehouse* integra armazenamento e processamento numa única plataforma, combinando a flexibilidade dos DLs com os mecanismos de controlo e desempenho analítico dos DWs, suportando simultaneamente BI, *Data Science* e ML.

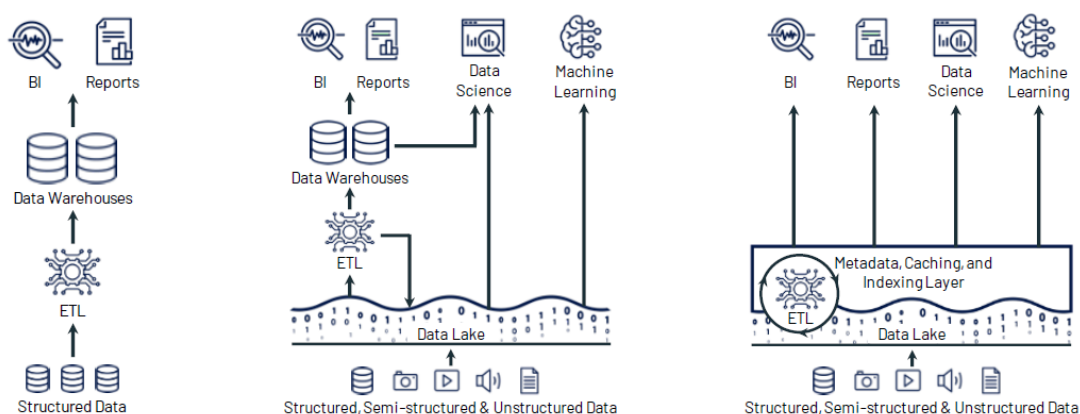


Figura 4. Evolução das arquiteturas analíticas: DW, DL e *Lakehouse* (extraído de Armbrust et al. (2021))

## 2.4. Representação Visual da Informação

A visualização de dados é fundamental para comunicar informação complexa de forma clara e acessível. De seguida, abordam-se conceitos fundamentais, princípios de design e tipos de gráficos. Posteriormente apresentam-se as ferramentas de visualização mais utilizadas de acordo com Aparicio et al. (2015).

### 2.4.1. Conceito e Importância da Visualização de Informação

A visualização de informação corresponde ao conjunto de técnicas que permitem transformar conjuntos complexos de dados em representações gráficas intuitivas, facilitando a identificação de padrões e tendências que, de outra forma, poderiam passar despercebidos (Islam & Jin, 2019). A sua relevância atual reflete-se na forma como influenciam decisões informadas em diversos setores, desde os negócios até à ciência, constituindo-se como uma ferramenta central em projetos de BI (Shadare et al., 2016). Adicionalmente, e de acordo com Unwin (2020), a visualização de informação é essencial na análise exploratória de

dados, pois permite detetar *outliers*, avaliar a qualidade dos dados e compreender as suas estruturas internas, especialmente nas fases iniciais da análise, quando ainda não existem hipóteses bem definidas (Unwin, 2020).

A visualização de informação desempenha um papel fundamental na comunicação dos resultados e, quando bem executada, permite transmitir de forma clara e acessível o conhecimento, mesmo a públicos com perfis variados (Aparicio et al., 2015). Combinando design gráfico, estatística e tecnologia, a construção de relatórios e *dashboards* torna os dados mais compreensíveis, facilitando o acesso à informação e apoiando a tomada de decisão na organização (Islam & Jin, 2019). Técnicas como gráficos de barras, de dispersão ou de linhas, aliadas ao uso adequado de cores e formas, facilitam a representação tanto de dados quantitativos como qualitativos (Aparicio et al., 2015).

A relevância da visualização de informação manifesta-se também nas suas aplicações práticas, nomeadamente na saúde pública, na área da energia e na deteção de fraude, ao facilitar a identificação de padrões (Shadare et al., 2016). Contudo, Unwin (2020) reforça que interpretar gráficos requer literacia visual. Por isso, a eficácia da visualização depende tanto da qualidade do gráfico quanto da contextualização da informação (Unwin, 2020).

#### 2.4.2. Fundamentos da Visualização de Informação

A visualização de informação envolve um processo de tradução visual em que as decisões de design assumem um papel determinante na forma como a informação é percebida e interpretada. A seleção do tipo de representação gráfica, bem como a utilização adequada de cores, formas e escalas, influencia diretamente a identificação de padrões, relações e anomalias nos dados, podendo potenciar ou limitar a eficácia comunicativa das representações visuais (Cleveland & McGill, 1984; Hullman & Diakopoulos, 2011).

É nesse contexto que os princípios da psicologia de *Gestalt* desempenham um papel importante na visualização de informação, pois explicam como o cérebro humano organiza estímulos visuais, facilitando o agrupamento e a leitura estrutural de elementos gráficos (Rosli & Cabrera, 2015). Entre os principais destacam-se (Rosli & Cabrera, 2015):

- **Proximidade (*Proximity*)** – Elementos próximos tendem a ser percebidos como pertencentes ao mesmo grupo, permitindo o agrupamento visual sem necessidade de marcadores explícitos;

- **Similaridade (*Similarity*)** – Objetos com características visuais semelhantes (como cor ou forma) quando são agrupados, favorecem a leitura categórica de conjunto de dados;
- **Fechamento (*Closure*)** – Quando figuras estão incompletas, o cérebro completa as lacunas, permitindo a identificação de padrões mesmo a partir de dados fragmentados;
- **Figura-Fundo (*Figure-Ground*)** – Este princípio estabelece a distinção entre o objeto em foco e o plano de fundo, essencial para direcionar a atenção do observador;
- **Unidade Percetiva (*Wholeness*)** – O sistema visual tende a integrar partes dispersas num todo. Este princípio reforça a ideia de que a visualização deve promover a compreensão global da estrutura dos dados, e não apenas a leitura isolada de elementos.

Estes princípios são processados de forma subconsciente e, quando aplicados de forma consciente no design de visualizações, aumentam significativamente a sua eficácia percetiva.

Segundo Shneiderman (2003), o princípio orientador da visualização de informação, defende a apresentação de uma visão geral seguida de funcionalidades de zoom e filtro, resultando na disponibilização de detalhes sob pedido. Esta abordagem orienta a construção de sistemas interativos e favorece uma exploração eficiente dos dados (Shneiderman, 2003). Complementando, Mackinlay (1986) propõe critérios de expressividade e efetividade para guiar a escolha de representações visuais com base nos tipos de dados e nos canais visuais mais adequados (posição, cor, tamanho) (Mackinlay, 1986).

Ware (2019) destaca que uma visualização mais eficaz é aquela que tira partido dos mecanismos naturais do cérebro para processar informação visual, como a segmentação, a atenção seletiva e a persistência percetiva. Esta última refere-se à capacidade de o cérebro manter brevemente uma imagem visual mesmo depois de esta desaparecer do campo de visão.

Contudo, a eficácia de uma visualização depende também da literacia visual dos seus utilizadores. Boy et al. (2014) desenvolveram testes baseados na Teoria de Resposta ao Item (TRI) para medir a capacidade de um indivíduo interpretar gráficos como linhas, barras e dispersões. Os autores destacam que a variabilidade na compreensão gráfica entre diferentes públicos pode comprometer a eficácia de sistemas visuais, sobretudo quando utilizadores em contextos críticos como saúde, política ou justiça (Boy et al., 2014). Paralelamente, Pandey

et al. (2015) alertam para os riscos associados a certas escolhas visuais, como o uso de escalas incompletas, distorções geométricas ou proporções inconsistentes, que podem induzir interpretações erradas, intencionais ou não (Pandey et al., 2015).

Um outro aspeto fundamental diz respeito ao papel retórico da visualização. Hullman & Diakopoulos (2011) argumentam que, em contextos narrativos, decisões editoriais como a ênfase visual, a exclusão seletiva de dados ou a utilização de metáforas gráficas podem orientar os utilizadores para interpretações específicas. Esta perspetiva aproxima a visualização das práticas jornalísticas e literárias, onde a narrativa desempenha um papel determinante na forma como a informação é evidenciada. Assim, o design da visualização deve procurar equilibrar a expressividade com responsabilidade ética, assegurando que os dados são representados de forma rigorosa e contextualizada (Hullman & Diakopoulos, 2011).

A avaliação da eficácia da visualização não deve limitar-se à análise de usabilidade. Plaisant (2004) defende uma abordagem mais abrangente, que inclua métricas de impacto, relevância prática e a aplicação efetiva em contextos reais de utilização. Estudos de caso, como os realizados com ferramentas de visualização interativas, mostram que a eficácia de uma visualização depende não apenas do conteúdo informativo que transmite, mas também da qualidade da experiência proporcionada ao utilizador. Estes estudos revelam, por exemplo, os benefícios obtidos pelos utilizadores, a frequência de utilização e a forma como a visualização contribui para decisões mais informadas (Plaisant, 2004).

### **2.4.3. Elementos Visuais Essenciais**

Os elementos visuais essenciais na visualização de informação, como a cor, a forma, a posição, a escala, o tipo de gráfico, a tipografia e a organização espacial constituem ferramentas fundamentais que influenciam diretamente a forma como a informação estatística é interpretada visualmente (Jiang, 2022). A aplicação estratégica destes componentes contribui para reduzir a carga cognitiva, facilitar a identificação de padrões e tornar a informação visualmente mais acessível (Jiang, 2022; Vanderplas et al., 2020).

A cor, em particular, não deve ser entendida apenas como um elemento decorativo. A sua aplicação influencia diretamente a legibilidade, o agrupamento perceptivo e a inclusão de diferentes perfis de utilizadores (Vanderplas et al., 2020). A investigação de Vanderplas et al. (2020) demonstra que a utilização simultânea de diferentes elementos visuais, como a cor

e a forma num gráfico de dispersão, facilita a identificação de padrões e torna as visualizações mais claras a utilizadores com limitações visuais ou dificuldades de processamento cognitivo, incluindo daltonismo e dislexia (Vanderplas et al., 2020). Esta perspetiva pode ser articulada com os princípios defendidos por Tufte (1983), que não rejeita o uso de elementos gráficos como a cor, desde que estes contribuam para a clareza, precisão e integridade da informação apresentada. Pelo contrário, o autor valoriza representações visuais que maximizem a leitura eficiente dos dados e defende que os elementos visuais devem ser utilizados com propósito informativo. Tufte (1983) alerta, no entanto, para os perigos da sobrecarga visual e do uso de chama a atenção para o excesso de elementos visuais que não acrescentam informação útil, referindo-se a estes como *chartjunk*. Estes elementos, segundo o autor, introduzem ruído e distraem da mensagem principal, comprometendo a compreensão dos dados. A sua proposta centra-se na maximização da *data-ink ratio*, promovendo uma visualização limpa e focada, na qual cada componente gráfico desempenha uma função clara na comunicação da informação.

Além da cor, atributos como a posição, o tamanho e a orientação gráfica atuam como elementos visuais que representam variáveis numéricas ou categóricas em formatos perceptivamente eficazes. O mapeamento destes atributos às variáveis constitui uma etapa central no *pipeline* de visualização, na qual se concretiza a transição entre os dados tratados e a sua representação visual. Esta fase exige uma correspondência adequada entre os objetivos analíticos e os canais visuais adotados, assegurando a clareza e a eficácia da comunicação gráfica. Conforme ilustrado na Figura 5, este processo integra uma sequência estruturada que inclui a importação, preparação, manipulação, mapeamento e visualização dos dados (Qin et al., 2020).

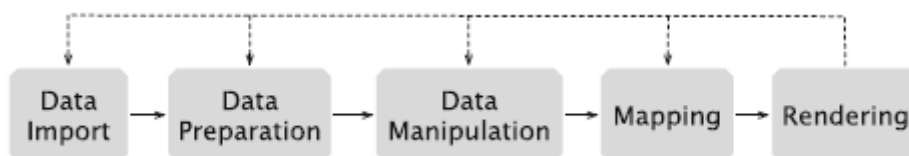


Figura 5. *Pipeline* de visualização de informação (extraído de Qin et al. (2020))

A escolha do tipo de gráfico representa também uma decisão importante no processo de visualização. Mais do que representar dados, um gráfico deve estruturar uma narrativa visual coerente. A gramática dos gráficos, permite articular dados, marcas e escalas com lógica e

consistência, consoante a complexidade e o propósito da análise em questão (Vanderplas et al., 2020).

Em contextos aplicados, como *dashboards* interativos, a utilização de padrões de design visual estruturados assume um papel determinante na eficácia da visualização. Soluções como a coordenação entre múltiplas vistas, a organização visual por categorias e a implementação de hierarquias de navegação contribuem para aumentar a legibilidade, orientar a exploração dos dados e apoiar a tomada de decisão em cenários analíticos complexos (Bach et al., 2023).

A eficácia da visualização depende ainda da sua clareza, autonomia e orientação para uma ideia principal. Green et al. (2023) salienta que uma boa visualização deve ser autoexplicativa e equilibrada, permitindo ao leitor apreender a mensagem de forma imediata. A seleção e combinação dos elementos visuais deve, por isso, ser guiada por critérios de simplicidade e consistência (Green et al., 2023).

#### **2.4.4. Tipos de Gráficos e as suas Aplicações**

A visualização de dados constitui uma abordagem essencial para explorar grandes volumes de informação, ao permitir que os dados sejam apresentados em formatos visuais que favorecem a análise, a descoberta de padrões e a comunicação eficaz dos resultados. A substituição de longas tabelas e listas por elementos gráficos contribui para uma compreensão mais rápida e acessível, especialmente em contextos onde a complexidade dos dados dificultaria a sua interpretação apenas em formato textual (Islam & Jin, 2019).

Entre os principais tipos de gráficos, os gráficos de linhas são especialmente eficazes para representar a evolução de dados ao longo do tempo.

Os gráficos de barras são fundamentais na comparação de categorias discretas e possuem variações como os formatos agrupados e empilhados, que facilitam a leitura de múltiplas variáveis simultaneamente (Luo et al., 2019).

Os histogramas são utilizados para representar a distribuição de frequência de variáveis contínuas, agrupando os dados em classes sem espaços entre as colunas (Slutsky, 2014). Para facilitar a interpretação dos dados, podem ser aplicadas técnicas como empilhamento, visualizações em fluxo ou linhas de tendência.

Gráficos de dispersão são utilizados na análise de relações entre duas variáveis numéricas, e podem ser enriquecidos com atributos visuais como cor, tamanho ou opacidade para representar múltiplas dimensões (Luo et al., 2019).

Já os gráficos circulares (“*pie charts*”) são recomendados para representar classes ou grupos de dados em proporção ao conjunto total. Cada fatia do gráfico corresponde a uma categoria e deve apresentar variações significativas. Em geral, recomenda-se limitar o número de categorias entre 3 e 10 para evitar confusão e manter a clareza visual. Quando bem utilizados, os gráficos circulares permitem visualizar de forma intuitiva a composição dos dados, sendo úteis para representar proporções (Slutsky, 2014).

Os diagramas de extremos e quartis (“*box plots*”), são utilizados para apresentar resumos estatísticos, como mediana, quartis, valores extremos e possíveis *outliers*. A estrutura visual deste tipo de gráfico permite ainda identificar o grau de dispersão dos dados e possíveis assimetrias na sua distribuição (Slutsky, 2014).

De seguida, apresenta-se a Tabela 1, na qual se encontra uma comparação dos principais tipos de gráficos utilizados em visualização de dados, juntamente com seus respectivos usos principais, características visuais, tipos de dados mais adequados e limitações observadas. Esta tabela foi elaborada com base nas classificações propostas por Tucakov (2025).

Tabela 1. Tipos de gráficos e respetivas características (adaptado de Tucakov (2025)).

| Tipo de Gráfico             | Uso principal                             | Caraterísticas Visuais   | Tipo de Dados Adequado | Limitações Observadas                                  |
|-----------------------------|---|--|------------------------|--|
| <b>Gráfico de Linhas</b>    | Tendências em séries temporais            | Pontos conectados por linhas indicam variações ao longo do tempo | Dados sequenciais      | Pode tornar-se confuso com muitas variáveis            |
| <b>Gráfico de Barras</b>    | Comparação entre categorias               | O comprimento das barras representa o valor                      | Dados categóricos      | Não é ideal para séries temporais ou muitas categorias |
| <b>Histograma</b>           | Distribuição de dados numéricos contínuos | Barras representam a frequência dos dados em intervalos          | Dados contínuos        | Não permite visualizar evolução temporal               |
| <b>Gráfico de dispersão</b> | Relação entre variáveis                   | Pontos representam   | Dados quantitativos    | Menos eficaz para dados categóricos                    |

|                                       |                            | pares de valores  | com duas variáveis  |   |
|---------------------------------------|----------------------------|---|---------------------|---|
| <b>Gráfico Circular</b>               | Proporções ou percentagens | Gráfico circular dividido em segmentos                                    | Relação parte-todo  | Pouco eficaz para comparação: ineficaz com muitos segmentos |
| <b>Diagrama de Extremos e Quartis</b> | Distribuição estatística   | Resumo visual dos quartis, mediana e valores extremos ( <i>outliers</i> ) | Dados quantitativos | Não mostra detalhadamente a distribuição completa dos dados |

#### 2.4.5. Ferramentas de Visualização

Num contexto marcado pelo aumento da complexidade e do volume de dados, as soluções de visualização tornaram-se essenciais para transformar dados em representações mais intuitivas, interativas e úteis para apoiar a tomada de decisão. Para além de facilitarem a identificação de padrões e tendências, estas soluções contribuem para tornar as análises mais acessíveis dentro das organizações, promovendo uma utilização mais alargada da informação (Islam & Jin, 2019).

Neste contexto, foram analisadas diversas plataformas de *Business Intelligence*, privilegiando aquelas posicionadas como líderes no setor e reconhecidas no mercado (Howson et al., 2019). As principais soluções consideradas foram (Howson et al., 2019):

- **Microsoft Power BI:** Plataforma de BI que disponibiliza funcionalidades de preparação de dados, exploração visual e *dashboards* interativos, podendo ser utilizada em ambiente *cloud* ou *on-premises* através do Power BI Report Server;
- **Tableau (Salesforce):** Plataforma de BI orientada à exploração visual e interativa de dados, permitindo que utilizadores criem, analisem e partilhem visualizações e *dashboards* de forma intuitiva. É utilizada tanto para análises realizadas diretamente por utilizadores de negócio como em contextos de BI corporativo, suportando a partilha de informação a partir de dados e modelos definidos;
- **Qlik Sense (Qlik):** Plataforma de BI baseada num motor associativo em memória, orientada à análise exploratória e ao desenvolvimento de aplicações analíticas

interativas. Suporta a integração de múltiplas fontes de dados e modelos complexos, sendo utilizada tanto por utilizadores finais como em contextos de BI corporativo;

- **Google Looker:** Plataforma de BI orientada para ambientes *cloud*, baseada na modelação de dados através da linguagem LookML e integrada com bases de dados analíticas modernas. A solução privilegia a definição de métricas e cálculos, bem como a reutilização destes em *dashboards* e aplicações analíticas, sendo utilizada tanto em cenários de análise realizada por utilizadores como em contextos de BI corporativo.

Este posicionamento é também evidenciado na Figura 6, que apresenta o *Magic Quadrant for Analytics and Business Intelligence Platforms*, onde o Power BI surge destacado como líder (Ganeshan et al., 2025).

Figure 1: Magic Quadrant for Analytics and Business Intelligence Platforms



Figura 6. Magic Quadrant for Analytics and Business Intelligence Platforms (extraído de Ganeshan et al. (2025))

## 2.5. Narrativas Baseadas em Dados

À medida que os dados se tornam mais complexos, é necessário comunicar de forma simples e eficaz. O *data storytelling* surge como uma abordagem que integra análise, visualização e narrativa para transmitir *insights*. Nos tópicos seguintes serão abordados o conceito, a

estrutura e os principais desafios desta prática, destacando o seu papel crescente na comunicação de dados.

### **2.5.1. Definição e Enquadramento**

O *data storytelling*, ou narrativa baseada em dados, tem-se afirmado como uma prática emergente que integra análise de dados, visualização e técnicas narrativas com o propósito de comunicar *insights* de forma clara, apelativa e impactante. Esta abordagem tem vindo a ganhar relevância à medida que aumentam a complexidade e o volume dos dados, tornando-se essencial para transmitir informações claras a públicos diversificados (Ren et al., 2023).

Do ponto de vista conceptual, distingue-se como uma forma adicional de visualização, ao integrar intencionalmente estruturas narrativas que desafiam expectativas e orientam a interpretação do público. Ao organizar os dados de forma estruturada, esta prática torna a informação mais acessível, envolvente e fácil de compreender, contribuindo para a aprendizagem e a retenção de conhecimento (Echeverria et al., 2018; Ren et al., 2023). Um estudo citado por Daradkeh & Atalla (2023) revela que apenas 5% das pessoas conseguem recordar estatísticas apresentadas de forma isolada, enquanto 63% retêm a informação quando esta é transmitida sob a forma de uma história, demonstrando o poder cognitivo das narrativas no contexto da comunicação de dados (Daradkeh & Atalla, 2023).

Segundo Matei & Hunter (2021), uma boa narrativa baseada em dados deve desafiar as expectativas do público, substituindo explicações habituais por causas inesperadas, mas fundamentadas, promovendo curiosidade, envolvimento emocional e aprendizagem ativa. Esta estrutura narrativa reforça a compreensão e facilita a construção de novos significados a partir dos dados apresentados (Matei & Hunter, 2021).

Adicionalmente, o processo de *data storytelling* é reconhecido como iterativo e composto por três etapas fundamentais: exploração dos dados, construção da narrativa e apresentação da história. Esta estrutura operacional exige competências multidisciplinares, que abrangem desde capacidades técnicas, como a análise e visualização de dados, até competências comunicacionais, nomeadamente a adaptação da mensagem ao público-alvo (Quenum et al., 2025).

### 2.5.2. Estrutura e Elementos do Data Storytelling

A estrutura do *data storytelling* assenta na articulação coordenada de três componentes fundamentais: os dados, a narrativa e as visualizações. Esta integração permite transformar informação técnica em mensagens claras, através de modelos que organizam e apresentam os conteúdos de forma estratégica (Daradkeh & Atalla, 2023).

Segundo os mesmos autores, o processo narrativo no *data storytelling* pode ser compreendido em três níveis distintos:

- **Modelo de análise:** responsável por extrair *insights* relevantes a partir dos dados não tratados;
- **Modelo de história:** organiza os *insights* numa sequência lógica e clara, adequada ao objetivo comunicacional;
- **Modelo de narração:** define o meio e o formato da entrega da história ao público, recorrendo a suportes como *dashboards*, relatórios ou vídeos.

Para assegurar uma narrativa completa e informativa, Daradkeh & Atalla (2023) propõem a utilização do modelo 5W1H, que integra seis elementos principais no processo de construção da história:

- **Quem? (Who)** – Identificar os *stakeholders* envolvidos e as suas motivações;
- **O quê? (What)** – Definir o problema central e os *insights* a comunicar;
- **Quando? (When)** - Estabelecer o intervalo temporal da análise, seja de curto ou longo prazo;
- **Onde? (Where)** – Determinar o enquadramento geográfico ou organizacional dos dados;
- **Porquê? (Why)** – Justificar a relevância da análise e o seu impacto na decisão;
- **Como? (How)** – Selecionar o formato e estilo narrativo mais eficaz para transmitir os dados.

Em alternativa, o modelo SPSN (*Situation, Problem, Solution, Next Steps*) organiza a narrativa em quatro etapas fundamentais (Tewari, 2019):

- **Situação (Situation)** - Apresentar o contexto atual e fornecer informações que enquadram o problema;

- **Problema** (*Problem*) - Identificar o desafio, destacando a sua relevância e impacto;
- **Solução** (*Solution*) – Propor soluções viáveis, explicando como podem ser aplicadas e os benefícios que proporcionam;
- **Próximos Passos** (*Next Steps*) - Indicar as ações seguintes a serem tomadas, incluindo cronogramas, recursos e responsabilidades.

Para complementar a estrutura lógica do SPSN, o modelo SUCCESS acrescenta um conjunto de princípios que reforçam a eficácia comunicacional da narrativa. Este modelo identifica fatores que tornam a comunicação mais clara e eficaz (Tewari, 2019):

- **Simplicidade** (*Simple*) - Focar na mensagem principal e evitar complexidade técnica desnecessária;
- **Surpresa** (*Unexpected*) - Evidenciar padrões esperados para captar a atenção do público;
- **Concretude** (*Concrete*) - Utilizar exemplos concretos que ajudem a visualizar a situação abordada;
- **Credibilidade** (*Credible*) - Apoiar a narrativa com dados ou validações externas que reforcem a confiança na mensagem;
- **Emoção** (*Emotional*) - Apelar a emoções humanas como medo, felicidade ou surpresa, tornando a história mais envolvente e relevante;
- **Histórias** (*Stories*) - Construir a narrativa com base numa estrutura clássica de história, incluindo herói, desafio e resolução, o que facilita a retenção da informação.

A aplicação destes princípios não se limita à construção textual da narrativa, estendendo-se também ao plano visual da comunicação de dados. Elementos como títulos orientados por mensagem, destaques cromáticos, anotações interpretativas e a eliminação de ruído visual são valorizados como estratégias que facilitam a leitura e a interpretação dos dados (Echeverria et al., 2018).

Conforme ilustrado na Figura 7, esta diferença é visível na comparação entre duas visualizações. À esquerda apresenta-se um gráfico tradicional, sem elementos narrativos. À direita observa-se uma versão com princípios de *data storytelling*, na qual se destaca um título descritivo, o realce visual da linha principal, a eliminação de elementos irrelevantes e a inclusão de anotações explicativas que orientam a interpretação. Esta versão segue os cinco princípios fundamentais para reforçar a comunicação visual: mensagem principal no título

(A), destaque cromático (B), evidência de dados-chave (C), remoção de elementos não informativos (D) e anotação contextual (E), conforme proposto por Echeverria et al. (2018).

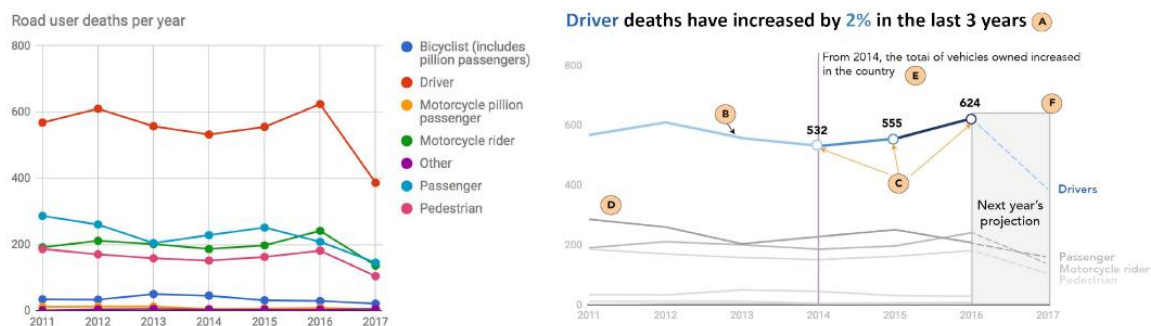


Figura 7. Exemplo com aplicação dos cinco princípios de *data storytelling* (A–E) (extraído de Echeverria et al. (2018))

### 2.5.3. Desafios e Limitações

Apesar do crescente interesse da comunidade científica, persiste uma indefinição teórica significativa quanto ao conceito de *data storytelling*. A ausência de uma definição padronizada tem conduzido a interpretações divergentes na literatura, dificultando a comparação entre estudos e a avaliação do seu impacto em diferentes domínios (Quenum et al., 2025; Ren et al., 2023). Esta lacuna reforça a necessidade de desenvolver quadros conceptuais mais claros, que permitam compreender os seus elementos estruturais, abordagens narrativas e critérios de eficácia.

Para além da indefinição conceitual, destacam-se diversos desafios práticos que comprometem a eficácia da narrativa com dados:

- **Excesso de dados irrelevantes** – Um dos princípios obstáculos é a dificuldade de seleccionar e estruturar os dados mais significativos, o que pode conduzir à sobrecarga informativa e dificultar a identificação da mensagem principal (Boldosova & Luoto, 2020);
- **Necessidade de competências narrativas** – A transformação de dados em histórias eficazes requer competências específicas de comunicação e narração, nem sempre presentes entre os analistas de dados, cuja formação tende a privilegiar aspetos técnicos (Dykes, 2019);
- **Subjetividade na interpretação** – A eficácia da narrativa pode variar consoante o perfil do público-alvo. A interpretação dos dados apresentados pode ser influenciada

por fatores culturais, contextuais ou emocionais, o que introduz um grau de subjetividade que pode comprometer a consistência da mensagem (Dykes, 2019).

Perante estes desafios, torna-se necessário seguir uma abordagem mais crítica e estruturada no uso do *data storytelling*, que considere não apenas a robustez técnica da análise, mas também a clareza da narrativa, a sua adequação ao público-alvo e o contexto interpretativo em que será recebida (Dykes, 2019).

## **2.6. E-commerce: Conceito e Relevância Analítica**

O comércio eletrônico (*e-commerce*) tem-se estabelecido como um dos pilares da economia digital moderna, redefinindo a dinâmica das transações comerciais à escala global. Segundo Jain et al. (2021), o *e-commerce* integra não apenas a compra e venda de produtos, mas também a gestão de dados, logística, marketing e serviços ao cliente, refletindo uma transformação abrangente e contínua no ambiente de negócios (Jain et al., 2021).

### **2.6.1. Conceito e Evolução do E-commerce**

O comércio eletrônico representa uma mudança na forma como consumidores e empresas interagem para realizar transações comerciais, sendo impulsionado por avanços tecnológicos desde meados da década de 1990. De acordo com Ntumba et al. (2023), esta transformação começou a ganhar expressão com o surgimento de empresas pioneiras como a Amazon, a eBay e a Dell, que estabeleceram as bases para uma nova era digital. A Amazon, por exemplo, foi inicialmente concebida como uma plataforma dedicada à venda de livros, mas rapidamente evoluiu para um *marketplace* generalista. A eBay destacou-se pela introdução de leilões virtuais, enquanto o PayPal trouxe soluções inovadoras de pagamento seguro, aumentando a confiança nas transações online. Essa transição tecnológica foi progressivamente moldada por inovações como a mobilidade, a IA, a Realidade Aumentada (AR) e a análise de grandes volumes de dados (Ntumba et al., 2023).

Além da sua evolução histórica, o conceito de *e-commerce* envolve dimensões técnicas e operacionais que asseguram o seu funcionamento enquanto sistema digital integrado. Tsai & Cheng (2012) destacam que o desenvolvimento tecnológico favoreceu a criação de canais digitais capazes de integrar fornecedores e consumidores, ampliando a eficiência e o alcance das transações. Nesse contexto, diversas inovações passaram a ser fundamentais para a evolução do *e-commerce*, nomeadamente:

- **Dispositivos móveis**, que possibilitaram o desenvolvimento do *mobile commerce*, permitindo aos consumidores efetuar compras em qualquer lugar e momento (Tsai & Cheng, 2012);
- **Gateways de pagamento**<sup>2</sup>, essenciais para assegurar transações digitais seguras e rápidas, apoiando a confiança dos utilizadores nas plataformas online (Jain et al., 2021);
- **Análise de dados**, que permite às empresas interpretar o comportamento dos consumidores e adaptar as suas estratégias em tempo real (Jain et al., 2021);
- **Redes sociais**, utilizadas pelas empresas para promover produtos, interagir com clientes e reforçar a fidelização através de feedback e estratégias de divulgação digital (Jain et al., 2021);
- **Impressão 3D**, que permite a produção personalizada de bens diretamente a partir de modelos digitais, transformando os modelos de produção e distribuição (Jain et al., 2021).

Do ponto de vista funcional, o *e-commerce* envolve não apenas a comercialização de produtos e serviços, mas também a transferência de informações e a execução de pagamentos em ambientes digitais. Esse modelo baseia-se numa infraestrutura tecnológica que permite a automação de processos e a integração entre sistemas internos e externos. Como afirmam Jain et al. (2021), uma das principais vantagens do comércio eletrónico para os vendedores é o aumento das vendas aliado à redução dos custos de operação e manutenção, impulsionando a competitividade no meio digital (Jain et al., 2021).

Complementando essa evolução, destaca-se o papel estratégico da vertente analítica. A integração de ferramentas de *Big Data* permite recolher e processar grandes volumes de informação em tempo real, promovendo decisões mais precisas em áreas como preços dinâmicos, segmentação e fidelização (Ghandour, 2015). Esta capacidade analítica é ampliada pelo uso de técnicas de ML e análises preditivas, que, segundo Riwayat et al. (2024) permitem extrair padrões relevantes dos dados e orientar decisões estratégicas. Os autores também alertam para os desafios relacionados à qualificação técnica e à gestão ética

---

<sup>2</sup> Gateway de pagamento: Serviço que facilita a transferência segura de dados de pagamento entre um site de *e-commerce* e um processador de pagamentos

e segura dos dados dos consumidores, salientando a importância de regulamentações e capacitação adequada nas equipas de *e-commerce* (Riwayat et al., 2024).

Dessa forma, o *e-commerce* consolidou-se como uma componente essencial da transformação digital, refletindo a interdependência entre tecnologia, análise de dados e estratégias orientadas ao consumidor.

### 2.6.2. Modelos de E-commerce

Os modelos de *e-commerce* podem ser classificados segundo dois critérios principais: o tipo de bens ou serviços comercializados e a natureza dos participantes envolvidos na transação, como empresas, consumidores ou entidades públicas (Joshi & Dumbre, 2017). Esta classificação permite identificar diferentes configurações de interação digital, que variam conforme os canais utilizados e os fluxos operacionais associados.

De acordo com Jain et al. (2021), os principais modelos de *e-commerce* são:

- ***Business-to-Business (B2B)***: Transações entre empresas, como fornecedores e retalhistas, geralmente suportadas por sistemas integrados de gestão de inventário e logística;
- ***Business-to-Consumer (B2C)***: Vendas diretas de empresas para consumidores finais, sendo o modelo mais visível no comércio eletrónico, especialmente no retalho online;
- ***Consumer-to-Consumer (C2C)***: Transações entre consumidores, frequentemente através de plataformas que facilitam a venda de bens usados ou trocas informais;
- ***Consumer-to-Business (C2B)***: Modelo em que indivíduos oferecem produtos ou serviços a empresas, como em plataformas de *freelancers* ou sites de comercialização de fotografias e conteúdos digitais licenciados.
- ***Business-to-Administration (B2A)*** e ***Consumer-to-Administration (C2A)***: Interações entre empresas ou cidadãos e entidades governamentais, abrangendo serviços como declarações fiscais, candidaturas a benefícios sociais ou agendamentos online com organismos públicos.

### **2.6.3. Benefícios e Desafios do E-commerce**

O *e-commerce* tem-se destacado como uma das maiores inovações no contexto dos negócios digitais, proporcionando vantagens significativas para consumidores e empresas.

Do ponto de vista do consumidor, a principal vantagem consiste na conveniência, expressa na possibilidade de realizar compras a qualquer hora e a partir de qualquer lugar, sem restrições geográficas ou horárias (Jain et al., 2021; Kale, 2022). A interação com opiniões de outros consumidores também permite avaliar a qualidade dos produtos antes da decisão de compra (Jain et al., 2021).

Para os vendedores, o *e-commerce* contribui para o aumento das receitas e para a redução dos custos operacionais e logísticos, levando a uma maior eficiência nos processos e a otimização dos recursos (Jain et al., 2021). Além disso, o uso de plataformas digitais permite uma melhor comunicação com os clientes e fornecedores, contribuindo para o fortalecimento das relações comerciais e da imagem da marca (Anzar, 2016). O acesso a dados analíticos também tem permitido às empresas compreender melhor o comportamento dos consumidores e ajustar as suas estratégias de marketing com maior precisão (Jain et al., 2021).

Contudo, os desafios associados ao comércio eletrônico são numerosos e relevantes. Questões como a segurança cibernética e a proteção da privacidade dos utilizadores são desafios importantes, uma vez que ataques e falhas de segurança podem comprometer a confiança dos consumidores (Jain et al., 2021).

Existem outros desafios, tais como, limitações no atendimento ao cliente, a impossibilidade de examinar fisicamente os produtos antes da compra e o tempo necessário para entrega, fatores que afetam a experiência de consumo (Kale, 2022). Barreiras sociais, como baixos níveis de literacia digital, tradições culturais consolidadas e desconfiança nas transações online, também dificultam o crescimento do *e-commerce*. (Kale, 2022).

## **2.7. Integração de Conceitos no Contexto Analítico do E-commerce**

A evolução dos sistemas de informação e a crescente digitalização das operações comerciais trouxeram novos desafios e oportunidades para o setor do *e-commerce*. Neste contexto, a capacidade de recolher, processar e interpretar grandes volumes de dados em tempo real

tornou-se uma vantagem competitiva essencial. Conceitos como Visualização de Dados, BI, *Data Storytelling*, *Dashboards* e KPIs emergem como pilares fundamentais para uma gestão orientada por dados. A integração destes elementos permite compreender melhor o comportamento dos consumidores, otimizar processos internos e definir estratégias mais eficazes. As secções seguintes exploram cada um destes conceitos e a sua aplicação prática no comércio eletrónico, destacando o seu contributo para decisões operacionais, táticas e estratégias mais informadas.

### **2.7.1. Visualização de Dados e Business Intelligence no E-commerce**

A visualização de dados tornou-se uma componente essencial da infraestrutura analítica das empresas de *e-commerce*, ao permitir representar graficamente grandes volumes de informação de forma clara e acessível. Esta abordagem favorece decisões operacionais, táticas e estratégicas, ao facilitar a identificação de padrões de consumo, comportamentos sazonais e tendências por meio de gráficos, *dashboards* interativos e representações visuais intuitivas (Bhargava et al., 2023).

No contexto do comércio eletrónico, os *dashboards* integram e consolidam métricas como vendas, tráfego, taxas de conversão e inventário, a partir de diversas fontes, incluindo plataformas de vendas, sistemas de CRM e redes sociais (Vamsi et al., 2025). Esta centralização de dados promove a análise em tempo real e melhora a gestão de campanhas, alocação de recursos e experiência do cliente.

Um exemplo prático desta aplicação é apresentado por Bhargava et al. (2023), que desenvolveram um *dashboard* interativo em Power BI com base em dados transacionais. A Figura 8 representa este *dashboard*, no qual se destacam múltiplos cartões com indicadores globais e gráficos como os circulares, de anel e *treemaps*. Estes elementos permitem explorar e comparar visualmente métricas de vendas por diferentes dimensões, como estado, ano, mercado, região e categoria. Os *slicers* incorporados na interface oferecem filtros dinâmicos que adaptam imediatamente os resultados visuais conforme as seleções feitas pelo utilizador. Esta funcionalidade facilita a análise segmentada e comparativa, permitindo, por exemplo, isolar o desempenho de uma determinada região ou categoria de produto e avaliar a sua evolução temporal. O gráfico *treemap*, em particular, destaca visualmente as concentrações de vendas por estado, utilizando a área de cada bloco para representar proporcionalmente o volume de vendas, o que ajuda a identificar rapidamente os mercados mais representativos. Este tipo de interface promove uma experiência analítica orientada à descoberta de padrões e à formulação de *insights* com impacto direto na gestão do negócio (Bhargava et al., 2023).

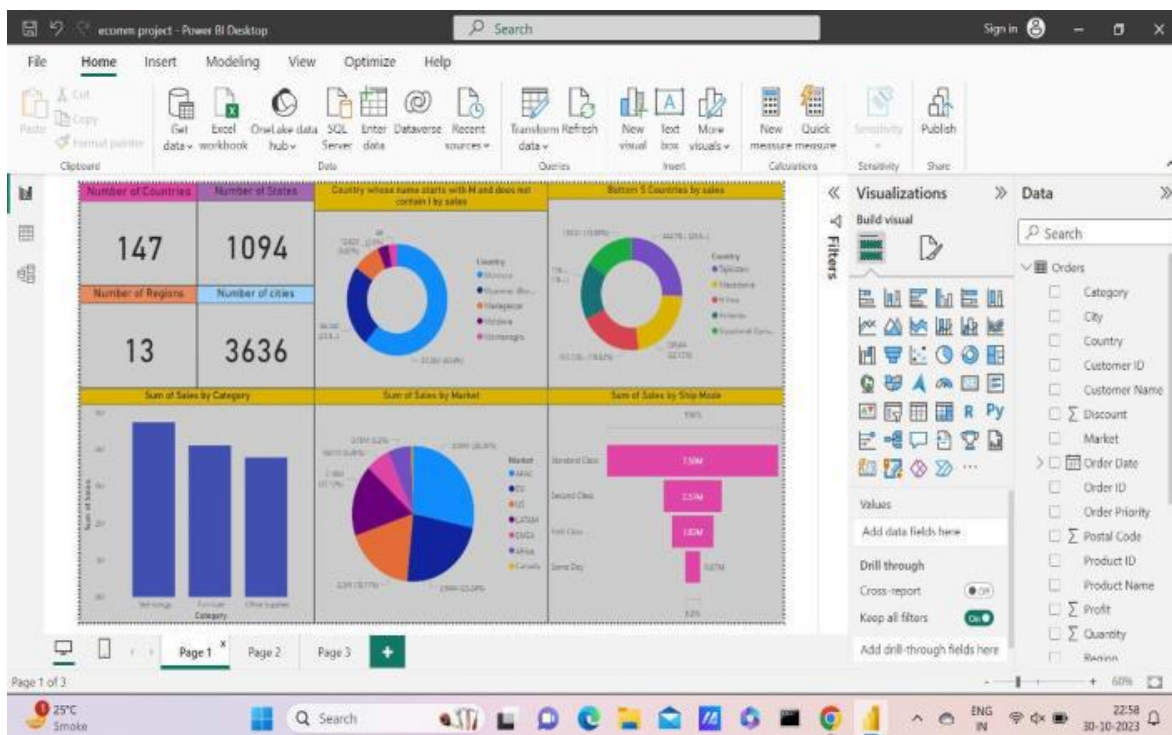


Figura 8. *Dashboard* de *e-commerce* em Power BI com filtros interativos (extraído de Bhargava et al. (2023))

A análise de comportamento do consumidor também beneficia diretamente da visualização de dados. Segundo Anwar et al. (2025), representações gráficas contribuem para compreender o impacto de fatores como sazonalidade, perfil demográfico e preferências nas decisões de compra, permitindo estratégias de segmentação mais precisas e personalizadas (Anwar et al., 2025).

Nesse enquadramento, o uso de BI no *e-commerce* assume um papel transformador. Pan et al. (2021) realçam que a conjugação entre BI e IA tem permitido reduzir custos operacionais, aumentar a personalização e reforçar a fidelização dos clientes (Pan et al., 2021). A visualização atua como ponte entre os modelos analíticos preditivos e a sua interpretação prática pelos decisores (Ferreira et al., 2019).

A integração entre sistemas de BI e plataformas de *e-commerce* também possibilita a extração de valor estratégico de dados operacionais, promovendo uma gestão de stocks mais eficiente, a leitura de padrões de consumo e a construção de relações mais eficazes com os clientes (Ferreira et al., 2019).

Ferramentas como Power BI, Tableau e Google Data Studio destacam-se pela sua flexibilidade na criação de visualizações personalizadas e interativas, adaptadas aos diferentes perfis de utilizador (Bhargava et al., 2023).

### **2.7.2. Integração do Data Storytelling no E-commerce**

A integração do *data storytelling* no *e-commerce* representa uma evolução estratégica na forma como as marcas se relacionam com os consumidores digitais. Ao tornar o conteúdo mais apelativo, o *data storytelling* facilita a comunicação e aproxima o utilizador da informação, contribuindo para uma experiência mais completa. De acordo com Karampournioti & Wiedmann (2021), esta abordagem melhora simultaneamente a perceção funcional e emocional das lojas online, aumentando a atratividade, a valorização da marca e a disposição dos consumidores para pagar mais. Os autores demonstram ainda que o *data storytelling* tem impacto tanto no processamento consciente como subconsciente da informação, influenciando positivamente a atitude para com a marca e as intenções de compra (Karampournioti & Wiedmann, 2021).

Um exemplo prático desta abordagem é o caso da marca *Coffee Circle*, que comercializa café e produtos relacionados. A empresa desenvolveu duas versões da sua loja online: uma página tradicional baseada em texto e outra com design interativo que recorre a técnicas de *scrollytelling*<sup>3</sup>. Ambas disponibilizam as mesmas informações sobre a missão e os valores da marca, como a sustentabilidade, a justiça no comércio e o apoio às comunidades produtoras de café. No entanto, a versão com *data storytelling* proporciona uma experiência

---

<sup>3</sup> *scrollytelling*: técnica narrativa que associa o *scroll* à apresentação progressiva de informação visual

visualmente mais rica, com elementos gráficos, animações e uma narrativa fluída que orienta o utilizador pela história da marca e pelos seus impactos sociais. A Figura 9 ilustra estas duas abordagens: a parte superior da figura representa a versão com *data storytelling* digital interativo, enquanto a parte inferior mostra a versão tradicional baseada em texto. Esta estratégia tem como objetivo criar associações positivas à marca e reforçar a ligação emocional com o consumidor, o que, segundo Karampournioti & Wiedmann (2021) contribui para uma maior intenção de compra e maior disposição para pagar um preço superior.

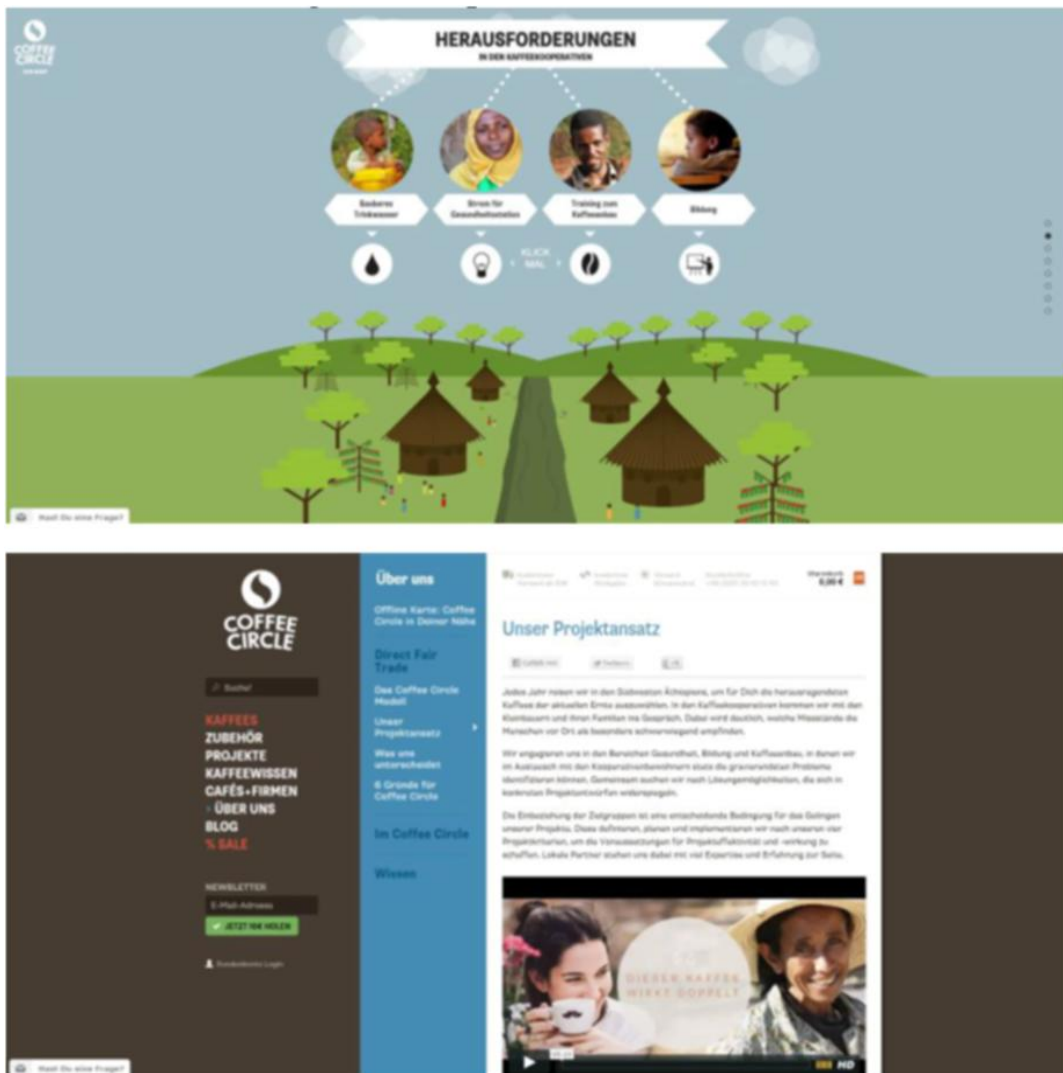


Figura 9. Coffee Circle: *data storytelling* vs. tradicional (extraído de Karampournioti & Wiedmann (2021))

Por outro lado Boldosova (2020), demonstra que, no contexto B2B, a combinação entre *data storytelling* e análise de *Big Data* contribui para uma comunicação mais eficaz do valor associado aos serviços inteligentes. Esta abordagem facilita a compreensão e a adoção dessas soluções digitais por parte dos clientes, enquanto ajuda a ultrapassar barreiras cognitivas. O

uso de narrativas cocriadas, baseadas em experiências reais e casos de sucesso, revela-se fundamental na construção de relações duradouras entre fornecedores e compradores (Boldosova, 2020).

Neste sentido, Herschel & Clements (2017) reforçam que o *data storytelling* no BI é fundamental para contextualizar dados analíticos e facilitar a sua compreensão pelos decisores. A estrutura narrativa permite transformar informação complexa em mensagens acessíveis, suportadas por visualizações que maximizam o impacto comunicacional das análises preditivas (Herschel & Clements, 2017).

Do ponto de vista da experiência do consumidor, Lundqvist et al. (2013) mostram que a exposição a uma história criada pela própria marca conduz a experiências mais positivas, aumenta a percepção de valor da marca e leva os consumidores a estarem dispostos a pagar mais pelos produtos. Estes resultados mostram o impacto do *data storytelling* na construção de percepções positivas da marca e na sua diferenciação num mercado competitivo. (Lundqvist et al., 2013).

M. Sharma (2019) destaca a importância do envolvimento narrativo nos ambientes do *e-commerce*, evidenciando que fatores como a imersão, a identificação com personagens e a coerência narrativa são cruciais para captar a atenção dos consumidores e prolongar o tempo de permanência nos sites. A narrativa funciona como uma ponte emocional entre o consumidor e a marca, criando ligações que ultrapassam a funcionalidade do produto. Estes elementos reforçam a importância do *data storytelling* como recurso estratégico para promover relações duradouras no contexto digital (M. Sharma, 2019).

### **2.7.3. KPIs no E-commerce: Operacionais, Estratégicos, Táticos e Financeiros**

KPIs ou Indicadores-Chave de Desempenho, são indicadores utilizados pelas organizações para estimar e reforçar o seu grau de sucesso em relação a metas previamente estabelecidas e de longo prazo (Velimirovića et al., 2010). Estes indicadores também são concebidos como métricas mensuráveis aplicadas para avaliar o desempenho organizacional no cumprimento de objetivos estratégicos (Ahmed et al., 2017).

No contexto da monitorização organizacional, os KPIs desempenham um papel essencial ao possibilitar o acompanhamento contínuo de tendências e ao apoiar a tomada de decisões baseadas em dados. Como demonstram Bodschwinn & Philipp (2024), a utilização eficaz

de uma aplicação web com painéis de controlo de KPIs facilita a análise direcionada de períodos específicos, melhora a acessibilidade à informação relevante e contribui para a melhoria da experiência do utilizador e da eficiência na gestão de dados institucionais (Bodschwinna & Philipp, 2024).

Além disso, os KPIs possibilitam uma comunicação mais clara e objetiva entre os diversos níveis hierárquicos da organização, permitindo alinhar as atividades operacionais com os objetivos estratégicos previamente definidos. A padronização e automatização do acompanhamento dos KPIs, como proposto por Bodschwinna & Philipp (2024), também contribui para a redução de erros humanos e para a otimização do tempo e dos recursos utilizados na análise de desempenho. Assim, os KPIs não apenas promovem uma cultura organizacional orientada por dados, mas também asseguram uma base sólida para a avaliação contínua, adaptação de estratégias e melhoria da qualidade institucional (Bodschwinna & Philipp, 2024).

No domínio do comércio eletrónico, esta abordagem orientada por dados adquire uma relevância ainda maior, considerando a complexidade e especificidade das atividades digitais. Keong Choong & Bacon (2024) observam que os sistemas contabilísticos tradicionais são inadequados para medir elementos críticos do *e-commerce*, como o fluxo de tráfego, a fiabilidade dos pagamentos eletrónicos e a utilização de dados provenientes das redes sociais. Esta limitação exige a utilização de sistemas de medição ajustados às realidades digitais (Keong Choong & Bacon, 2024). Complementarmente, Ahmed et al. (2017) sublinham que a escolha criteriosa de KPIs, com base em ferramentas analíticas como o *Google Analytics* e técnicas de ML, é essencial para garantir decisões eficazes num ambiente em constante transformação (Ahmed et al., 2017).

Com base nesta lógica de adaptação às exigências digitais, os KPIs utilizados no comércio eletrónico podem ser organizados em três níveis distintos, de acordo com as suas funções no processo de gestão: operacionais, táticos, estratégicos e financeiros. A seguir, descrevem-se alguns dos principais indicadores utilizados na monitorização do desempenho neste contexto.

### Indicadores Operacionais:

- **Visitors:** Quantidade de visitantes que acessam o site num determinado período, permitindo avaliar o alcance e a visibilidade da plataforma (Nepal, 2017).
- **Orders:** Número total de encomendas concluídas, funcionando como um indicador direto da atividade comercial e da capacidade de conversão da loja online (Nepal, 2017).
- **Conversion Rate (CR):** Proporção de visitantes que realizaram uma ação desejada, como uma compra (Ahmed et al., 2017; Keong Choong & Bacon, 2024).
- **Refund and Return Rate:** Representa a percentagem de devolução em relação ao total de vendas realizadas (Keong Choong & Bacon, 2024).
- **Bounce Rate:** Refere-se à percentagem de visitantes que abandonam o site após visualizarem uma página (Ahmed et al., 2017; Keong Choong & Bacon, 2024).
- **Unique Visitors:** Refere-se ao número de visitantes distintos que acessam o site num determinado período (Ahmed et al., 2017).

### Indicadores Estratégicos:

- **Customer Lifetime Value (CLV):** Estima o valor financeiro total que um cliente pode gerar ao longo do tempo (Keong Choong & Bacon, 2024).
- **Return of Advertising Spend (ROAS):** Mede o retorno financeiro direto obtido a partir do investimento em anúncios (Keong Choong & Bacon, 2024).
- **Net Customer Margin (NCM):** Representa a margem líquida gerada por um cliente ao longo do tempo, tendo em conta os custos associados à sua aquisição e manutenção (Keong Choong & Bacon, 2024).
- **Customer Visit Frequency:** Indica o grau de fidelização e retenção de clientes ao longo do tempo, sendo fundamental para decisões estratégicas de relacionamentos e lealdade (Nepal, 2017).

### Indicadores Táticos:

- **Click-through Rate (CTR):** Avalia a eficácia dos anúncios digitais ao medir a percentagem de vezes que os utilizadores clicam num anúncio em relação ao número total de visualizações (Keong Choong & Bacon, 2024).

- **Cost per Click (CPC):** Mede o custo médio por cada clique obtido em campanhas de publicidade digital (Keong Choong & Bacon, 2024).
- **Cost per Acquisition (CPA):** Indica o custo médio necessário para adquirir um cliente durante uma campanha de marketing (Keong Choong & Bacon, 2024).
- **Average Time Spent:** Auxilia na análise do comportamento do utilizador e permite ajustes táticos na estrutura de navegação e conteúdo do site (Ahmed et al., 2017; Nepal, 2017).
- **Shopping Cart Abandonment:** Indica pontos de desistência durante o processo de compra, sendo útil para identificar obstáculos que impedem a conclusão da transação (Ahmed et al., 2017; MUNTEAN et al., 2016; Nepal, 2017).
- **Email Subscribers:** Avalia o desempenho das campanhas por email com base na percentagem de utilizadores que realizaram uma compra após interagirem com os conteúdos recebidos (Keong Choong & Bacon, 2024).
- **Mobile Users:** Mede a proporção de transações concluídas através de dispositivos móveis, refletindo a relevância crescente do *mobile commerce* (Keong Choong & Bacon, 2024).
- **Average Revenue per Visitor (ARPV):** Indica o valor médio da receita gerada por cada visitante do site, permitindo avaliar a eficácia geral do desempenho comercial da plataforma (MUNTEAN et al., 2016).
- **Average Items per Order (AIPO):** Indica quantos produtos, em média, são comprados em cada encomenda, ajudando a entender o volume habitual de compras por pedido (MUNTEAN et al., 2016).
- **Average Page Views per Visit:** Permite avaliar o nível de interesse e envolvimento dos utilizadores com o conteúdo do site (Ahmed et al., 2017).
- **Returning Visitors:** Indica o grau de retenção e o potencial de fidelização dos utilizadores (Ahmed et al., 2017).

#### **Indicadores Financeiros:**

- **Revenue:** KPI financeiro essencial, refletindo o volume de negócios obtido através das vendas (Nepal, 2017).
- **Average Order Value (AOV):** Contribui para decisões financeiras estratégicas, nomeadamente na definição de preços, planeamento de promoções e avaliação da rentabilidade associada a cada pedido (MUNTEAN et al., 2016; Nepal, 2017).

- **Return on Investment (ROI):** Mede o retorno financeiro obtido sobre os investimentos realizados (Keong Choong & Bacon, 2024).
- **Net Profit (NP):** Representa o lucro líquido após a dedução de todos os custos operacionais e despesas (Keong Choong & Bacon, 2024).
- **Contribution Margin (CM):** Representa o lucro obtido antes da dedução dos custos fixos (Keong Choong & Bacon, 2024).

Estes indicadores permitem monitorizar, comparar e gerir o progresso organizacional, fornecendo informações claras e quantificáveis sobre a eficácia das estratégias e processos adotados (Bodschwinna & Philipp, 2024).

Ahmed et al. (2017) sublinham que a má escolha dos KPIs, como medir tudo indiscriminadamente, pode levar a interpretações erradas e decisões pouco eficazes (Ahmed et al., 2017). Já Ishaq Bhatti & Awan (2014) defendem que indicadores como qualidade, custo, flexibilidade, satisfação do cliente, entre outros, têm impacto significativo na performance global das organizações (Ishaq Bhatti & Awan, 2014). No contexto do *e-commerce*, a definição de KPIs torna-se ainda mais relevante, dada a natureza dinâmica e competitiva do ambiente digital (Ahmed et al., 2017).

#### **2.7.4. Classificação dos Dashboards: Operacional, Tático e Estratégico no E-commerce**

A crescente complexidade dos ambientes organizacionais e a velocidade com que os dados são gerados exigem ferramentas que permitam uma análise estruturada e eficiente da informação. Neste contexto, os *dashboards*, ou painéis de controlo, destacam-se como instrumentos essenciais para sintetizar grandes volumes de dados e apoiar decisões em diferentes níveis hierárquicos (Hayward, 2024). A sua utilização proporciona uma visão clara, integrada e atualizada do desempenho organizacional, permitindo identificar rapidamente áreas críticas ou oportunidades de melhoria. No entanto, nem todos os *dashboards* têm o mesmo propósito ou são dirigidos ao mesmo público. De acordo com Hayward (2024), a sua eficácia depende do alinhamento entre os objetivos do utilizador, o tipo de decisão a suportar e a frequência com que os dados são consultados. Com base nestes critérios, é possível classificar os *dashboards* em três categorias principais: estratégicos, táticos e operacionais, cada uma com funções e profundidades analíticas distintas.

Os *dashboards* estratégicos destinam-se à alta administração, como diretores e executivos, e têm como objetivo fornecer uma visão agregada e de longo prazo do desempenho da organização. São focados no acompanhamento de objetivos globais, metas corporativas e KPIs estratégicos. Estes *dashboards* são frequentemente atualizados numa base mensal ou trimestral, permitindo avaliar o progresso face ao planeamento estratégico e apoiar decisões estruturais com impacto institucional alargado.

Já os *dashboards* táticos operam a um nível intermédio da gestão, sendo utilizados por chefias de departamento e gestores de equipa. Estes *dashboards* traduzem os objetivos estratégicos em metas operacionais concretas, oferecendo uma visão detalhada do desempenho de processos, projetos ou áreas específicas. A sua frequência de atualização tende a ser semanal ou mensal, servindo como instrumento de monitorização do progresso das atividades e suporte à coordenação entre equipas.

Por fim, os *dashboards* operacionais têm como finalidade monitorizar as atividades do dia a dia e identificar de forma rápida quaisquer desvios ou falhas nos processos. São utilizados por equipas técnicas e operacionais e caracterizam-se pela atualização constante (em tempo real ou diária), permitindo uma resposta rápida a eventos críticos. Estes *dashboards* focam-se em indicadores como produtividade, taxas de erro, tempo de resposta ou volume de trabalho, sendo essenciais para assegurar a eficiência no funcionamento da organização.

De acordo com o guia de boas práticas publicado por Hayward (2024), a distinção entre os diferentes tipos de *dashboards* está ligada às perguntas que se pretende responder, ao tipo de utilizadores e à frequência com que a informação precisa de ser consultada. Esta abordagem facilita a personalização dos *dashboards* conforme os níveis hierárquicos da organização e reforça a eficácia da gestão baseada em dados.

Para ilustrar esta lógica, a Tabela 2 apresenta uma adaptação de uma tabela comparativa disponibilizada por Hayward (2024), que ajuda a selecionar o tipo de *dashboard* mais adequado em função de questões-chave como “que problema estamos a tentar resolver?”, “quem irá utilizar o *dashboard*?” ou “que lacunas existem no nosso desempenho?”. Esta estrutura fornece orientações práticas para alinhar os *dashboards* aos objetivos específicos da organização e ao perfil dos seus utilizadores, promovendo maior eficácia na análise de dados e na tomada de decisão.

Tabela 2. Tipos de *Dashboard*. (adaptado de Hayward (2024))

| <b>Questão orientadora</b>                      | <b>Dashboard Operacional</b>  | <b>Dashboard Tático</b>  | <b>Dashboard Estratégico</b>   |
|---|---|--|--|
| <b>Que problema estamos a tentar resolver?</b>  | Necessidade de acesso rápido a dados sensíveis ao tempo e visibilidade imediata | Necessidade de identificar tendências e obter análises mais aprofundadas | Monitorização de KPIs de topo e alinhamento com objetivos institucionais |
| <b>Quem irá utilizar o dashboard?</b>           | Gestores operacionais e equipas de execução                                     | Analistas e gestores intermédios   | Diretores e executivos   |
| <b>Que lacunas existem no nosso desempenho?</b> | Indicadores de desempenho diário e falhas operacionais                          | Questões de desempenho recorrentes e acompanhamento semanal              | Visão geral do desempenho mensal ou trimestral                           |
| <b>Quais são os nossos objetivos?</b>           | Garantir estabilidade nos processos e rastrear metas operacionais               | Definir metas analíticas e aumentar a visibilidade sobre processos-chave | Acompanhar objetivos estratégicos e suportar decisões a longo prazo      |

### 3. Metodologia do Caso de Estudo

O presente capítulo descreve a metodologia adotada no desenvolvimento do projeto, apresentando os dados utilizados, as ferramentas aplicadas e as principais decisões metodológicas que orientaram o trabalho.

O estudo segue uma abordagem aplicada, recorrendo a métodos qualitativos e quantitativos. A componente qualitativa esteve associada ao estudo prévio, baseado na análise de literatura científica e na observação de práticas existentes nas áreas de BI e visualização de dados. A componente quantitativa centrou-se na análise e exploração dos dados utilizados no projeto, com vista à extração de *insights* relevantes.

Os conceitos analisados no estudo prévio, nomeadamente os princípios de visualização de dados e os elementos visuais essenciais discutidos na Secção 2.4.3, influenciaram as decisões metodológicas adotadas. Em particular, aspetos como a utilização da cor, a escolha adequada dos tipos de gráfico, a organização espacial das visualizações e a preocupação com a redução da carga cognitiva foram considerados ao longo do processo de conceção e implementação dos *dashboards* (Jiang, 2022; Tufte, 1983; Vanderplas et al., 2020).

Com base neste enquadramento, a metodologia do projeto foi estruturada em etapas sequenciais, que incluíram a preparação e transformação dos dados, a definição do modelo de dados, o desenho conceptual das visualizações e a implementação final dos *dashboards* na ferramenta de BI.

#### 3.1. Enquadramento Geral do Caso de Estudo

Num contexto organizacional cada vez mais orientado por dados, é essencial transformar grandes volumes de informação em conhecimento útil para suportar decisões fundamentadas. Neste projeto, pretende-se aplicar os princípios do BI ao domínio do *e-commerce*, com destaque para a visualização de dados e o *data storytelling* como ferramentas estratégicas na comunicação de *insights* relevantes.

O caso de estudo baseia-se no desenvolvimento de um sistema analítico capaz de agregar, transformar e interpretar dados comerciais provenientes de uma plataforma de vendas online.

Para esse fim, foi selecionado o *dataset* público “[Brazilian E-Commerce](#)” da Olist, disponibilizado na plataforma Kaggle (Sionek, 2018).

Este conjunto de dados contém cerca de 100 mil registros de encomendas efetuadas entre 2016 e 2018 em *marketplaces* brasileiros, abrangendo múltiplas dimensões: estados das encomendas, preços, métodos de pagamento, tempos de entrega, localização dos clientes, características dos produtos e avaliações feitas pelos consumidores.

Os dados encontram-se anonimizados, assegurando a confidencialidade das entidades envolvidas.

A diversidade de variáveis presentes neste *dataset* permite realizar análises detalhadas e contextualizadas, com o objetivo de identificar padrões de comportamento dos consumidores, avaliar a eficiência logística e compreender a experiência de compra a partir de comentários e avaliações. O projeto propõe não só a criação de *dashboards* interativos que apoiem a tomada de decisão, mas também a integração de técnicas narrativas que tornem a comunicação dos resultados mais clara e adaptada a diferentes perfis de *stakeholders*.

Desta forma, o estudo procura demonstrar o potencial do BI, da visualização e do *data storytelling* como pilares para uma gestão mais inteligente e orientada por dados, reforçando a importância da análise visual e narrativa na extração de valor dos dados no comércio eletrônico.

### 3.2. Caracterização dos Dados Fonte

Esta Secção apresenta os ficheiros que compõem o *dataset* original, fornecido em formato CSV. Cada ficheiro representa uma entidade distinta do domínio do *e-commerce* e descreve diferentes etapas do processo de compra, desde o pedido inicial até à entrega e avaliação do produto.

De seguida, apresentam-se os principais ficheiros disponíveis no repositório, com a descrição do seu conteúdo e finalidade analítica:

- **olist\_orders\_dataset.csv:** Contém os dados principais de cada encomenda, incluindo datas de compra, envio, entrega e estado do pedido.
- **olist\_order\_items\_dataset.csv:** Lista os produtos associados a cada encomenda, com referência ao vendedor, valor dos itens e custo de envio.

- **olist\_products\_dataset.csv:** Reúne informações sobre os produtos comercializados, como categoria, dimensões e peso.
- **olist\_customers\_dataset.csv:** Agrega dados de identificação dos clientes, incluindo código de cliente, cidade e estado.
- **olist\_order\_payments\_dataset.csv:** Detalha os pagamentos efetuados em cada encomenda, com informação sobre o método de pagamento, número de prestações e valor pago.
- **olist\_order\_reviews\_dataset.csv:** Inclui as avaliações submetidas pelos clientes, com classificação numérica, título, comentário textual e data da submissão.
- **olist\_sellers\_dataset.csv:** Regista os dados dos vendedores, com códigos identificadores e localização geográfica.
- **product\_category\_name\_translation.csv:** Tabela auxiliar que apresenta a correspondência entre as categorias de produto em português e a respetiva tradução em inglês.
- **geolocation\_dataset.csv:** Contém coordenadas geográficas associadas a códigos postais, permitindo realizar análises espaciais e agrupamentos regionais.

Estes ficheiros constituem a base de dados original sobre a qual assenta a modelação descrita na secção seguinte.

### 3.3. Ferramentas e Tecnologias Utilizadas

A implementação do projeto recorreu a um conjunto de ferramentas tecnológicas que asseguraram a execução das diferentes etapas do processo analítico, desde a preparação dos dados até à construção de *dashboards*. A escolha destas ferramentas teve em consideração não só a sua adequação funcional aos objetivos do projeto, como também a facilidade de utilização, a sua presença consolidada no mercado e o suporte a um desenvolvimento ágil e integrado.

A fase de *Exploratory Data Analysis* (EDA) e o processo de ETL foram desenvolvidos no ambiente *Jupyter Notebook*, recorrendo à linguagem *Python*. Esta combinação foi escolhida por ser prática e por disponibilizar bibliotecas para manipulação, processamento e

visualização de dados, como *pandas*, *numpy*, *matplotlib* e *seaborn*, que facilitam a análise e transformação da informação.

A escolha desta linguagem e das respectivas bibliotecas foi também baseada pela experiência prévia na sua utilização, uma vez que são utilizadas regularmente em contexto profissional, o que contribuiu para uma maior familiaridade com o seu funcionamento e com a adoção de boas práticas de desenvolvimento e análise de dados.

Após a análise comparativa das plataformas realizada no estudo prévio, a escolha incidiu no Microsoft Power BI, que foi utilizado como a principal ferramenta de visualização e análise de dados do projeto, permitindo a construção de *dashboards* interativos e a análise dos resultados obtidos. A sua escolha deveu-se, em grande parte, à disponibilidade de uma versão gratuita com funcionalidades suficientes para responder aos objetivos do projeto, sendo especialmente adequada a contextos académicos.

## 4. Desenho do Sistema

Este capítulo descreve o processo de concepção e desenvolvimento do sistema de BI aplicado ao *e-commerce* da Olist. São apresentadas as etapas que orientaram o desenho da solução, desde a análise exploratória dos dados, que permitiu compreender a sua estrutura e qualidade, até à definição dos requisitos, das perguntas de negócio e das *user stories*, resultando na definição dos KPIs, na organização da narrativa analítica, na modelação dimensional e na arquitetura do sistema de BI.

O objetivo é documentar de forma estruturada o percurso seguido desde a compreensão inicial dos dados até à construção do modelo analítico que serve de base aos *dashboards* e às análises estratégicas desenvolvidas no projeto.

### 4.1. Análise Exploratória de Dados

Antes de definir os requisitos do sistema e estruturar o modelo dimensional, foi realizada uma análise exploratória dos dados, com o objetivo de conhecer melhor os diferentes ficheiros disponíveis no *dataset* da Olist, identificar padrões relevantes, avaliar a qualidade dos dados e apoiar decisões futuras no desenho do sistema (Sionek, 2018).

A análise foi realizada sobre os principais ficheiros disponibilizados pela Olist, incluindo dados de encomendas, clientes, pagamentos, produtos, vendedores, localizações geográficas e avaliações de clientes. Foram utilizadas bibliotecas como *pandas*, *matplotlib*, *seaborn* e *plotly*, permitindo gerar estatísticas descritivas, gráficos de distribuição etc.

Esta análise revelou aspetos relevantes sobre a estrutura e o conteúdo dos dados, como a presença de valores nulos, distribuições assimétricas, padrões geográficos e relações entre variáveis, que serviram de base para decisões sobre o tratamento dos dados, a definição de indicadores e a modelação da informação. Nos subcapítulos seguintes são apresentados exemplos dessa análise, com destaque para a estrutura dos *datasets*, padrões identificados e implicações para o sistema a desenvolver.

#### 4.1.1. Identificação e Descrição das Variáveis dos Conjunto de Dados

Como ponto de partida para a análise exploratória, foi realizado um estudo das colunas de cada tabela, dos respetivos tipos de dados e da descrição associada a cada variável, de forma a compreender a estrutura e o conteúdo dos ficheiros disponibilizados no *dataset* da Olist.

Nesta Secção apresenta-se a estrutura de cada conjunto de dados incluído na análise, com a identificação dos nomes das variáveis (colunas), os respetivos tipos de dados e uma breve descrição do significado de cada coluna. Esta informação é essencial para compreender a natureza dos dados disponíveis, apoiar a construção do modelo analítico e orientar os processos de limpeza, transformação e integração dos dados.

A descrição detalhada de cada *dataset*, incluindo o nome das variáveis, o tipo de dado e uma breve explicação do seu significado, encontra-se disponível no Anexo A .

#### 4.1.2. Resumo dos Datasets

Nesta Secção foi realizado um levantamento das principais características estruturais de cada um dos ficheiros que compõem o *dataset* da Olist, incluindo o número total de registos (linhas), o número de variáveis (colunas), a quantidade de valores nulos e a verificação de possíveis duplicados. Esta análise permitiu avaliar a qualidade e a consistência dos dados antes da realização das etapas de limpeza e transformação.

A Tabela 3 apresenta um resumo com esta informação:

Tabela 3. Resumo estrutural dos ficheiros do *dataset* Olist

| Dataset              | Nº linhas | Nº colunas | Valores Nulos | Duplicados |
|----------------------|-----------|------------|---------------|------------|
| Customers            | 99441     | 5          | 0             | Não        |
| Geolocation          | 1000163   | 5          | 0             | Não        |
| Orders               | 99441     | 8          | 4908          | Não        |
| Order Items          | 112650    | 7          | 0             | Não        |
| Order Payments       | 103886    | 5          | 0             | Não        |
| Order Reviews        | 99224     | 7          | 145930        | Sim        |
| Products             | 32951     | 9          | 2448          | Não        |
| Category_translation | 71        | 2          | 0             | Não        |

|                |      |   |   |     |
|----------------|------|---|---|-----|
| <b>Sellers</b> | 3095 | 4 | 0 | Não |
|----------------|------|---|---|-----|

Esta análise permitiu identificar quais os ficheiros com maior volume de dados, bem como a presença de valores em falta e de registos duplicados, que necessitaram de tratamento durante a preparação dos dados.

No caso particular do *dataset Customers*, verificou-se que um mesmo identificador de cliente único (*customer\_unique\_id*) pode estar associado a diferentes *customer\_id*, refletindo múltiplos registos por cliente no sistema.

Já no *dataset Order\_reviews* foi verificada a presença de registos duplicados, cuja origem e tratamento serão abordados mais à frente, na fase de limpeza e transformação.

#### 4.1.3. Análise de Valores Nulos

Após a análise estrutural, procedeu-se à análise detalhada da distribuição de valores nulos por coluna em cada *dataset*. Esta análise é fundamental para definir estratégias adequadas de tratamento e correção de dados em falta.

A Tabela 4 apresenta as colunas com valores nulos.

Tabela 4. Distribuição de valores nulos por coluna

| <b>Dataset</b>       | <b>Coluna</b>                 | <b>Valores nulos</b> |
|----------------------|-------------------------------|----------------------|
| <b>Orders</b>        | order_approved_at             | 160                  |
|                      | order_delivered_carrier_date  | 1783                 |
|                      | order_delivered_customer_date | 2965                 |
| <b>Order Reviews</b> | review_comment_title          | 87656                |
|                      | review_comment_message        | 58274                |
| <b>Products</b>      | product_category_name         | 610                  |
|                      | product_name_length           | 610                  |
|                      | product_description_length    | 610                  |
|                      | product_photos_qty            | 610                  |
|                      | product_weight_g              | 2                    |
|                      | product_length_cm             | 2                    |
|                      | product height cm             | 2                    |
| product_width_cm     | 2                             |                      |

Verifica-se que o *dataset Orders* apresenta valores em falta em colunas relacionadas com o processo de entrega, nomeadamente nas datas de aprovação e de expedição da encomenda, o que indica a existência de pedidos que não chegaram a ser concluídos.

No *dataset Order\_reviews*, as colunas de texto associadas aos comentários dos clientes contêm uma quantidade elevada de valores nulos, uma vez que muitos consumidores optaram por atribuir apenas uma pontuação sem deixar comentário.

Por fim, o *dataset Products* apresenta lacunas em variáveis relacionadas com características físicas dos produtos, como peso e dimensões.

Essas inconsistências foram consideradas na definição das regras de tratamento de dados aplicadas na fase de pré-processamento e transformação.

#### 4.1.4. Análise de Integridade Referencial

A validação da integridade referencial foi realizada para garantir a consistência dos relacionamentos entre os *datasets* através das *foreign keys*. Esta análise verifica se todas as referências apontam para registos válidos nas tabelas de destino.

A Tabela 5 apresenta os resultados desta validação:

Tabela 5. Validação de integridade referencial

| Tabela Origem         | Foreign Key | Tabela Destino | Total Registos | Válidos | Órfãos | % Integridade |
|-----------------------|-------------|----------------|----------------|---------|--------|---------------|
| <b>Orders</b>         | customer_id | Customers      | 99441          | 99441   | 0      | 100%          |
| <b>Order Items</b>    | order_id    | Orders         | 112650         | 112650  | 0      | 100%          |
| <b>Order Items</b>    | product_id  | Products       | 112650         | 112650  | 0      | 100%          |
| <b>Order Items</b>    | seller_id   | Sellers        | 112650         | 112650  | 0      | 100%          |
| <b>Order Payments</b> | order_id    | Orders         | 103886         | 103886  | 0      | 100%          |
| <b>Order Reviews</b>  | order_id    | Orders         | 99224          | 99224   | 0      | 100%          |

A análise revelou que todos os seis relacionamentos apresentam 100% de integridade referencial. Este resultado indica uma elevada qualidade dos dados e facilita os processos subsequentes de integração e análise, uma vez que não será necessário tratar inconsistências nas relações entre tabelas.

#### 4.1.5. Análise Descritiva dos Campos Numéricos

Após a validação estrutural e referencial dos *datasets*, procedeu-se à análise descritiva das variáveis numéricas, com o objetivo de compreender a sua distribuição, amplitude de valores e coerência com o contexto do negócio. Esta análise permite detetar valores atípicos, potenciais erros de registo e variáveis com baixa variabilidade, que poderão influenciar os resultados das análises exploratórias e dos modelos analíticos posteriores.

Os resultados da análise estatística descritiva encontram-se resumidos na Tabela 6 que apresenta, para cada variável numérica relevante, os respetivos valores mínimo, máximo, média e desvio padrão.

Tabela 6. Estatísticas descritivas das variáveis numéricas relevantes

| Dataset        | Coluna                     | Mínimo | Máximo    | Média    | Desvio Padrão |
|----------------|----------------------------|--------|-----------|----------|---------------|
| Order Items    | price                      | 0,85   | 6.375,00  | 120,65   | 183,63        |
|                | freight_value              | 0,00   | 409,68    | 19,99    | 15,81         |
| Order Payments | payment_sequential         | 1,00   | 29,00     | 1,09     | 0,71          |
|                | payment_installments       | 0,00   | 24,00     | 2,85     | 2,69          |
|                | payment_value              | 0,00   | 13.664,08 | 154,10   | 217,49        |
| Order Reviews  | review_score               | 1,00   | 5,00      | 4,09     | 1,35          |
| Products       | product_name_length        | 5,00   | 76,00     | 48,48    | 10,25         |
|                | product_description_length | 4,00   | 3.992,00  | 771,50   | 635,12        |
|                | product_photos_qty         | 1,00   | 20,00     | 2,19     | 1,74          |
|                | product_weight_g           | 0,00   | 40.425,00 | 2.276,47 | 4.282,04      |
|                | product_length_cm          | 7,00   | 105,00    | 30,82    | 16,91         |
|                | product_height_cm          | 2,00   | 105,00    | 16,94    | 13,64         |
|                | product_width_cm           | 6,00   | 118,00    | 23,20    | 12,08         |

Após a observação dos resultados apresentados na Tabela 6, verifica-se que as variáveis numéricas apresentam elevada heterogeneidade entre os diferentes domínios do *dataset*, refletindo a diversidade de produtos, valores de transação e padrões de comportamento dos consumidores.

No caso das variáveis *price* e *freight\_value*, observa-se uma distribuição assimétrica à direita, com a maioria dos registos concentrados em valores reduzidos e a presença de *outliers* que elevam a média. Já nas variáveis associadas a pagamentos, nota-se uma predominância de parcelamentos curtos (1 a 3 prestações) e valores médios de pagamento moderados, o que sugere um perfil de consumo típico do comércio eletrónico.

A variável *review\_score* apresenta média elevada (4,09) e baixa dispersão, indicando uma tendência geral de avaliações positivas por parte dos clientes. Por sua vez, as variáveis do *dataset products* apresentam variação significativa em termos de peso, dimensões e comprimento das descrições, refletindo a diversidade do portfólio de produtos disponibilizados na plataforma.

Estas observações encontram-se comprovadas graficamente no Anexo B onde são apresentadas as distribuições das principais variáveis numéricas analisadas.

#### 4.1.6. Análise de Correlações Entre Variáveis Numéricas

Após a análise descritiva, procedeu-se ao estudo das correlações entre variáveis numéricas com o objetivo de identificar relações estatísticas relevantes entre atributos que possam refletir padrões de negócio, comportamentos do consumidor ou características físicas dos produtos.

A Figura 10 representa a relação entre o preço do produto e o valor dos portes no *dataset Order\_items*.

Verificou-se uma correlação positiva moderada ( $r = 0,41$ ), indicando que produtos de maior preço tendem a estar associados a custos de envio superiores.

A concentração de pontos na zona inferior esquerda do gráfico demonstra que a maioria dos produtos apresenta preços reduzidos e custos de envio baixos, mas observa-se uma tendência ascendente coerente com o peso e o volume de produtos mais caros, que implicam maior custo logístico.

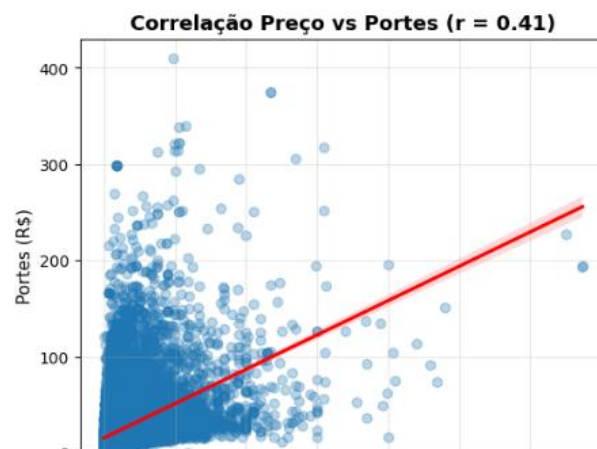


Figura 10. Correlação entre o preço do produto e o valor dos portes de envio

A Figura 11 apresenta a matriz de correlação entre as variáveis de peso, comprimentos, altura e largura dos produtos.

As correlações observadas são positivas e moderadas a fortes ( $0,48 \leq r \leq 0,56$ ), evidenciando a consistência estrutural do *dataset* e confirmando que produtos com maiores dimensões tendem a ter maior peso.

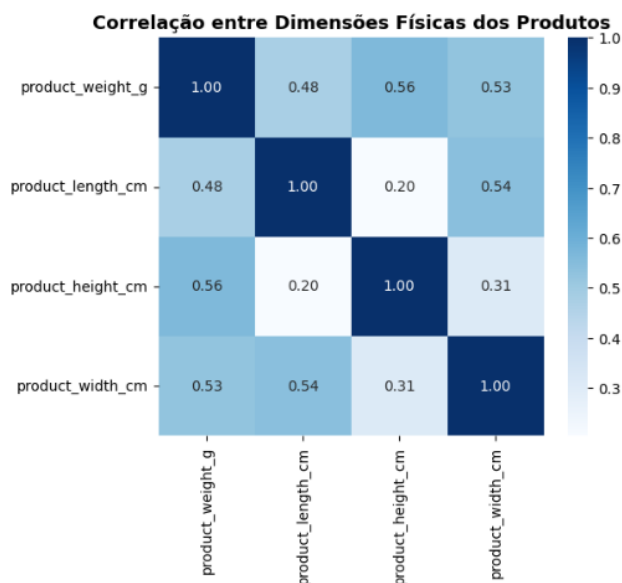


Figura 11. Correlação entre Dimensões Físicas dos Produtos

A Figura 12 mostra a relação entre o valor total pago e o número de prestações registrado no *dataset Order\_payments*.

Foi observada uma correlação positiva moderada ( $r = 0,33$ ), evidenciando que compras de maior valor tendem a ser pagas mais em mais prestações.

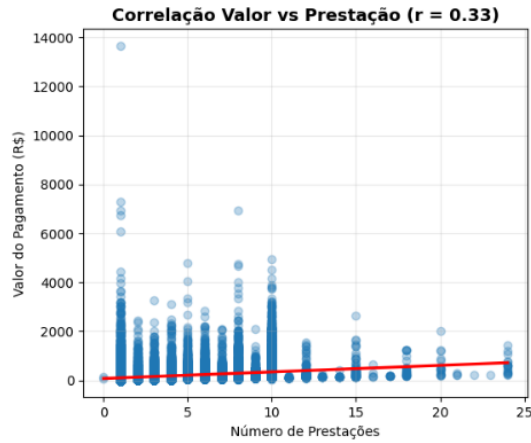


Figura 12. Correlação entre o valor do pagamento e o número de prestações

A Figura 13 apresenta a correlação entre a pontuação atribuída nas avaliações dos clientes (*review\_score*) e o tempo total de entrega (*delivery\_time\_days*), calculado pela diferença entre a data de envio e a data de entrega do pedido.

Foi identificado uma correlação negativa moderada ( $r = -0,33$ ), o que indica que entrega mais rápidas estão associadas a melhores classificações por parte dos clientes.

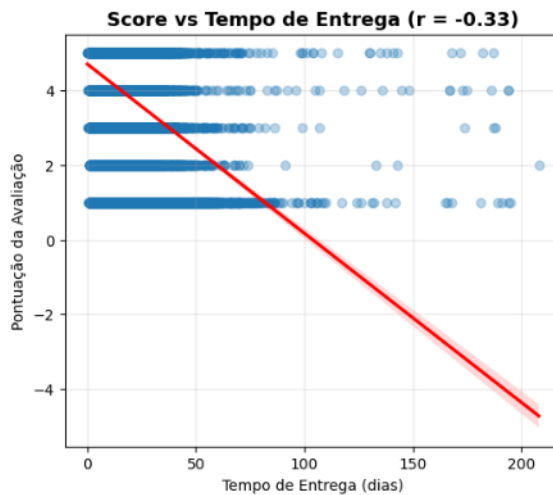


Figura 13. Correlação entre a pontuação das avaliações e o tempo de entrega

#### 4.1.7. Análise Geral de Pedidos e Pagamentos

Nesta Secção procede-se à caracterização geral dos pedidos e dos métodos de pagamento registados na base de dados, com o objetivo de compreender o comportamento operacional da plataforma e as preferências de pagamento dos consumidores.

A Figura 14 apresenta a distribuição dos estados dos pedidos. Observa-se uma predominância do estado *delivered*, que representa a grande maioria das encomendas registadas, demonstrando um nível elevado de encomendas entregues. Os restantes estados, como *shipped*, *cancelado*, *processing* ou *invoiced*, têm pouca representatividade, indicando que a maioria das transações é concluída com sucesso, o que demonstra a eficácia do processo logístico.

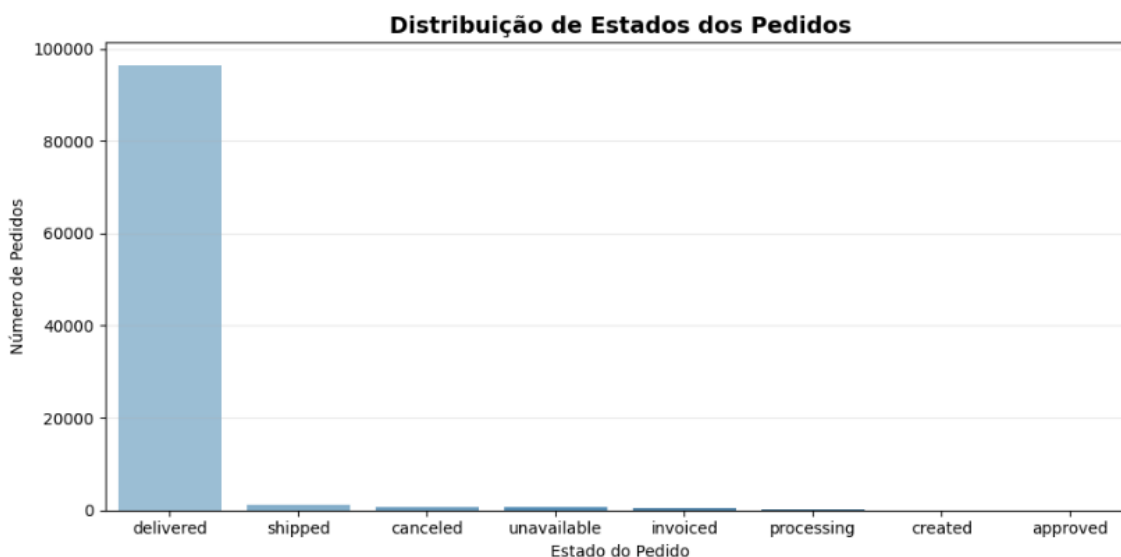


Figura 14. Distribuição dos estados dos pedidos

A Figura 15 ilustra a distribuição dos métodos de pagamento utilizados. Verifica-se uma preferência pelo cartão de crédito, responsável por cerca de 74% das transações. Seguem-se o boleto bancário (19%) e o voucher (5,6%), enquanto os métodos *debit\_card* e *not\_defined* possuem relevância mínima.

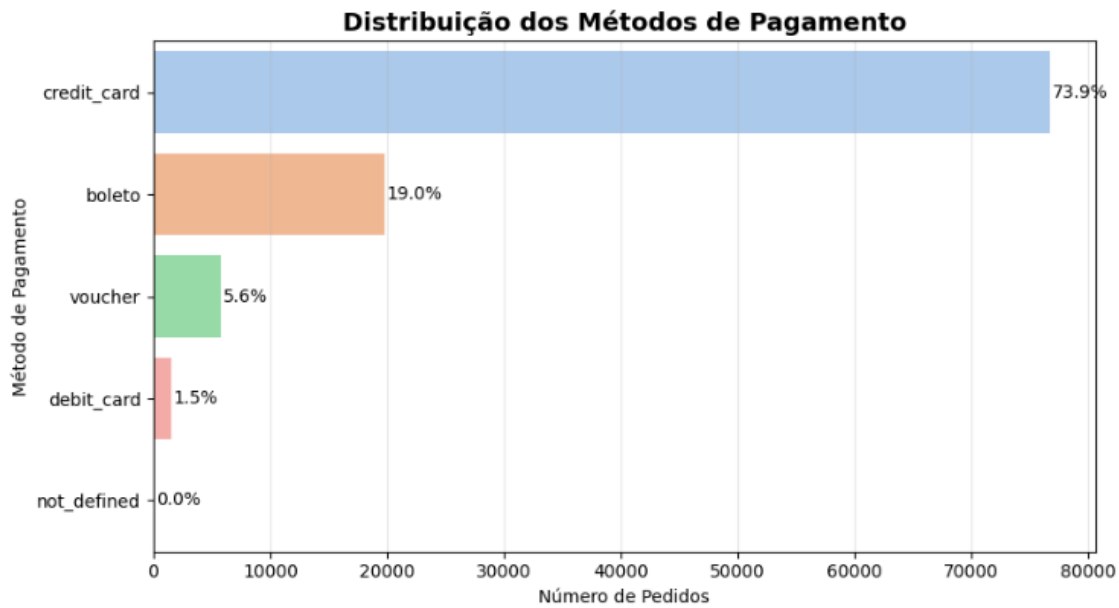


Figura 15. Distribuição dos métodos de pagamento

#### 4.1.8. Análise Temporal das Vendas e Entregas

A análise temporal tem como objetivo avaliar a evolução da atividade de vendas e o desempenho operacional das entregas ao longo do período abrangido pelo conjunto de dados.

A Figura 16 representa a evolução mensal do número de encomendas entre setembro de 2016 e agosto de 2018.

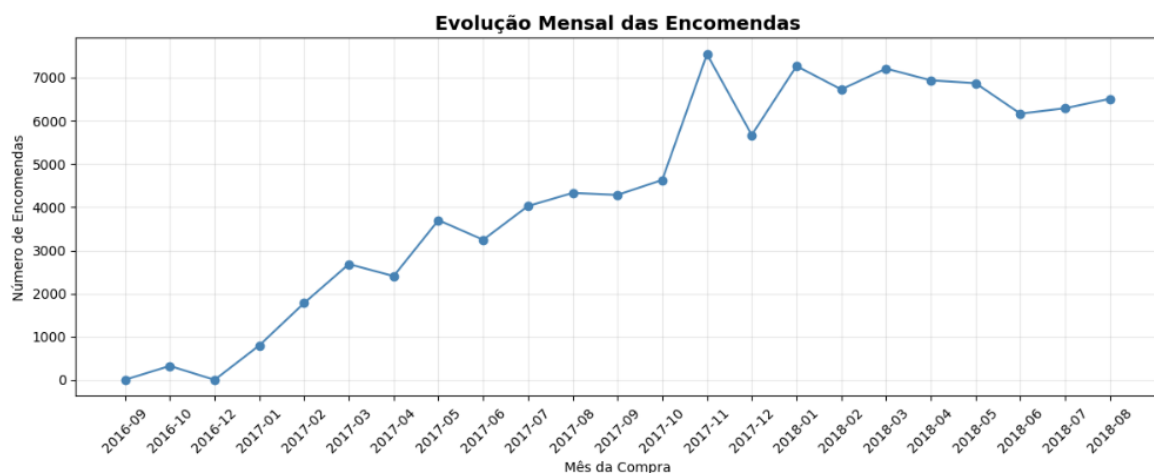


Figura 16. Evolução mensal do número de encomendas entre 2016 e 2018

Observa-se uma trajetória ascendente no volume de vendas, especialmente entre meados de 2017 e o início de 2018, altura em que o número mensal de encomendas ultrapassa as 7.000 unidades. Esta tendência reflete a expansão e consolidação do *marketplace*, possivelmente

influenciada por campanhas promocionais, períodos sazonais e um aumento da adesão de clientes.

A Figura 17 apresenta a evolução mensal do tempo médio de entrega (em dias).

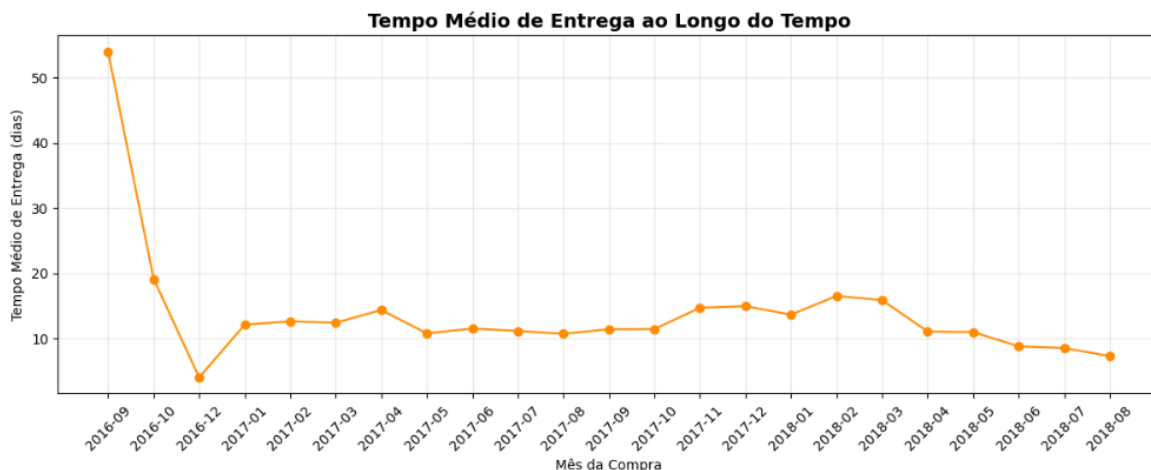


Figura 17. Tempo médio de entrega mensal (em dias)

Embora os primeiros meses de 2016 revelem maior variabilidade, provavelmente devido a um número reduzido de registos, a tendência estabiliza a partir de 2017, mantendo-se entre 10 e 15 dias. Esta consistência demonstra um desempenho logístico estável, capaz de suportar o aumento da procura sem comprometer os prazos da entrega.

#### 4.1.9. Análise das Categorias de Produto e da Satisfação dos Clientes

Após a análise temporal das vendas e entregas, procedeu-se à avaliação das categorias de produto com o objetivo de compreender o portefólio comercial da plataforma, os padrões de consumo e os níveis de satisfação dos clientes.

Esta análise permite identificar quais as categorias com maior procura, bem como aquelas que apresentam preços médios elevados ou melhor avaliação por parte dos consumidores.

A Figura 18 apresenta as dez categorias de produto com maior número de encomendas, acompanhadas do respetivo preço médio e da pontuação média atribuída nas avaliações dos clientes.

Observa-se que as categorias *bed\_bath\_table*, *health\_beauty* e *sports\_leisure* destacam-se em volume de vendas, refletindo a procura por produtos de uso quotidiano e bem-estar pessoal.

Em termos de preço médio, verifica-se uma maior variabilidade entre as categorias analisadas. Categorias como *watches\_gifts* e *auto* apresentam valores médios mais elevados, sugerindo produtos de maior valor agregado, enquanto *bed\_bath\_table* e *health\_beauty* se situam em faixas de preço mais acessíveis.

Relativamente à satisfação do cliente, as pontuações médias situam-se entre 3,6 e 4,1, o que indica uma perceção globalmente positiva, ainda que com margem para melhoria na experiência pós-compra.

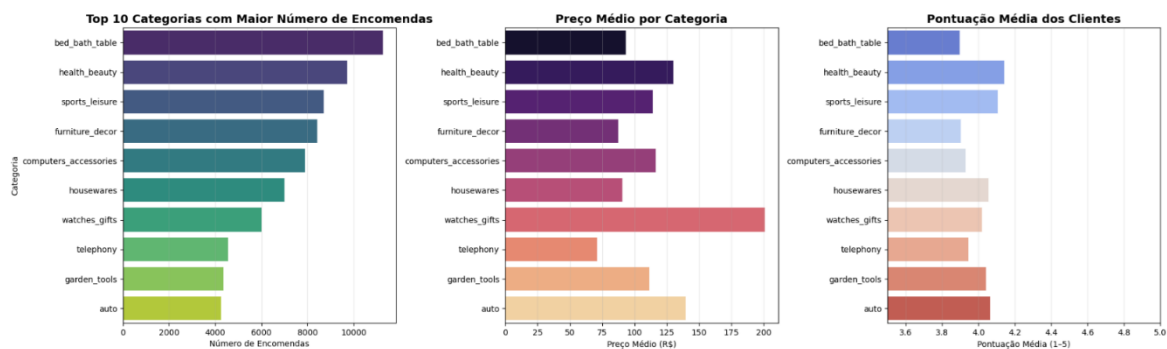


Figura 18. Top 10 Categorias de Produto por Vendas e Satisfação

#### 4.1.10. Análise de Clientes e Localização

Após a caracterização dos produtos e pagamentos, procedeu-se à análise geográfica e comportamental dos clientes, com o objetivo de compreender a distribuição da base de consumidores e o seu impacto nas encomendas realizadas.

A Figura 19 apresenta os 10 estados com maior número de clientes, evidenciando uma forte concentração em São Paulo (SP), que representa a maioria dos utilizadores registados na plataforma. Seguem-se Rio de Janeiro (RJ) e Minas Gerais (MG), reforçando a predominância da região Sudeste no comércio eletrónico nacional.

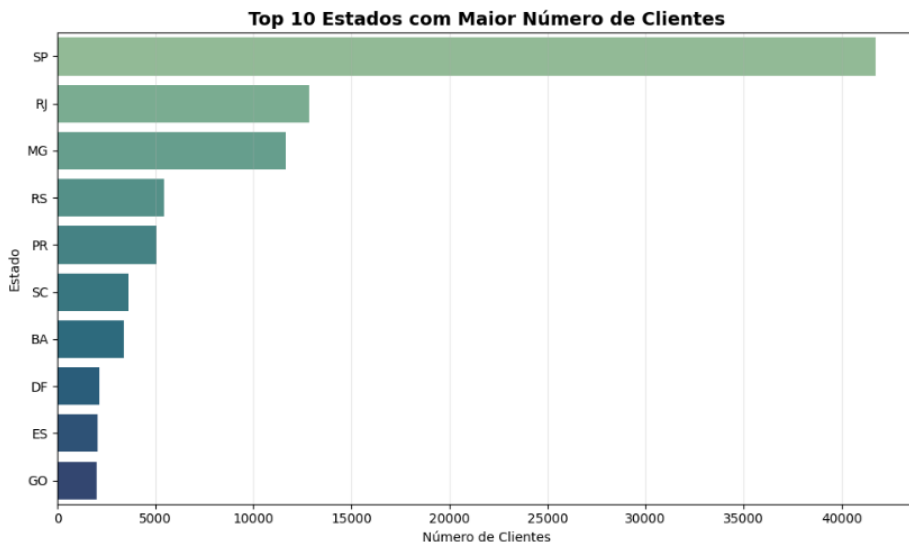


Figura 19. Distribuição dos 10 estados com maior número de clientes registrados

A Figura 20 mostra os 10 estados com maior volume de encomendas, mantendo um padrão semelhante ao da distribuição de clientes. A coincidência entre os estados mais populosos e os que apresentam mais compras reflete uma correlação direta entre densidade de utilizadores e volume de transações.

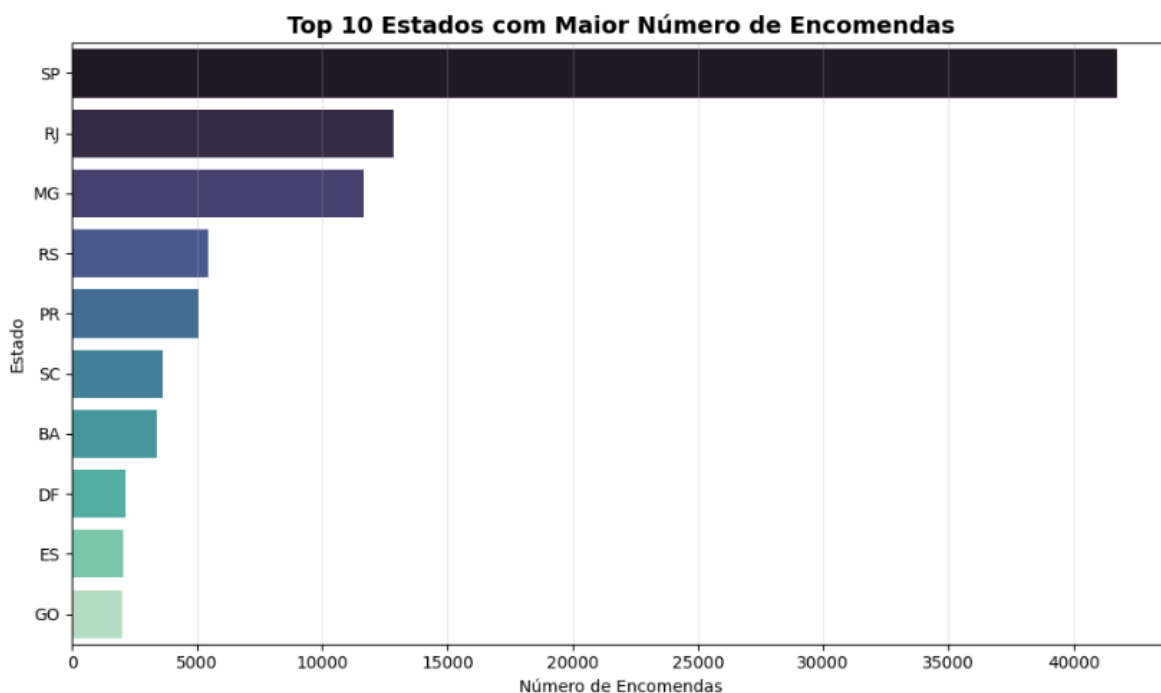


Figura 20. Distribuição dos 10 estados com maior número de encomendas realizadas



#### 4.2.1. Perguntas de Negócio

As perguntas de negócio orientam a concepção dos *dashboards* e a seleção dos indicadores de desempenho do sistema de BI, assegurando que a análise de dados responde a necessidades estratégicas bem definidas. Para que estas perguntas sejam claras, mensuráveis e úteis, recorreu-se à metodologia SMART, descrita por (Boogaard, 2023), que estabelece que os objetivos devem ser Específicos, Mensuráveis, Atingíveis, Relevantes e Temporais (Boogaard, 2023).

Segundo esse artigo, “SMART” corresponde a *Specific, Measurable, Achievable, Relevant e Time-Bound*, ou seja, Específicas, Mensuráveis, Atingíveis, Relevantes e Temporais. A definição destes parâmetros“ ajuda a garantir que os objetivos são alcançáveis dentro de um determinado prazo, elimina generalidades e facilita o acompanhamento do progresso e a identificação de marcos em falta” (Boogaard, 2023).

Aplicando esta abordagem, cada pergunta de negócio foi formulada de modo a:

- Ser específica, indicando a área a analisar;
- Ser mensurável, permitindo acompanhar resultados com dados do conjunto Olist;
- Ser atingível, compatível com a informação disponível;
- Ser relevante, alinhada com os objetivos estratégicos do *e-commerce*;
- E estar associada a um período temporal bem definido;

De seguida, na Tabela 7, apresentam-se as principais perguntas de negócio, elaboradas de acordo com os princípios SMART e que servirão de base à definição de KPIs e ao desenho dos *dashboards*.

Tabela 7. Perguntas de negócio que suportam a definição de KPIs e *dashboards*

| ID  | Pergunta de Negócio   | Área de Análise        |
|-----|---|------------------------|
| Q01 | Como evoluiu o desempenho global ao longo do tempo, em termos de encomendas, receita e lucro?                                     | Vendas                 |
| Q02 | Que categorias e produtos mais contribuem para as vendas e o lucro?   | Produtos               |
| Q03 | Qual é o desempenho logístico em termos de tempo de entrega e cumprimento de prazos, e como varia ao longo do tempo e por região? | Logística              |
| Q04 | Como se caracteriza o comportamento dos clientes em termos de volume, recorrência, perfil e padrões temporais de compra?          | Clientes               |
| Q05 | Como se distribui o desempenho económico pelos vendedores e qual o grau de concentração do lucro no <i>marketplace</i> ?          | Vendedores             |
| Q06 | Qual é o nível de satisfação dos clientes e de que forma este se relaciona com o desempenho das entregas?                         | Avaliações / Logística |

#### 4.2.2. User Stories

Nesta Secção apresentam-se as *User Stories* que descrevem as principais necessidades analíticas e objetivos de negócio considerados no desenvolvimento do sistema de BI. Estas histórias foram definidas com o objetivo de traduzir, de forma simples e objetiva, as questões de análise e os resultados esperados a partir dos *dashboards* construídos, evidenciando o valor acrescentado que cada funcionalidade proporciona ao processo de monitorização e apoio à decisão.

Na Tabela 8 são apresentadas as principais *User Stories* identificadas, organizadas por objetivo analítico e respetivo benefício para o processo de análise e tomada de decisão.

Tabela 8. Definição de *User Stories* para o Desenvolvimento do Sistema de BI

| ID   | Objetivo  | Benefício   |
|------|---|---|
| US01 | Analisar o desempenho logístico através do tempo médio de entrega e do cumprimento de prazos por estado e ao longo do tempo | Identificar ineficiências logísticas e apoiar ações de melhoria operacional     |
|      | Avaliar o desempenho das categorias e subcategorias de  | Apoiar decisões sobre gestão do portefólio de produtos e estratégias comerciais |

|             |  |   |
|-------------|--|---|
| <b>US02</b> | produto em termos de volume de vendas, lucro e recorrência   |   |
| <b>US03</b> | Analisar a evolução temporal do desempenho global do negócio, incluindo encomendas, receita e lucro            | Monitorizar o crescimento do <i>e-commerce</i> e identificar padrões sazonais                     |
| <b>US04</b> | Aceder a indicadores-chave do negócio, como receita total, número de encomendas e avaliação média dos clientes | Suportar decisões estratégicas de forma rápida e informada  |
| <b>US05</b> | Analisar o comportamento dos clientes quanto à dimensão da base, recorrência e padrões temporais de compra     | Apoiar estratégias de retenção, fidelização e segmentação de clientes                             |
| <b>US06</b> | Avaliar a satisfação dos clientes e a sua relação com o desempenho das entregas                                | Compreender os fatores que influenciam a experiência do cliente e melhorar a qualidade do serviço |
| <b>US07</b> | Analisar o desempenho económico dos vendedores e a concentração das vendas no <i>marketplace</i>               | Identificar dependências excessivas e riscos associados a determinados vendedores                 |

#### 4.2.3. Requisitos Funcionais e Não Funcionais

Nesta Secção apresentam-se os requisitos funcionais e não funcionais definidos para o desenvolvimento do sistema de BI, com foco nos *dashboards* destinados à análise do *e-commerce* da Olist. Estes requisitos foram identificados a partir das necessidades expressas nas *User Stories* e visam garantir que a solução final responde de forma eficaz aos objetivos operacionais, táticos e estratégicos da organização.

Segundo Rebrova (2024), os requisitos funcionais correspondem às funcionalidades que a solução deve possuir para garantir que se comporta conforme esperado, permitindo aos utilizadores realizar tarefas específicas. Já os requisitos não funcionais referem-se às condições gerais que a solução deve cumprir para funcionar corretamente, assegurando atributos como usabilidade, fiabilidade e eficiência (Rebrova, 2024).

De seguida, na Tabela 9, são apresentados os principais requisitos funcionais identificados para o desenvolvimento do sistema de BI, definidos a partir das necessidades expressas nas *User Stories*.

Tabela 9. Requisitos Funcionais

| ID   | Requisito Funcional                  | Descrição   |
|------|--------------------------------------|---|
| RF01 | Filtros e Segmentadores              | Permitir filtrar a informação por ano de análise, categoria de produto e outras dimensões relevantes, garantindo consistência dos filtros entre páginas |
| RF02 | Indicadores Principais               | Disponibilizar indicadores-chave como receita total, lucro, número de encomendas, número de clientes, valor médio por encomenda e avaliação média       |
| RF03 | Análise Temporal                     | Apresentar a evolução dos principais indicadores ao longo do tempo, permitindo análise por ano, trimestre e mês   |
| RF04 | Análise Logística                    | Mostrar o tempo médio de entrega, a percentagem de entregas no prazo e de entregas atrasadas, com possibilidade de segmentação temporal e geográfica    |
| RF05 | Análise Geográfica                   | Visualizar o desempenho do negócio por estado, incluindo volume de encomendas, taxa de atraso e tempo médio de entrega                                  |
| RF06 | Análise de Produtos                  | Analisar o desempenho das categorias e subcategorias de produto em termos de vendas, lucro e recorrência de compras                                     |
| RF07 | Análise de Pagamentos                | Apresentar a distribuição das encomendas por método de pagamento e número de prestações   |
| RF08 | Avaliações dos Clientes              | Mostrar a avaliação média, a distribuição das classificações e a sua evolução temporal  |
| RF09 | Relação entre Logística e Satisfação | Permitir analisar a relação entre o cumprimento dos prazos de entrega e as avaliações atribuídas pelos clientes   |
| RF10 | Interatividade entre Visualizações   | Garantir que a interação com uma visualização atualiza automaticamente os restantes elementos da página   |
| RF11 | Navegação entre Páginas              | Permitir a navegação entre os diferentes <i>dashboards</i> através de um menu comum e consistente   |
| RF12 | Apoio à Exploração da Informação     | Disponibilizar funcionalidades como <i>tooltips</i> , alternância de visualizações e controlo da granularidade temporal                                 |

Complementarmente, foram também definidos os requisitos não funcionais que o sistema de BI deve cumprir. Estes requisitos não estão relacionados diretamente com as funcionalidades disponibilizadas, mas sim com as características de qualidade e desempenho que garantem a usabilidade, a fiabilidade e a consistência da solução. A Tabela 10 apresenta os principais requisitos não funcionais considerados essenciais para o sistema desenvolvido.

Tabela 10. Requisitos Não Funcionais

| ID    | Requisito Não Funcional | Descrição   |
|-------|-------------------------|---|
| RNF01 | Usabilidade             | A interface deve ser intuitiva e de fácil utilização, permitindo que utilizadores sem conhecimentos técnicos avançados consigam explorar os <i>dashboards</i> de forma autónoma |
| RNF02 | Desempenho              | O tempo de resposta na aplicação de filtros, interações e carregamento das visualizações deve ser adequado, não comprometendo a experiência do utilizador                       |
| RNF03 | Fiabilidade             | Os <i>dashboards</i> devem apresentar resultados consistentes e coerentes com os dados de origem, garantindo confiança na informação analisada                                  |
| RNF04 | Compatibilidade         | A solução deve ser acessível em diferentes dispositivos (desktop, portátil e tablet), mantendo uma visualização adequada dos <i>dashboards</i>                                  |

#### 4.2.4. KPIS e Métricas Relevantes

Os KPIs e as métricas associadas constituem a base para monitorizar e avaliar o desempenho do sistema de BI aplicado ao *e-commerce* da Olist. Depois de definidas as perguntas de negócio, as *User Stories* e os requisitos funcionais e não funcionais, torna-se essencial identificar os indicadores que permitirão medir, de forma objetiva e contínua, o cumprimento dos objetivos estratégicos, táticos e operacionais.

De seguida, apresenta-se a Tabela 11 com os principais KPIs e métricas identificados para o sistema de BI da Olist, agrupados de acordo com as áreas de análise que servirão de base para a construção dos *dashboards* e para a monitorização contínua do negócio.

Tabela 11. KPIs e métricas relevantes para o sistema de BI da Olist

| ID  | Área de Análise | KPI / Métrica        | Descrição   |
|-----|-----------------|----------------------|---|
| K01 | Vendas          | Receita Total        | Soma do valor das encomendas concluídas no período em análise |
| K02 | Vendas          | Número de Encomendas | Total de pedidos concluídos                                   |
| K03 | Vendas          | Lucro Total          | Valor total de lucro gerado pelas encomendas concluídas       |

|            |               |                                |   |
|------------|---------------|--------------------------------|---|
| <b>K04</b> | Vendas        | AOV                            | Receita total dividida pelo número de encomendas                      |
| <b>K05</b> | Vendas        | Evolução Temporal de Vendas    | Variação temporal da receita e das encomendas (ano, trimestre e mês)  |
| <b>K06</b> | Produtos      | Volume de Vendas por Categoria | Quantidade de unidades vendidas por categoria de produto              |
| <b>K07</b> | Produtos      | Lucro por Categoria            | Contributo de cada categoria para o lucro total                       |
| <b>K08</b> | Produtos      | Recorrência de Compras         | Classificação dos produtos segundo a frequência de compra             |
| <b>K09</b> | Clientes      | Número de Clientes             | Total de clientes distintos no período analisado                      |
| <b>K10</b> | Clientes      | Clientes Novos vs. Recorrentes | Distribuição entre clientes de primeira compra e clientes recorrentes |
| <b>K11</b> | Clientes      | Receita Média por Cliente      | Receita média gerada por cliente                                      |
| <b>K12</b> | Logística     | Tempo Médio de Entrega         | Média de dias entre a compra e a entrega da encomenda                 |
| <b>K13</b> | Logística     | Taxa de Entregas no Prazo      | Percentagem de encomendas entregues até à data prevista               |
| <b>K14</b> | Logística     | Taxa de Entregas Atrasadas     | Percentagem de encomendas entregues fora do prazo                     |
| <b>K15</b> | Avaliações    | Avaliação Média                | Média das classificações atribuídas pelos clientes                    |
| <b>K16</b> | Avaliações    | Distribuição das Avaliações    | Percentagem de avaliações por nível (1 a 5 estrelas)                  |
| <b>K17</b> | Avaliações    | Relação Entrega–Avaliação      | Análise da relação entre atraso na entrega e avaliação atribuída      |
| <b>K18</b> | Vendedores    | Número de Vendedores Ativos    | Total de vendedores com vendas registadas                             |
| <b>K19</b> | Vendedores    | Lucro por Vendedor             | Lucro total gerado por vendedor                                       |
| <b>K20</b> | Vendedores    | Concentração do Lucro          | Percentagem do lucro concentrada nos principais vendedores            |
| <b>K21</b> | Cancelamentos | Taxa de Cancelamento           | Percentagem de encomendas canceladas face ao total                    |

### 4.3. Definição da Narrativa Analítica

Nesta Secção é definida a estrutura narrativa que orienta a análise dos *dashboards* desenvolvidos. O objetivo é definir o enquadramento conceptual e metodológico que orienta a interpretação dos indicadores de desempenho, garantindo uma abordagem coerente e orientada à decisão.

Como discutido na Secção 2.5.2, o *data storytelling* resulta da articulação entre dados, narrativa e visualizações, permitindo comunicar resultados analíticos de forma clara, contextualizada e orientada a objetivos (Daradkeh & Atalla, 2023). Neste trabalho, os *dashboards* desenvolvidos constituem o suporte visual da narrativa, enquanto a narrativa aqui definida determina a sequência lógica, o foco interpretativo e as questões orientadoras que estruturam a exploração dos dados nas secções seguintes.

O problema central abordado nesta análise prende-se com a necessidade de compreender o desempenho global do *e-commerce* da Olist, identificando fatores que explicam o crescimento do negócio, o comportamento dos clientes, a eficiência logística, a concentração do desempenho económico e o impacto destes elementos na satisfação do cliente. Face à diversidade e volume dos dados disponíveis, torna-se essencial uma abordagem estruturada que permita não apenas descrever resultados, mas também explicar relações e apoiar decisões informadas.

Neste contexto, a narrativa é estruturada com base no modelo 5W1H, proposto por Daradkeh & Atalla (2023) assegurando que a narrativa responde às principais questões associadas ao processo de decisão:

- **Who:** principais intervenientes do *marketplace*, nomeadamente clientes e vendedores;
- **What:** desempenho comercial, logístico e nível de satisfação;
- **When:** evolução temporal entre 2016 e 2018, com identificação de sazonalidade;
- **Where:** distribuição geográfica por estado e categoria;
- **Why:** compreender os fatores que influenciam crescimento, eficiência e satisfação;
- **How:** utilização de *dashboards* como meio de exploração e comunicação dos dados.

Em paralelo, a estrutura da análise segue implicitamente o modelo SPSN (Tewari, 2019). Inicialmente, é apresentado o contexto atual do negócio (situação), seguido da identificação de padrões e desafios, como a concentração do lucro, a pressão logística em períodos de pico e a baixa taxa de recorrência de clientes (problema). A solução consiste na utilização dos *dashboards* desenvolvidos como apoio à análise e à decisão, permitindo explorar os dados de forma integrada. Por fim, a análise conduz à identificação de oportunidades de melhoria e recomendações, que poderão servir de base a ações futuras (próximos passos).

Adicionalmente, a apresentação dos resultados procura respeitar os princípios do modelo SUCCESS, privilegiando uma narrativa simples, concreta e credível, suportada por dados quantitativos e visualizações consistentes (Tewari, 2019). Sempre que possível, a análise destaca padrões inesperados, relações relevantes e implicações práticas para o negócio, reforçando o papel do *data storytelling* como meio de comunicação eficaz entre dados e decisão.

Deste modo, os *insights* que se pretende obter nesta Secção centram-se em:

- compreender os motores de crescimento do negócio;
- identificar padrões de consumo e comportamento dos clientes;
- avaliar o impacto da logística na satisfação do cliente;
- analisar a concentração do desempenho económico por produtos e vendedores;
- apoiar a definição de estratégias orientadas para eficiência, fidelização e sustentabilidade do *e-commerce*.

#### **4.4. Modelo Dimensional**

O *dataset* original da Olist apresenta uma estrutura relacional adequada à exploração inicial dos dados, baseada em tabelas normalizadas interligadas por chaves técnicas, conforme ilustrado na Figura 22. Embora esta organização seja suficiente para análises exploratórias, revela limitações quando aplicada a um sistema de BI orientado à análise multidimensional, à definição de métricas e ao desempenho analítico.

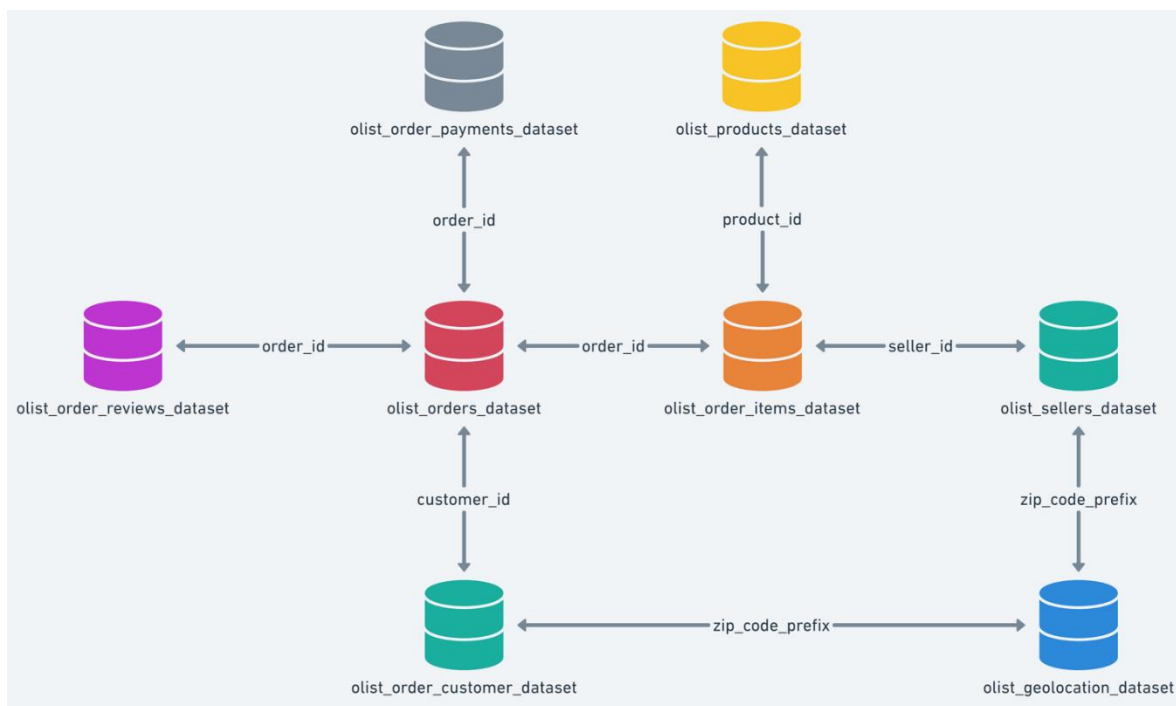


Figura 22. Modelo relacional original do *dataset* Olist

Em particular, o modelo original não separa explicitamente factos e dimensões, não normaliza o tempo como uma dimensão independente e concentra múltiplos eventos temporais em colunas dispersas, o que dificulta análises comparativas ao longo do tempo, a reutilização de métricas e a construção eficiente de *dashboards* analíticos. Adicionalmente, a ausência de hierarquias temporais limita análises agregadas por períodos como semana, mês, trimestre ou hora do dia.

Com base nestas limitações, foi definido um novo modelo dimensional orientado pelas perguntas de negócio, user stories e KPIs identificados na Secção 4.2. O modelo adotado segue inicialmente o esquema em estrela (*star schema*), sendo posteriormente expandido para uma constelação de estrelas (*galaxy schema*), de modo a integrar múltiplos processos de negócio interligados ao longo do ciclo de vida de uma encomenda.

O modelo proposto é constituído por quatro tabelas factuais, *FactOrders*, *FactOrderItems*, *FactPayments* e *FactReviews*, e por oito dimensões partilhadas, *DimCustomer*, *DimSeller*, *DimProduct*, *DimPaymentMethod*, *DimCategory*, *DimSubCategory*, *DimCalendar* e *DimHour*. Esta reorganização permite representar os diferentes níveis de granularidade do negócio, desde a encomenda até ao item, ao pagamento e à avaliação do cliente.

Uma das principais mais-valias do modelo desenvolvido face ao modelo original reside na introdução explícita das dimensões temporais *DimCalendar* e *DimHour*. No *dataset* original,

as informações temporais encontram-se distribuídas por múltiplos campos de data e hora nas tabelas de encomendas e avaliações, sem uma estrutura comum. A criação destas dimensões permite normalizar o tempo, suportando análises consistentes por dia, semana, mês, trimestre, ano e hora do dia, bem como a reutilização destas hierarquias temporais em todas as tabelas factuais. Esta abordagem simplifica a construção de métricas temporais e melhora o desempenho analítico, ao evitar cálculos repetidos em tempo de execução.

A tabela *FactOrders* representa o processo de compra ao nível da encomenda, com uma granularidade de uma linha por pedido. Para além das referências temporais associadas às várias fases do pedido, foram criados campos calculados que não existiam no modelo original, como *days\_to\_approve*, *days\_to\_ship*, *days\_to\_deliver*, *days\_late\_delivery* e *is\_late\_delivery*. Estes indicadores permitem avaliar de forma direta o desempenho operacional e logístico, transformando dados transacionais brutos em métricas analíticas prontas a consumir.

A *FactOrderItems* apresenta um nível de detalhe ao nível do produto vendido, permitindo análises mais detalhadas de preços, portes e margens. O cálculo de métricas como *total\_item\_value* e *margin\_estimate* acrescenta valor ao modelo, ao integrar regras de negócio na camada dimensional e simplificar as análises posteriores.

As tabelas *FactPayments* e *FactReviews* representam processos de negócio distintos que continuam associados à tabela de encomendas através das dimensões partilhadas. A sua separação no modelo permite analisar métodos de pagamento, prestações, valores pagos e avaliações dos clientes de forma mais específica, mantendo a possibilidade de cruzamento analítico. No caso das avaliações, a introdução do indicador *days\_to\_respond* acrescenta uma dimensão temporal adicional, permitindo analisar o tempo de resposta dos clientes no processo de feedback.

As dimensões partilhadas fornecem o contexto necessário para a análise integrada. A *DimProduct*, em conjunto com *DimCategory* e *DimSubCategory*, explicita a hierarquia de produtos, que no modelo original não existia. A *DimCustomer* e a *DimSeller* normalizam a informação geográfica e demográfica, enquanto a separação entre dimensões e factos melhora a legibilidade do modelo e facilita a sua extensão futura.

Em síntese, o modelo dimensional proposto representa uma evolução face ao modelo original da Olist, ao transformar uma estrutura relacional orientada a dados operacionais

num modelo analítico otimizado para BI. O Anexo C apresenta o modelo dimensional final, evidenciando as relações entre tabelas factuais e dimensões que suportam uma análise integrada, coerente e extensível do negócio.

#### **4.5. Arquitetura do Sistema de BI**

A arquitetura do sistema de BI desenvolvido para o *e-commerce* da Olist foi concebida de forma a garantir consistência, qualidade e desempenho analítico, assegurando que os dados provenientes do *dataset* original são preparados e disponibilizados de forma adequada para suporte à análise e à tomada de decisão.

Uma vez que os dados utilizados são estáticos e não sofrem atualizações ao longo do tempo, optou-se por uma abordagem simplificada, centrada na limpeza e transformação dos ficheiros CSV originais do *dataset*. Não foi, por isso, implementado um processo de ETL clássico nem uma camada intermédia de armazenamento, como um DW. Após o tratamento dos dados, foram gerados novos ficheiros CSV, já consolidados e preparados para análise.

Estes ficheiros foram posteriormente importados diretamente para o Power BI, onde foi realizada a modelação dos dados necessária à construção dos *dashboards* analíticos.

A Figura 23 ilustra o fluxo geral do sistema, desde os ficheiros de dados de origem até à camada de visualização analítica no Power BI.

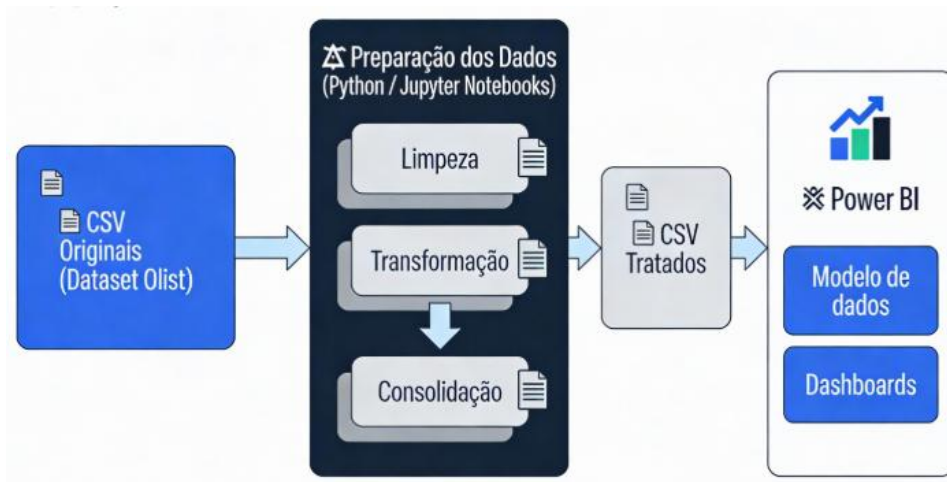


Figura 23. Arquitetura geral do sistema de BI da Olist

A Figura 24 apresenta um excerto do modelo de dados implementado no Power BI, evidenciando algumas das principais tabelas e as respectivas relações, que suportam a análise realizada nos *dashboards*.

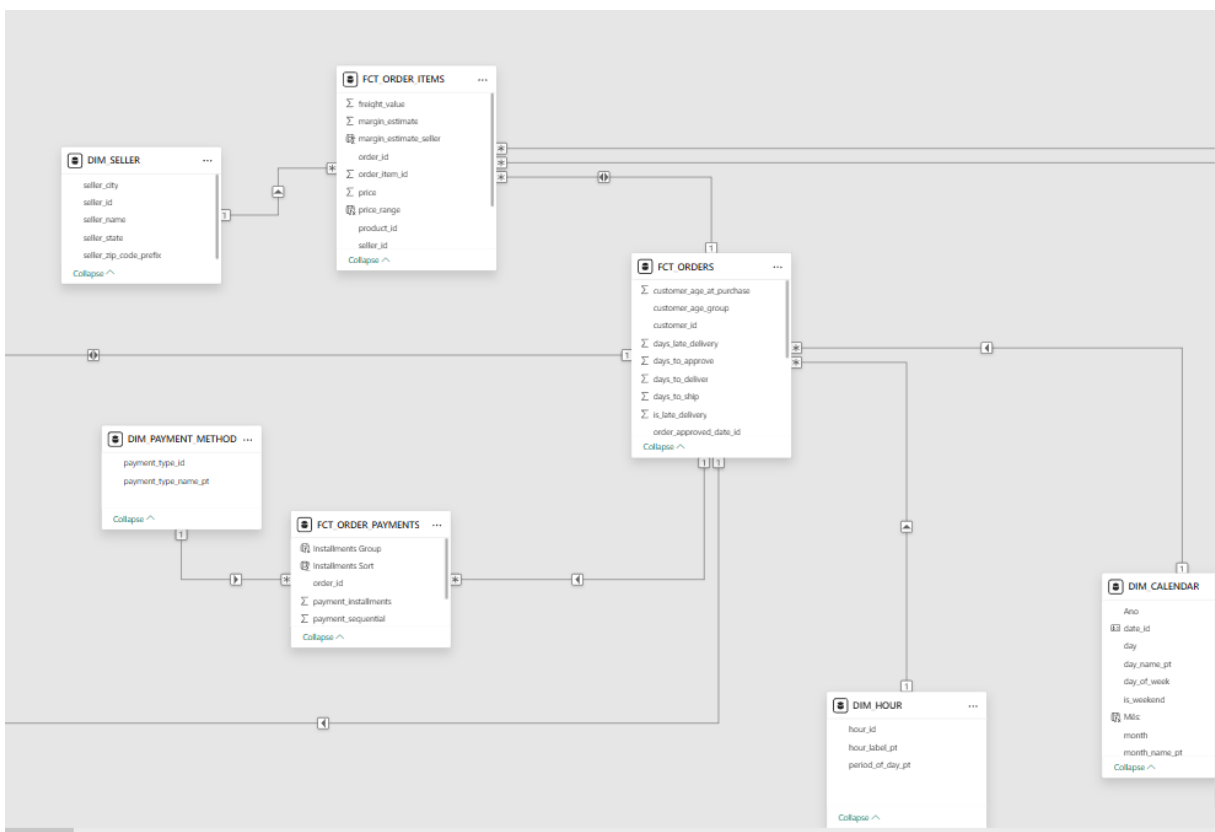


Figura 24. Excerto do modelo de dados desenvolvido no Power BI

## 5. Processo de Preparação, Limpeza e Transformação dos Dados

O presente capítulo descreve o processo de preparação, limpeza e transformação com o objetivo de suportar o modelo dimensional apresentado no capítulo anterior.

Os dados utilizados provêm dos diferentes ficheiros CSV disponibilizados no *dataset* do *e-commerce* da Olist, tendo sido selecionadas apenas as fontes consideradas relevantes para a análise. O ficheiro *geolocation\_dataset.csv* não foi utilizado, uma vez que a informação relativa ao código postal, cidade e estado já se encontra presente nas tabelas de clientes e vendedores, tornando a sua utilização redundante.

O processo de preparação dos dados incidiu na limpeza e normalização da informação, de forma a garantir a consistência e integridade das tabelas dimensionais e factuais definidas no modelo dimensional. Este processo incluiu, entre outras operações, a remoção de inconsistências, a padronização de formatos e a criação das chaves necessárias para estabelecer as relações entre as diferentes tabelas.

Como resultado deste processo, foram gerados novos ficheiros de dados tratados, que serviram de base para a construção do modelo de dados no Power BI e para o desenvolvimento dos *dashboards* apresentados nos capítulos seguintes.

### 5.1. Preparação de Dados de Origem

Esta fase incidiu na seleção e preparação dos ficheiros de dados de origem que compõem o *dataset* público da Olist, disponibilizado na plataforma Kaggle. O conjunto de dados é constituído por oito ficheiros em formato CSV, cada um representando uma componente distinta do processo de *e-commerce*, desde a realização dos pedidos até às avaliações efetuadas pelos clientes.

Os ficheiros originais, que representam os dados no seu estado inicial (*raw data*), foram organizados numa estrutura de diretórios local, sendo armazenados na pasta *01\_Raw\_Data*. Posteriormente, estes ficheiros foram carregados para o ambiente de trabalho *Jupyter Notebook*, onde foram lidos utilizando a linguagem *Python* e a biblioteca *pandas*.

## 5.2. Limpeza e Transformação de Dados

Após a preparação dos dados de origem, procedeu-se à limpeza e transformação da informação, com o objetivo de garantir a sua consistência, integridade e adequação ao modelo dimensional definido. Esta etapa centrou-se no tratamento e normalização dos dados provenientes das tabelas *raw*, assegurando que a informação utilizada na análise se encontrava corretamente estruturada e preparada.

Este processo foi desenvolvido no ambiente *Jupyter Notebook*, recorrendo à linguagem *Python* e a bibliotecas especializadas em manipulação e processamento de dados, nomeadamente *pandas* e *numpy*.

Para efeitos de organização e manutenção do trabalho desenvolvido, foi criada uma estrutura de diretórios local designada *02\_Transformations*, destinada a armazenar os *notebooks* responsáveis pelas diferentes etapas do processo de transformação.

Nos subcapítulos seguintes são apresentados os principais procedimentos aplicados no processo de preparação e transformação dos dados do *dataset* da Olist, incluindo operações gerais de validação, tratamento de valores em falta, normalização de atributos e desenvolvimento de dimensões temporais.

### 5.2.1. Procedimentos Gerais

Nesta fase procedeu-se à leitura dos ficheiros originais armazenados na pasta *01\_Raw\_Data*, correspondentes aos diferentes *datasets* do *marketplace* Olist. Foram realizadas verificações de qualidade transversais a todas as tabelas, incluindo análise da estrutura, contagem de registos, validação de chaves primárias e deteção de valores nulos e duplicados.

Adicionalmente, procedeu-se à conversão de tipos de dados (nomeadamente *datas* e valores monetários), à decomposição de atributos temporais para integração com dimensões de tempo e à criação de métricas derivadas relevantes para análise, como indicadores de desempenho logístico e estimativas de margem.

Estes procedimentos asseguraram a integridade, coerência e compatibilidade dos dados com o modelo dimensional definido.

### 5.2.2. Tratamento de Valores por Preencher

Durante a preparação dos dados foram identificados valores em falta em diferentes atributos, exigindo estratégias diferenciadas de tratamento.

Por exemplo, na dimensão de produtos, os valores nulos na variável *product\_category\_name* foram substituídos por uma categoria genérica designada “Outro”, garantindo que todos os registos permanecessem classificados. Já nas variáveis relacionadas com dimensões físicas (peso, comprimento, altura e largura), os valores ausentes foram substituídos pela mediana da respetiva coluna, garantindo maior consistência na análise.

### 5.2.3. Normalização de Dados

A normalização incidiu principalmente sobre atributos geográficos e categóricos, garantindo maior consistência e uniformidade.

Foi realizada a padronização dos nomes das cidades e estados através da integração com ficheiros de referência oficiais (*cities.json* e *states.json*), eliminando variações ortográficas e discrepâncias de formatação. A Figura 25 apresenta um exemplo ilustrativo do processo de normalização e da criação da chave composta utilizada para garantir consistência geográfica.

## 5. Normalização de Cidades e Estados

```
# Função auxiliar para normalizar texto
def normalize_text(text):
    if pd.isna(text):
        return np.nan
    text = text.strip().lower()
    text = unicodedata.normalize("NFKD", text).encode("ASCII", "ignore").decode("utf-8")
    return text

# Aplicar normalização
df_customers["customer_city_norm"] = df_customers["customer_city"].apply(normalize_text)
df_customers["city_key"] = (
    df_customers["customer_city_norm"] + "_" + df_customers["customer_state"]
)
```

Figura 25. Normalização dos nomes de cidades

Adicionalmente, os métodos de pagamento foram mapeados e traduzidos para designações em português, garantindo coerência interpretativa.

Procedeu-se ainda à uniformização de estados de encomenda e à criação de identificadores únicos compostos quando necessário, assegurando integridade referencial no modelo dimensional.

#### **5.2.4. Desenvolvimento da Dimensão Calendário e Dimensão Hora**

De forma a suportar análises temporais consistentes e flexíveis, foi criada uma dimensão de calendário abrangendo todo o intervalo temporal presente nos dados transacionais. Esta dimensão inclui atributos hierárquicos como dia, mês, trimestre, ano, semana e indicador de fim de semana.

Complementarmente, foi desenvolvida uma dimensão de horas contendo todos os minutos do dia, possibilitando análises com maior granularidade temporal.

### **5.3. Importação de Dados no Power BI**

Após a fase de limpeza e transformação, os ficheiros CSV finais foram organizados na pasta *03\_Output* e importados para o Power BI. Esta etapa permitiu disponibilizar os dados tratados na ferramenta de visualização, servindo de base à construção do modelo de dados e dos *dashboards* analíticos.

No Power BI, foi realizada a modelação dos dados, estabelecendo-se as relações entre as tabelas de acordo com o modelo definido.

No Anexo D é apresentada uma imagem ilustrativa do modelo de dados desenvolvido no Power BI.

## 6. Desenvolvimento dos Dashboards e Análise de Resultados

Este capítulo descreve o desenvolvimento dos *dashboards* do sistema de BI aplicado ao *e-commerce* da Olist e a análise dos resultados obtidos a partir da sua utilização. São apresentadas as principais etapas da concepção e implementação dos painéis, bem como as funcionalidades transversais que suportam a navegação, a interação e a exploração da informação.

Adicionalmente, é realizada uma análise integrada dos resultados com base numa abordagem de *data storytelling*, permitindo interpretar os principais indicadores, identificar padrões e tendências relevantes e transformar a informação visual em conhecimento útil para o apoio à tomada de decisão.

### 6.1. Concepção dos Dashboards: Mockups

Antes do desenvolvimento dos *dashboards*, foram concebidos diferentes *mockups* com o objetivo de apoiar a fase de planeamento e orientar a implementação da solução final. No total, foram desenvolvidos seis *mockups*, correspondentes aos diferentes painéis previstos. Neste trabalho, são apresentados dois desses *mockups*, nomeadamente os painéis Sumário e Produtos, por serem representativos da estrutura e dos princípios de design adotados. Estes *mockups* encontram-se incluídos no Anexo E .

A concepção dos *mockups* teve como principal objetivo garantir um estilo visual consistente entre os diferentes painéis, evitando confusão na interpretação da informação e facilitando a navegação entre *dashboards*.

Os *mockups* foram desenvolvidos com base em princípios como a organização da informação por categorias, a definição de hierarquias visuais e a coordenação entre diferentes vistas. A hierarquia visual foi utilizada para destacar os indicadores mais relevantes e orientar a leitura inicial do *dashboard*, enquanto a coordenação entre diferentes vistas permitiu que múltiplas visualizações apresentassem perspetivas complementares dos mesmos dados. Estes princípios contribuem para uma exploração mais clara e estruturada da

informação, apoiando a análise em contextos mais complexos, em linha com o referido por Bach et al. (2023).

Durante esta fase, procurou-se também assegurar que cada painel fosse simples e fácil de compreender, mantendo uma identidade visual comum em toda a solução. A clareza da informação, a simplicidade na disposição dos elementos e a consistência no uso de cores e posicionamento dos indicadores foram considerados aspetos essenciais. Segundo Green et al. (2023), uma visualização eficaz deve permitir ao utilizador compreender rapidamente a mensagem principal.

## 6.2. Implementação dos Dashboards

A implementação dos *dashboards* teve início com a integração dos dados provenientes de ficheiros CSV, correspondentes às diferentes áreas de análise consideradas no projeto. Estes ficheiros correspondem ao output da etapa de limpeza e transformação de dados, encontrando-se organizados na pasta *03\_Output*, e serviram de base para a construção da solução analítica, garantindo uma visão integrada da informação necessária ao suporte à tomada de decisão.

Após a integração dos ficheiros CSV, foi construído um modelo de dados, com o objetivo de organizar a informação e permitir a correta ligação entre as diferentes entidades. Este modelo serviu de base para o cálculo dos indicadores e para a construção das visualizações utilizadas nos *dashboards*. O modelo de dados implementado encontra-se apresentado no Anexo D .

Com base no modelo de dados e nos *mockups* definidos na fase de conceção, foram implementadas seis páginas de *dashboards*, organizadas por áreas temáticas distintas, de forma a facilitar a navegação e a análise da informação. Esta organização permitiu separar a análise global de análises mais específicas.

A navegação entre as diferentes páginas é realizada através de um menu lateral comum a todos os *dashboards*, permitindo ao utilizador alternar facilmente entre as diferentes áreas de análise. Este menu assegura uma estrutura de navegação simples e consistente, contribuindo para uma experiência de utilização mais intuitiva.

De forma a garantir uma experiência de utilização uniforme, procurou-se manter um estilo visual consistente entre todas as páginas dos *dashboards*, alinhado com a identidade visual

da Olist. A disposição dos elementos, o tipo de visualizações e a utilização de cores foram mantidos de forma semelhante ao longo das diferentes páginas, facilitando a leitura da informação e evitando confusão na navegação.

A página *Sumário* apresenta uma visão geral do desempenho do negócio. Inclui indicadores principais no topo e gráficos de apoio que permitem acompanhar a evolução dos dados ao longo do tempo e observar informação por categoria e por estado. Na Figura 26 é apresentada a página *Sumário*.

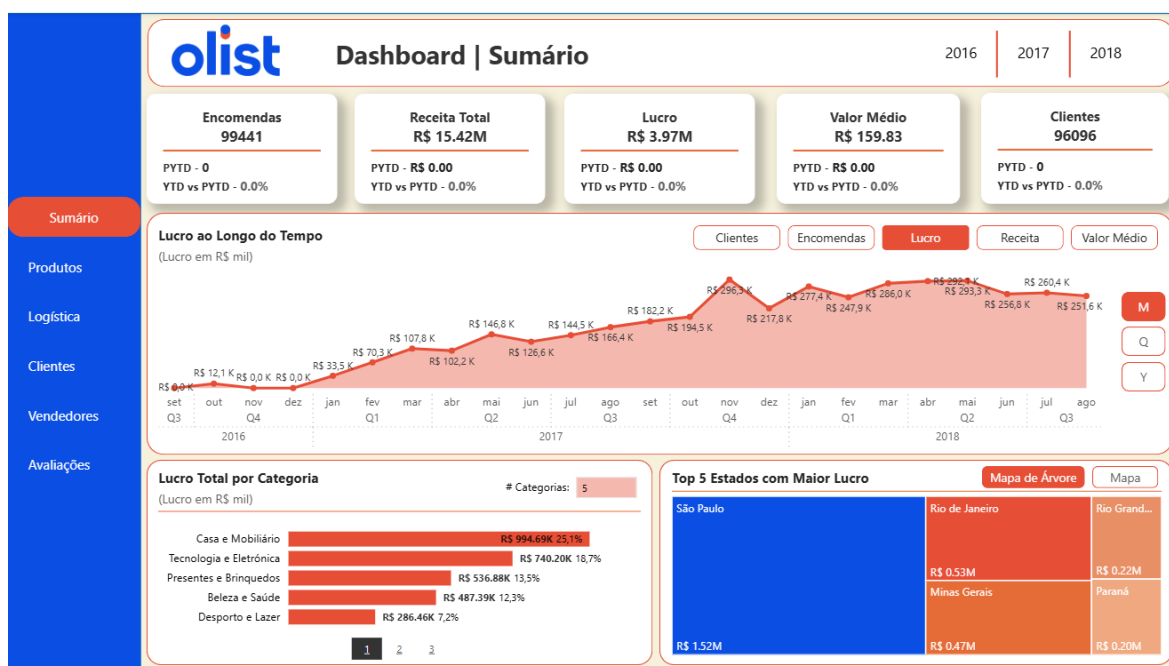


Figura 26. Dashboard Sumário

A página *Produtos* foi concebida para a análise do desempenho dos produtos. Nesta página são apresentados gráficos que permitem comparar categorias e subcategorias, analisar a recorrência de compras e acompanhar a evolução dos resultados ao longo do tempo. São também apresentadas tabelas de apoio que permitem observar a informação de forma mais detalhada. Na Figura 27 é apresentada a página *Produtos*.



Figura 27. Dashboard Produtos

A página *Logística* foi desenvolvida para apoiar a análise do processo de entrega das encomendas. Nesta página são apresentados indicadores relacionados com o número de pedidos, o tempo médio de entrega, a percentagem de entregas no prazo e a percentagem de entregas atrasadas.

Adicionalmente, são incluídos gráficos que permitem acompanhar a evolução da taxa de atraso ao longo do tempo e analisar a relação entre atrasos e volume de encomendas por estado. É também apresentada uma tabela que permite observar estes indicadores por categoria e respetivas subcategorias, facilitando uma análise mais detalhada do desempenho logístico. Na Figura 28 é apresentada a página *Logística*.

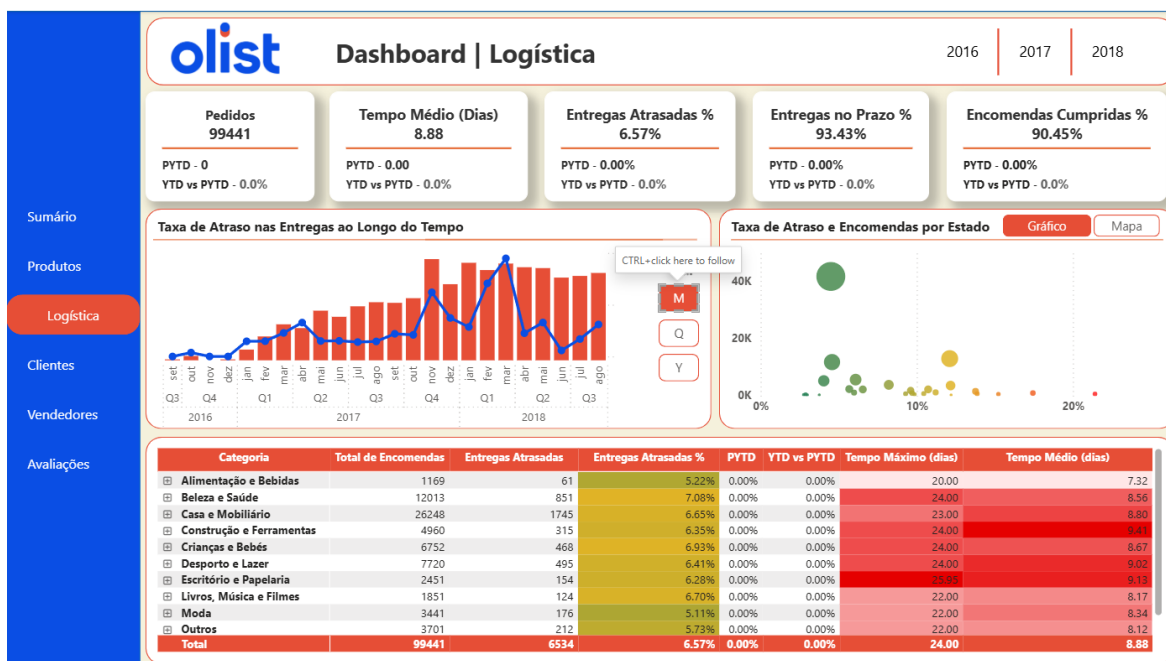


Figura 28. Dashboard Logística

A página *Clientes* foi desenvolvida para apoiar a análise do comportamento dos clientes. Nesta página são apresentados indicadores relacionados com o número de clientes, clientes recorrentes, receita total, receita média por cliente e taxa de cancelamento. São também incluídos gráficos que permitem analisar a distribuição dos clientes por faixa etária, a média de encomendas ao longo do mês, bem como o padrão de compras por dia da semana e intervalo horário, permitindo identificar os períodos de maior e menor atividade ao longo da semana. Adicionalmente, é apresentada a distribuição das encomendas por método de pagamento, assim como uma tabela de apoio que possibilita a observação desta informação de forma mais detalhada. Na Figura 29 é apresentada a página *Clientes*.

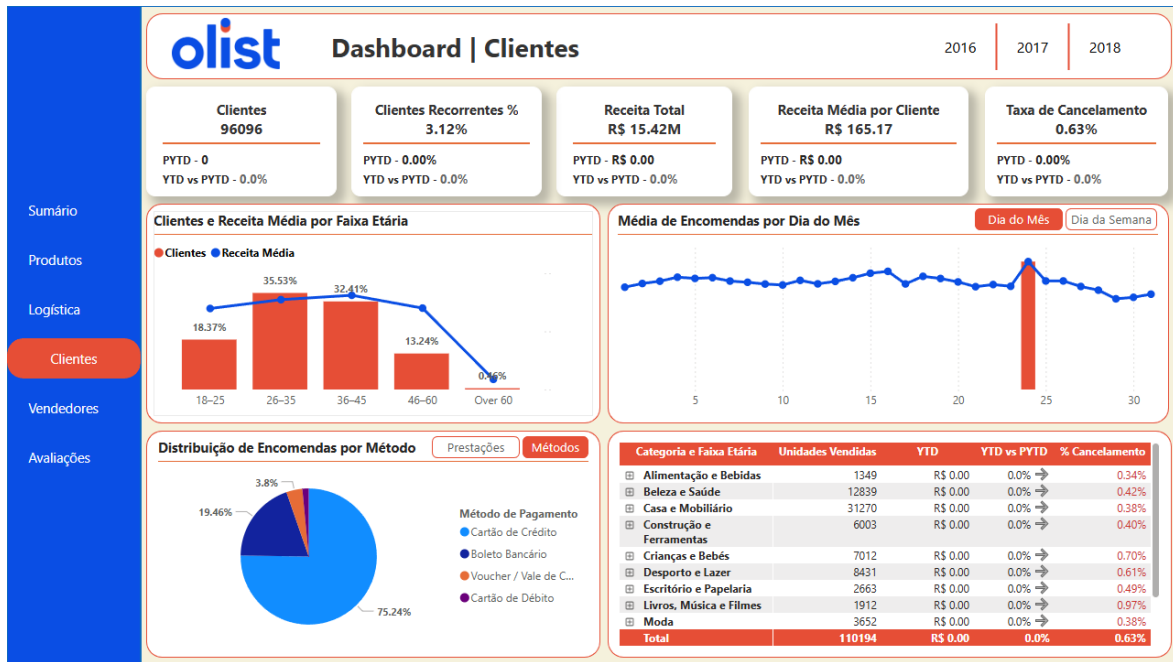


Figura 29. Dashboard Clientes

A página *Vendedores* apresenta informação sobre o desempenho dos vendedores. Inclui indicadores relacionados com lucro, número de encomendas e lucro médio, bem como uma tabela com os vendedores com maior lucro e um gráfico que permite acompanhar a evolução do lucro ao longo do tempo. Na Figura 30 é apresentada a página *Vendedores*.

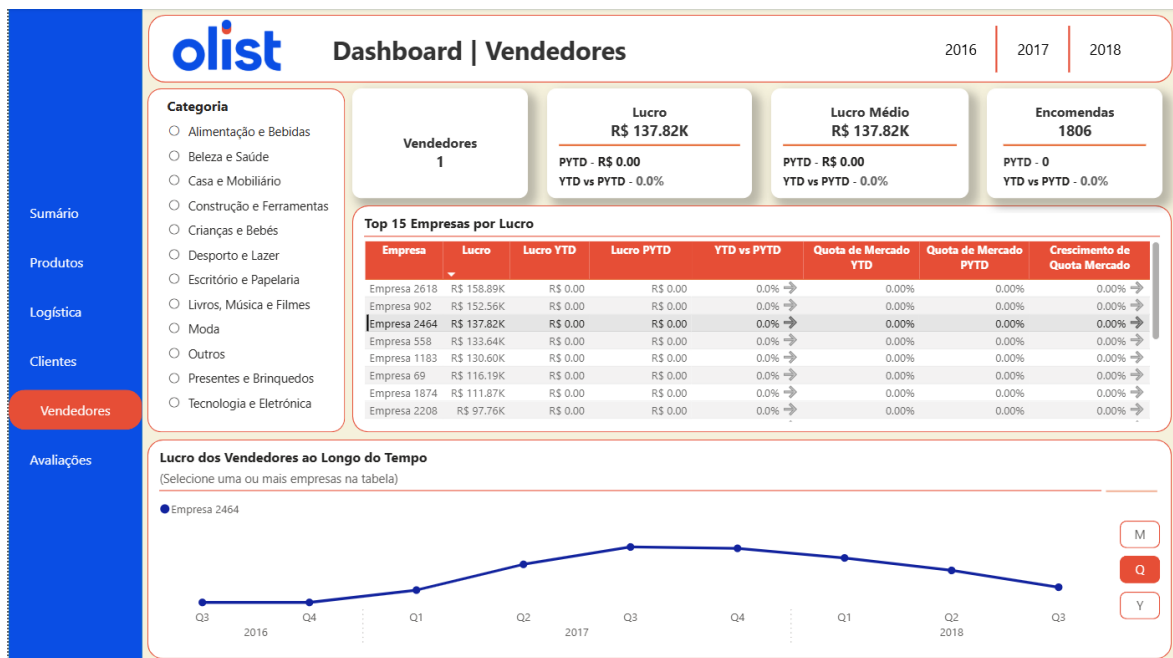


Figura 30. Dashboard Vendedores

A página *Avaliações* apresenta informação relacionada com a avaliação dos clientes às encomendas. Inclui indicadores como a avaliação média e a distribuição das classificações, bem como gráficos que permitem acompanhar a evolução das avaliações ao longo do tempo e analisar a relação entre as avaliações e o desempenho das entregas. Na Figura 31 é apresentada a página *Avaliações*.

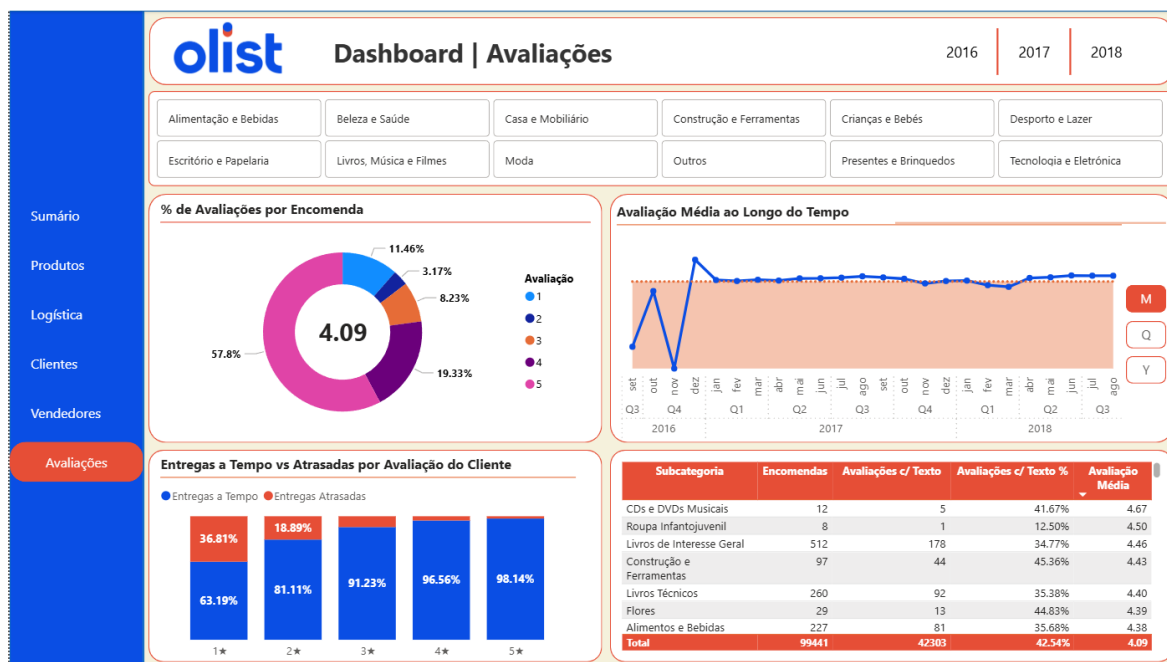


Figura 31. *Dashboard Avaliações*

Para efeitos de testes, os *dashboards* foram publicados no serviço do Power BI, permitindo validar a navegação, as visualizações e a consistência da informação apresentada.

### 6.3. Funcionalidades Transversais

Para além da estrutura das páginas apresentada na Secção 6.2, foram implementadas funcionalidades transversais comuns aos diferentes *dashboards*, com o objetivo de facilitar a navegação, melhorar a interação com os dados e tornar a análise mais flexível. Estas funcionalidades permitem ao utilizador explorar a informação ao longo das diferentes páginas, adaptando a visualização dos dados de acordo com as suas necessidades de análise.

As funcionalidades transversais incluem mecanismos de navegação, filtros e segmentadores globais, interações entre visualizações, opções de controlo temporal e funcionalidades de apoio à exploração da informação. Nos subcapítulos seguintes são descritas estas

funcionalidades, ilustrando de que forma contribuem para uma utilização mais intuitiva e eficaz dos *dashboards* desenvolvidos.

### 6.3.1. Navegação e Acesso às Páginas

Conforme explicado na Secção 6.2, a navegação entre as diferentes páginas dos *dashboards* é realizada através de um menu lateral comum a todo o relatório. Este menu permite ao utilizador aceder facilmente às diferentes áreas de análise, assegurando uma navegação simples, consistente e intuitiva ao longo de toda a solução.

### 6.3.2. Filtros e Segmentadores Globais

Foi implementado um filtro comum a todas as páginas que permite seleccionar o ano de análise, estando disponíveis os anos de 2016, 2017 e 2018, correspondentes aos períodos para os quais existem registos de compras. Este filtro possibilita a análise dos dados relativos a um ano específico, bem como a comparação entre diferentes anos. Na Figura 32 é apresentado o filtro global de ano.

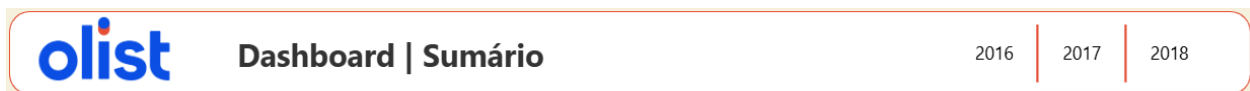


Figura 32. Filtro global de ano

Adicionalmente, nos painéis *Produtos*, *Vendedores* e *Avaliações* foi incluído um filtro comum de categorias, permitindo ao utilizador restringir a análise a categorias específicas. Esta funcionalidade facilita a exploração dos dados de forma mais direccionada, adaptando a análise aos interesses do utilizador. Na Figura 33 é apresentado um exemplo do filtro de categorias.



Figura 33. Segmentador de categorias

### 6.3.3. Interações entre visualizações

Os *dashboards* incluem interações entre visualizações, permitindo que a seleção de um elemento num gráfico influencie automaticamente os restantes elementos da página. Estas interações facilitam a exploração dos dados e permitem aprofundar a análise sem

necessidade de aplicar filtros adicionais. Na Figura 34 é apresentado um exemplo de interação entre visualizações.



Figura 34. Interação entre visualizações e filtros no painel Produtos

Em algumas páginas, foram também implementadas zonas que permitem alternar entre diferentes tipos de gráficos, oferecendo várias opções de visualização da informação sem ocupar espaço adicional no *dashboard*. Esta abordagem contribui para uma apresentação mais organizada e flexível dos dados. Na Figura 35 é apresentado um exemplo de alternância entre visualizações.

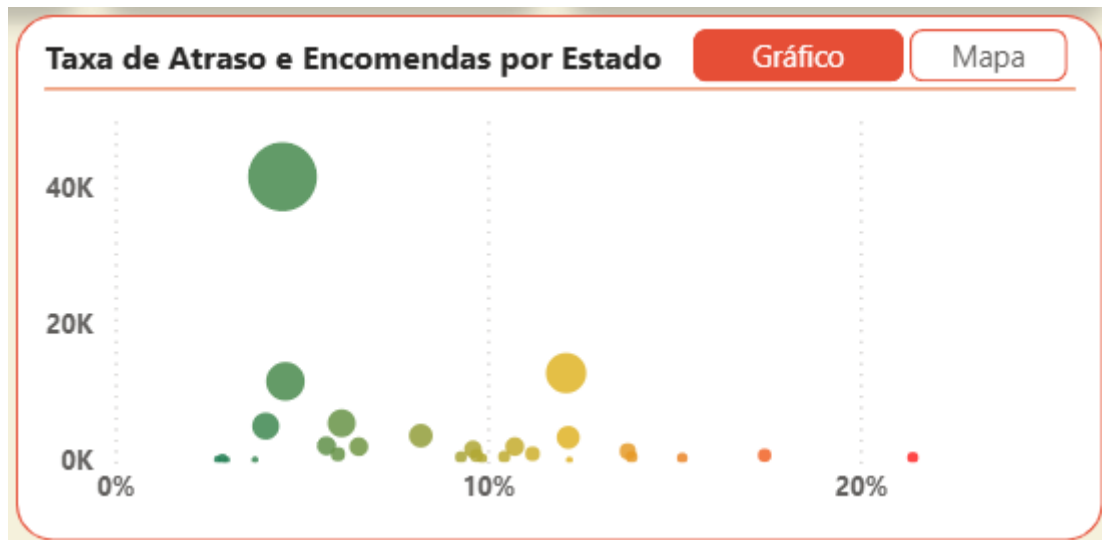


Figura 35. Alternância entre modos de visualização (gráfico e mapa)

### 6.3.4. Controlo Temporal

Nos painéis *Sumário*, *Produtos*, *Logística*, *Vendedores* e *Avaliações* foram incluídos gráficos com botões que permitem analisar os dados segundo diferentes níveis temporais. O utilizador pode optar por visualizar os dados por mês, incluindo ano, trimestre e mês, por trimestre, incluindo trimestre e ano, ou por ano. Esta funcionalidade foi implementada recorrendo à estrutura de hierarquias disponibilizada pelo Power BI. Na Figura 36 é apresentado um exemplo do controlo de granularidade temporal.

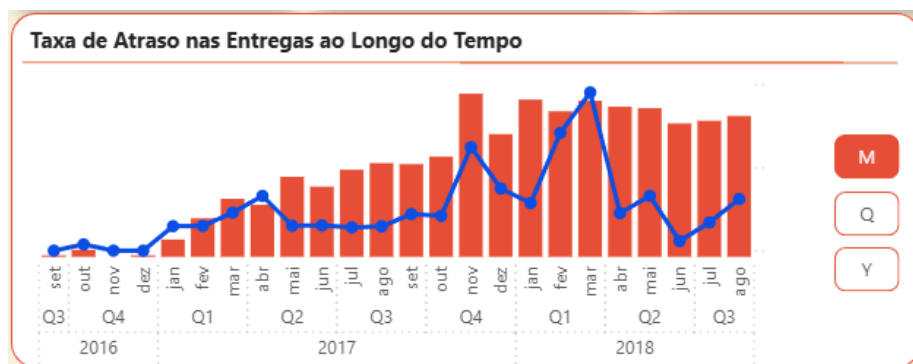


Figura 36. Exemplo de controlo temporal num gráfico

### 6.3.5. Apoio à Exploração da Informação

No painel *Sumário*, no gráfico de lucro total por categoria, foi implementada uma funcionalidade de paginação dinâmica que permite ao utilizador definir quantas categorias pretende visualizar em simultâneo. Esta abordagem facilita a leitura do gráfico quando existe

um número elevado de categorias, evitando a sobrecarga visual. Na Figura 37 é apresentada a paginação dinâmica aplicada ao gráfico de categorias.

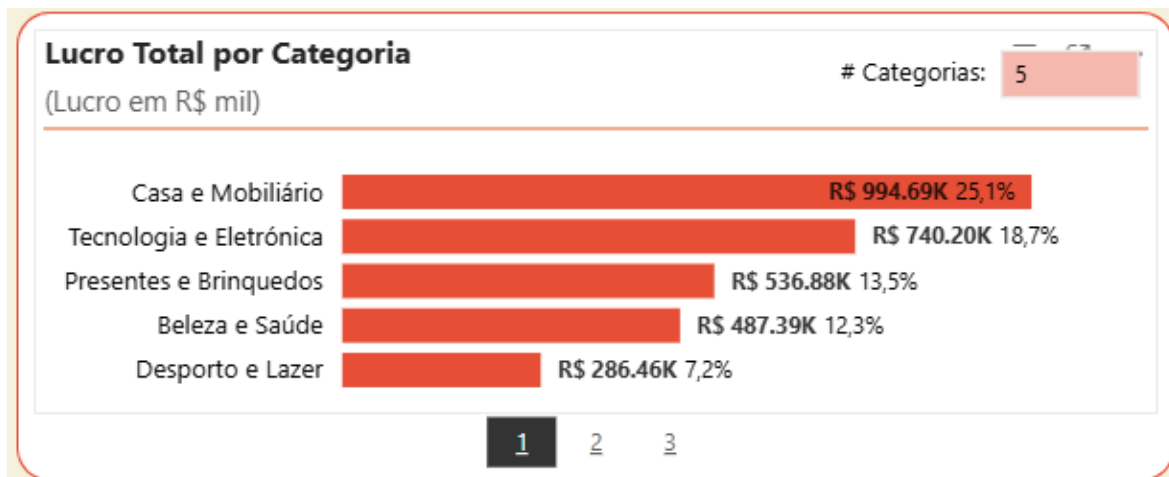


Figura 37. Paginação dinâmica aplicada ao gráfico de lucro por categoria

Por sua vez, no painel *Produtos*, foi incluído um filtro de pesquisa textual que permite localizar rapidamente uma subcategoria específica, solução adotada devido ao elevado número de subcategorias existentes. Esta funcionalidade melhora a usabilidade e reduz o esforço necessário para encontrar informação relevante. A Figura 38 apresenta o filtro textual de subcategorias.

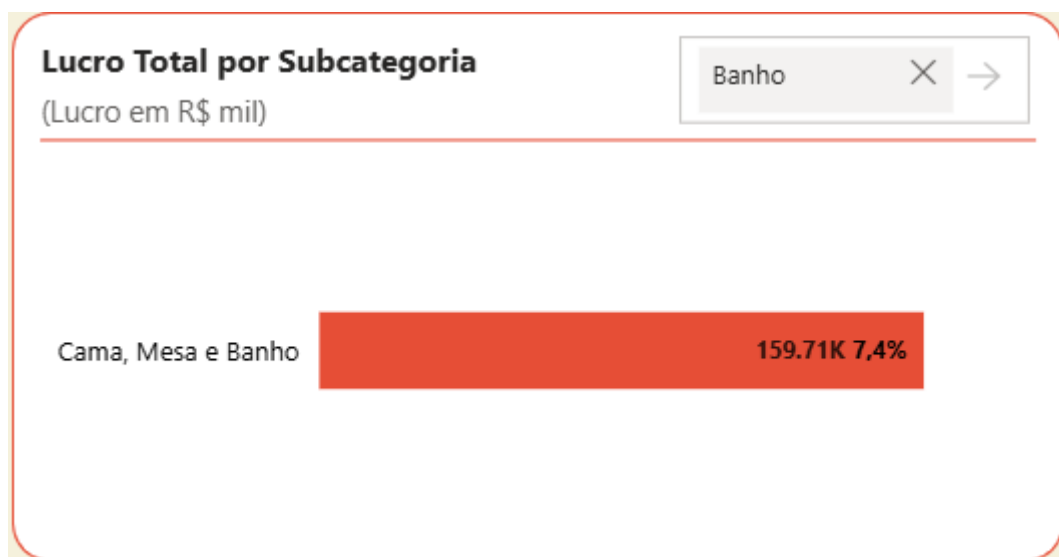


Figura 38. Filtro textual de subcategorias no painel Produtos

### 6.3.6. Tooltips e Informação de Apoio

De forma a enriquecer a análise e fornecer informação adicional sem sobrecarregar as visualizações principais, foram utilizadas *tooltips* nos *dashboards* desenvolvidos. Estas

*tooltips* permitem apresentar informação contextual quando o utilizador interage com os elementos gráficos, apoiando a interpretação dos dados de forma intuitiva.

Em alguns casos, foram implementadas *tooltips* personalizadas, permitindo apresentar indicadores adicionais relevantes para a análise, como variações temporais, valores acumulados e informação detalhada por categoria ou subcategoria. Esta abordagem possibilita aprofundar a análise sem necessidade de recorrer a novos gráficos ou páginas adicionais, fornecendo informação contextual de forma clara e não intrusiva. Na Figura 39 e Figura 40 são apresentadas *tooltips* personalizadas aplicadas a diferentes visualizações dos *dashboards*.

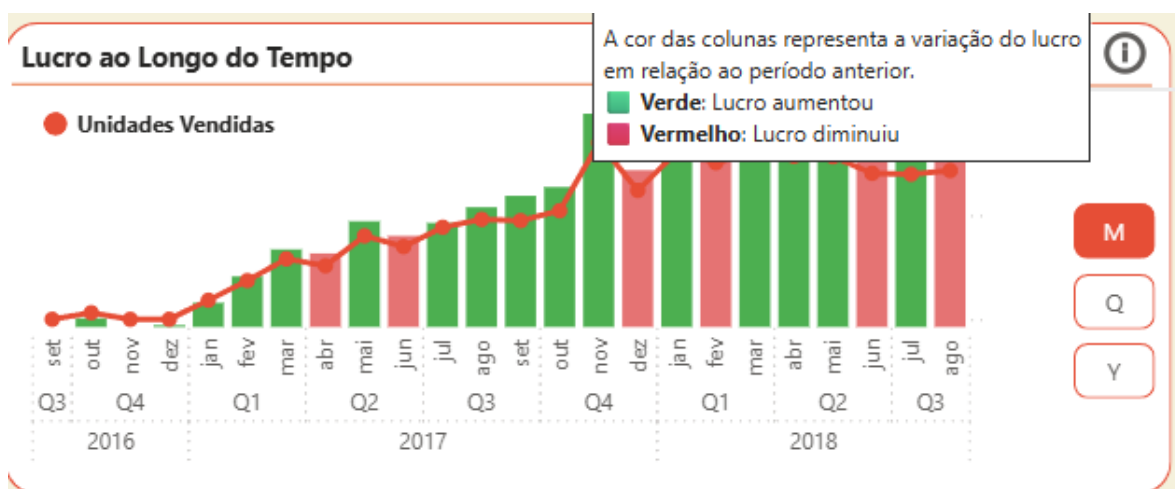


Figura 39. *Tooltip* personalizada com informação adicional de análise temporal

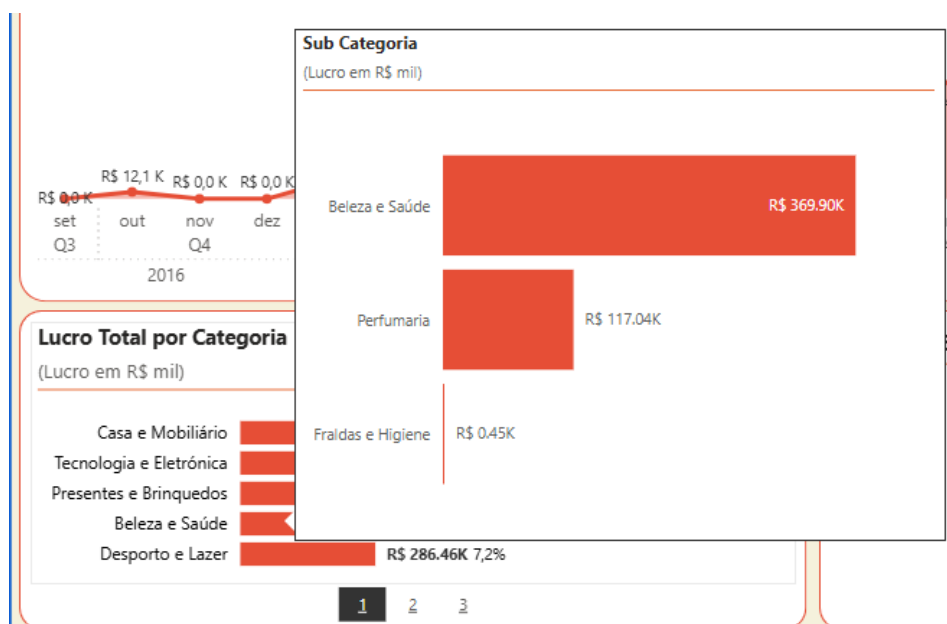


Figura 40. *Tooltip* personalizada com informação detalhada por categoria/subcategoria

Adicionalmente, foram utilizadas *tooltips* padrão do Power BI em visualizações onde a informação adicional apresentada é suficiente para apoiar a leitura dos dados. Estas *tooltips* permitem visualizar valores exatos e percentagens associadas aos elementos selecionados.

Na Figura 41 é apresentada uma *tooltip* padrão utilizada numa visualização da distribuição de encomendas por método de pagamento.

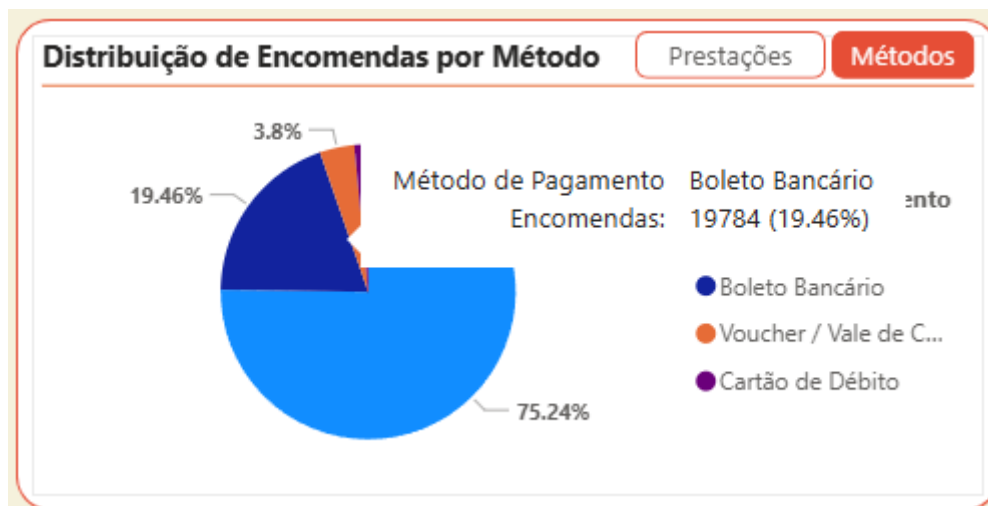


Figura 41. *Tooltip* padrão na distribuição de encomendas por método de pagamento

## 6.4. Conjunto de Medidas e Indicadores

O desenvolvimento dos *dashboards* exigiu a definição de um conjunto abrangente e estruturado de medidas e indicadores, capaz de suportar as diferentes análises apresentadas ao longo das várias páginas do relatório. As métricas foram concebidas de forma consistente, assegurando coerência entre os *dashboards* e permitindo comparações fiáveis entre períodos, categorias e áreas de negócio.

As medidas implementadas incluem indicadores base, como receita, lucro, número de encomendas, unidades vendidas e número de clientes, bem como indicadores derivados que permitem analisar o desempenho ao longo do tempo e avaliar variações entre diferentes períodos. Para este efeito, foram desenvolvidas métricas acumuladas, como *Year-to-Date* (YTD) e *Previous Year-to-Date* (PYTD), bem como métricas de variação percentual, possibilitando a comparação entre períodos homólogos e a identificação de tendências de crescimento ou de quebra de desempenho.

No caso das métricas financeiras, foram aplicadas regras de negócio específicas com o objetivo de garantir o rigor da análise. Em particular, os indicadores de receita, lucro e AOV consideram apenas encomendas efetivamente entregues e pagas, excluindo encomendas canceladas ou não concluídas. Esta decisão assegura que os resultados financeiros apresentados refletem apenas transações finalizadas, contribuindo para uma análise mais realista e fiável do desempenho do negócio.

Para além dos indicadores financeiros, foram também desenvolvidas medidas específicas para a análise do comportamento dos clientes, como o número de clientes recorrentes, a receita média por cliente (*Average Revenue per Customer – ARPC*) e métricas associadas à recorrência de compras. Estes indicadores permitem compreender padrões de fidelização e avaliar o contributo dos clientes para os resultados globais.

No domínio logístico, foram implementadas métricas relacionadas com o desempenho das entregas, incluindo tempo médio de entrega, entregas atrasadas, entregas no prazo e a taxa de cumprimento de encomendas (*On-Time In-Full - OTIF*). Estes indicadores permitem avaliar a eficiência do processo de distribuição e identificar potenciais áreas de melhoria no serviço prestado.

Relativamente à experiência do cliente, foram desenvolvidas métricas associadas às avaliações das encomendas, incluindo a avaliação média, a distribuição das classificações e o número de avaliações com comentários, permitindo analisar não apenas o nível de satisfação dos clientes, mas também a qualidade do feedback fornecido.

De forma a facilitar a utilização e manutenção do modelo, as medidas foram organizadas por grupos funcionais, refletindo as diferentes áreas de análise abordadas nos *dashboards*. Esta organização contribuiu para uma utilização consistente das métricas ao longo do projeto, promovendo a reutilização dos indicadores e assegurando coerência na apresentação e interpretação dos resultados.

## **6.5. Estruturas de Apoio à Análise**

De forma a apoiar a construção de algumas visualizações e facilitar a interpretação dos resultados, foram criadas dimensões auxiliares específicas. Estas estruturas não fazem parte das fontes de dados originais, tendo sido desenvolvidas com o objetivo de organizar a

informação em categorias analíticas mais adequadas ao processo de análise e ao *data storytelling*.

A dimensão *Dim\_Product\_Recurrence* foi utilizada para classificar os produtos de acordo com a recorrência de compras, permitindo agrupar os produtos em diferentes níveis de frequência, como compras únicas ou compras repetidas. Esta dimensão foi aplicada na visualização de recorrência de compras por produto, possibilitando analisar o comportamento dos clientes em termos de repetição de compra e identificar produtos com maior grau de fidelização. Na Figura 42 é apresentada a visualização de recorrência de compras por produto.



Figura 42. Recorrência de compras por produto

Por sua vez, a dimensão *Dim\_Price\_Range* foi criada para agrupar os produtos por intervalos de preço previamente definidos. Esta estrutura permitiu analisar a distribuição dos produtos por faixas de preço, facilitando a comparação entre diferentes níveis de valor e apoiando a análise do posicionamento dos produtos em termos de preço. Na Figura 43 é apresentada a distribuição de produtos por faixa de preço.

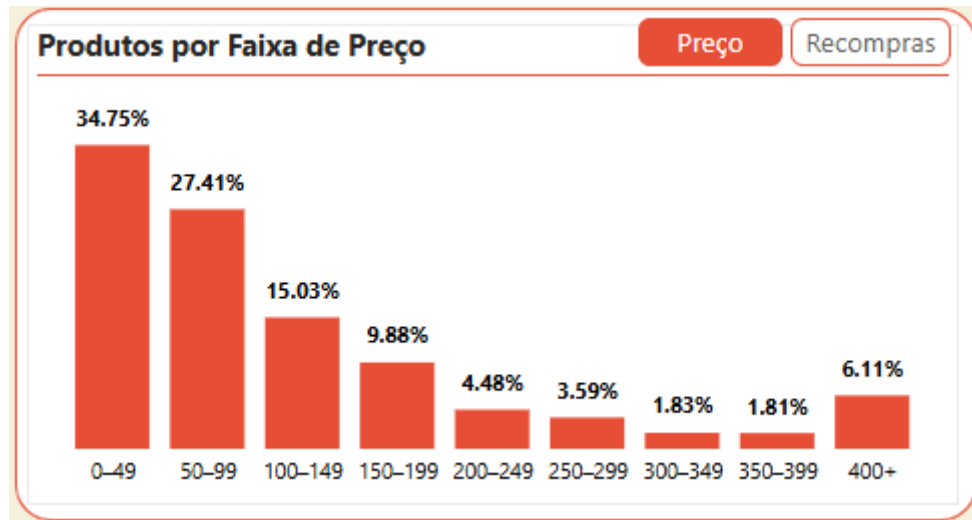


Figura 43. Distribuição de produtos por faixa de preço

## 6.6. Interatividade e Uso de Bookmarks

Conforme descrito nas Secções 6.3.3 e 6.3.4, em vários painéis foram implementadas zonas que permitem alternar entre diferentes formas de visualização da informação e controlar a dimensão temporal da análise. Estas funcionalidades foram implementadas com recurso a *bookmarks* do Power BI, que permitem definir diferentes estados de visualização dentro da mesma área do *dashboard*.

No painel Sumário, ilustrado na Figura 44, os *bookmarks* permitem alternar entre diferentes perspetivas de análise, nomeadamente *Clientes*, *Encomendas*, *Lucro*, *Receita* e *Valor Médio*. Cada opção corresponde a uma vista alternativa da informação, mantendo uma apresentação consistente e facilitando a comparação entre indicadores.

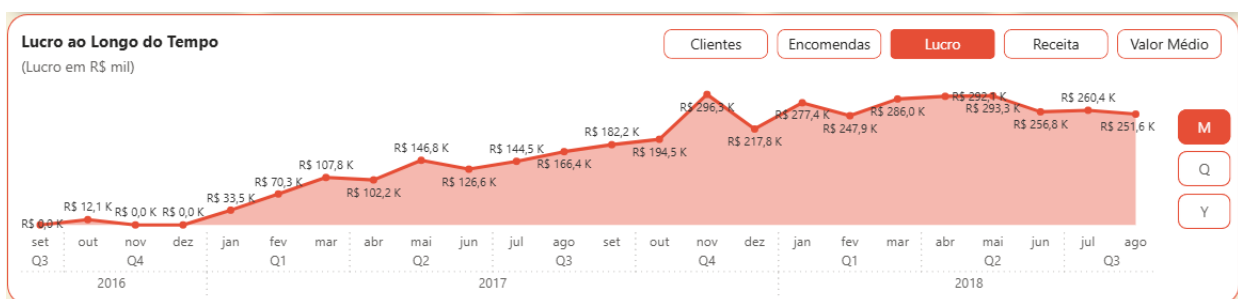


Figura 44. Alternância de perspetivas de análise no gráfico temporal com recurso a *bookmarks*

Adicionalmente, nesta zona do *dashboard* foi integrado o controlo de granularidade temporal, permitindo analisar os dados por mês, trimestre ou ano. A combinação entre alternância de perspetivas e controlo temporal reforça a flexibilidade da análise, permitindo ao utilizador explorar a informação com diferentes níveis de detalhe.

O uso de *bookmarks* contribuiu, assim, para aumentar a interatividade dos *dashboards*, oferecendo múltiplas opções de visualização sem comprometer a organização do *layout*.

## 6.7. Análise de Resultados

Nos subcapítulos seguintes procede-se à exploração dos *insights* previamente definidos na Secção 4.3, utilizando as diferentes páginas dos *dashboards* como suporte analítico. A análise desenvolve-se segundo a narrativa estruturada anteriormente, assegurando alinhamento entre o enquadramento conceptual e a interpretação dos resultados.

A abordagem inicia-se com uma visão global do desempenho do negócio, permitindo enquadrar a evolução do *e-commerce* Olist ao longo do período em estudo e identificar os principais padrões de crescimento que servem de base às análises subsequentes.

### 6.7.1. Visão Global do Negócio

A análise da página *Sumário* permite compreender a evolução global do *e-commerce* Olist ao longo do período em estudo, evidenciando um crescimento consistente do negócio entre 2016 e 2018. Este crescimento manifesta-se de forma transversal nos principais indicadores de desempenho, nomeadamente no número de clientes, no volume de encomendas, na receita gerada e no lucro obtido, refletindo a progressiva consolidação da plataforma no mercado.

O aumento do volume total de encomendas, que ultrapassa as 99 mil, ocorre em paralelo com a expansão da base de clientes, que atinge cerca de 96 mil utilizadores. Esta evolução indica que o crescimento do negócio é impulsionado pela aquisição de novos clientes, sobretudo numa fase inicial mais acelerada entre 2016 e 2017. Em 2018, observa-se um crescimento mais moderado, indicativo de uma transição para uma fase de maior estabilidade e maturidade do *e-commerce*.

A análise temporal revela a existência de períodos de maior intensidade transacional, destacando-se o mês de novembro de 2017 como o momento de maior volume de encomendas. Na Figura 45 é possível observar um pico de receita neste período, correspondente a um crescimento mensal de aproximadamente 53,55% em relação ao mês anterior, refletindo-se diretamente no aumento da atividade transacional e sugerindo a influência de fatores sazonais, como campanhas promocionais e datas comerciais relevantes.

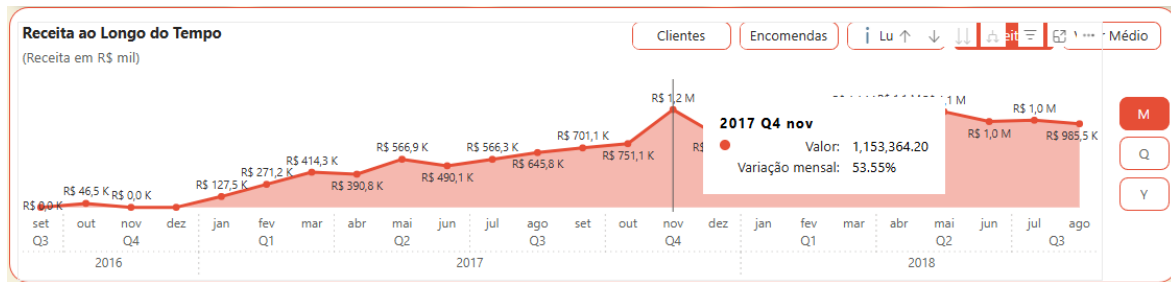


Figura 45. Receita mensal e pico em novembro de 2017

No conjunto do período analisado, a receita acumulada atinge aproximadamente R\$ 15,4 milhões, evidenciando a capacidade do negócio em converter crescimento operacional em desempenho financeiro.

O lucro acompanha diretamente a evolução da receita, totalizando cerca de R\$ 4,0 milhões.

Dado que esta métrica corresponde a uma percentagem fixa do valor da encomenda, a sua variação acompanha o comportamento da receita ao longo do tempo, não permitindo, por si só, identificar alterações na eficiência operacional ou na estrutura de custos do negócio.

Por fim, o valor médio por encomenda mantém-se relativamente estável ao longo de todo o período, situando-se em torno dos R\$ 160. A estabilidade deste indicador sugere que o crescimento da receita resulta principalmente do aumento do volume de encomendas, e não de variações no valor médio das transações.

Em síntese, a análise da página *Sumário* evidencia que o crescimento do *e-commerce* Olist ao longo do período em estudo é essencialmente suportado pelo aumento do volume de encomendas e da base de clientes, com um papel relevante da sazonalidade em períodos específicos. Este enquadramento global constitui o ponto de partida para as análises subsequentes, permitindo interpretar de forma integrada os resultados observados nas restantes dimensões do negócio.

### 6.7.2. Desempenho dos Produtos

A análise da página *Produtos* permite aprofundar o desempenho do portefólio do *e-commerce*, identificando padrões de venda, concentração de lucro, recorrência de compras e distribuição por faixas de preço. No período em análise, foram vendidas 110 194 unidades, evidenciando a dimensão e diversidade do catálogo de produtos disponibilizado na plataforma.

A evolução temporal das vendas mostra a existência de períodos de maior intensidade comercial, destacando-se o mês de novembro de 2017 como o momento com maior volume de vendas, registrando 8 475 unidades vendidas. Conforme evidenciado na Figura 46, este pico de vendas coincide com o período de maior receita, reforçando a existência de um efeito de sazonalidade associado a campanhas promocionais e datas comerciais relevantes. Este comportamento confirma a importância estratégica deste período para a maximização do desempenho comercial do negócio.

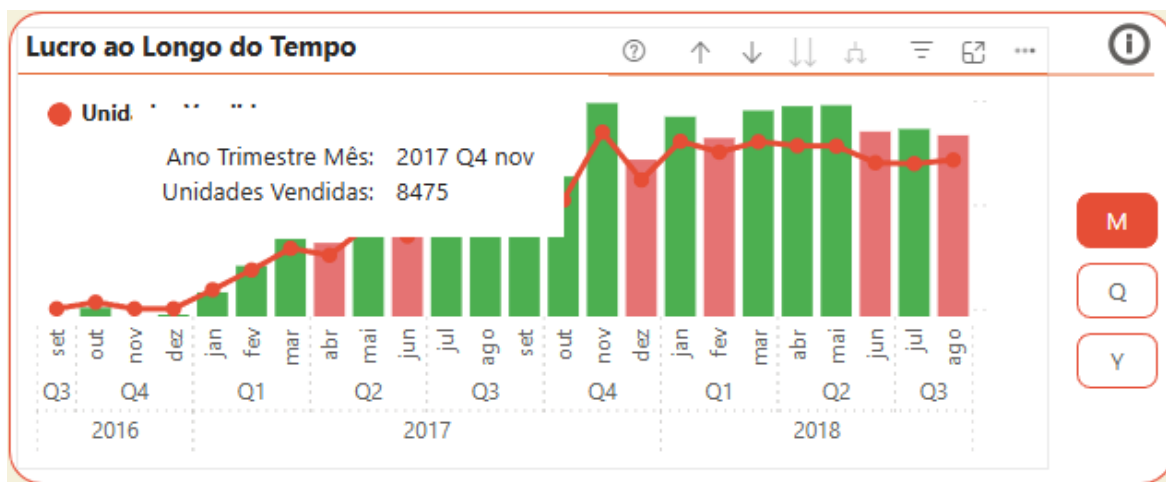


Figura 46. Evolução mensal das unidades vendidas, com pico em novembro de 2017

Relativamente à distribuição do lucro por categoria, observa-se uma clara concentração dos resultados financeiros em algumas áreas específicas do portfólio. Como ilustrado na Figura 47, a categoria Casa e Mobiliário destaca-se como a categoria com maior contributo para o lucro total, com aproximadamente R\$ 994,7 mil, correspondendo a 25,1% do lucro total, seguida por Tecnologia e Eletrónica (R\$ 740,2 mil - 18,7%) e Presentes e Brinquedos (R\$ 536,9 mil - 13,5%). Esta concentração mostra que uma parte reduzida do catálogo é responsável por uma parte significativa dos resultados financeiros, reforçando a importância de gerir estrategicamente a oferta de produtos.

Ao nível das subcategorias, verifica-se que Casa, Mesa e Banho apresenta o maior volume de vendas, com 10 953 unidades vendidas, evidenciando uma elevada procura e uma contribuição relevante para o desempenho global da plataforma. Esta informação, detalhada na tabela de apoio da página, sugere oportunidades de reforço da oferta, otimização de stock e maior destaque comercial desta subcategoria.

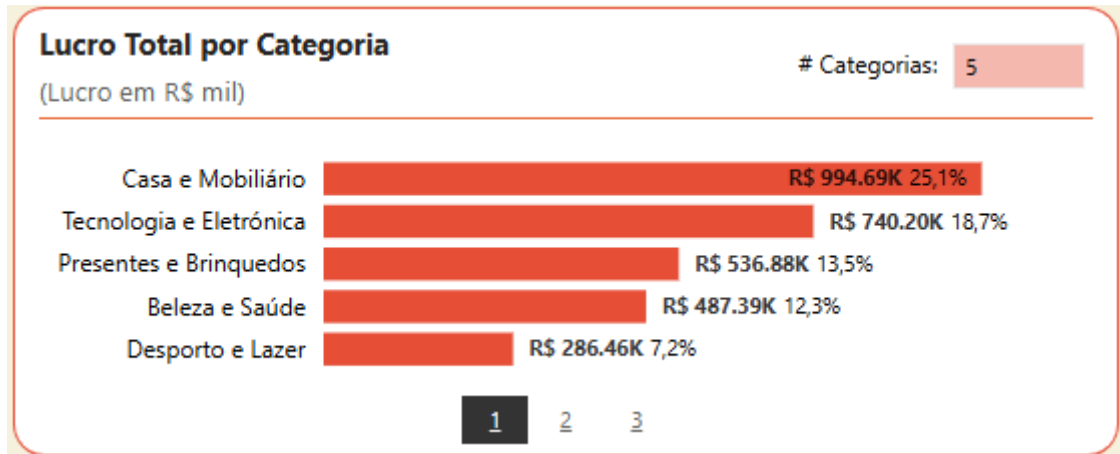


Figura 47. Distribuição do lucro total por categoria de produto

No que diz respeito à recorrência de compras, a Figura 48 evidencia que a maioria dos produtos é adquirida apenas uma única vez, representando cerca de 59,4% do total. Aproximadamente 30,6% dos produtos são comprados entre 2 e 5 vezes, enquanto as compras com elevada recorrência representam uma percentagem residual. Este padrão indica que, apesar do elevado volume de vendas, existe margem para o desenvolvimento de estratégias de fidelização e recompra, especialmente em categorias com maior potencial de repetição de compra.

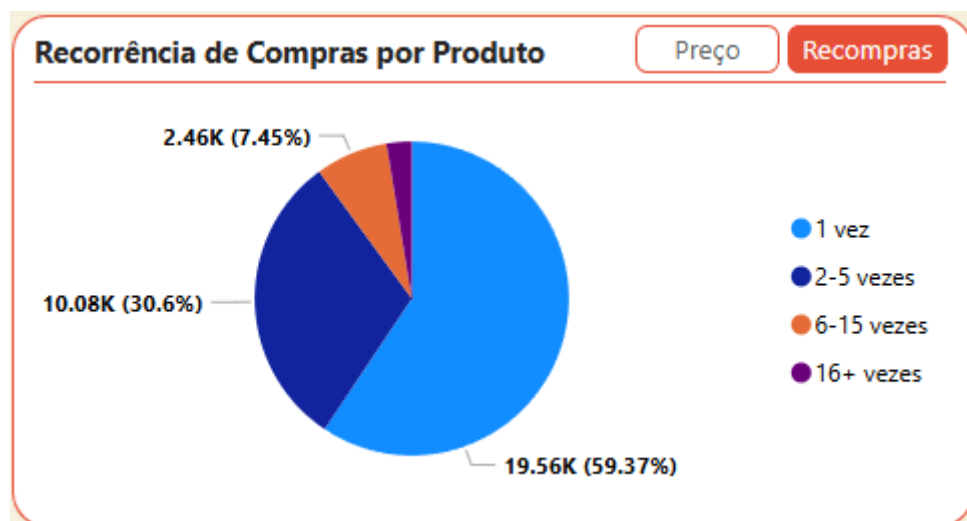


Figura 48. Recorrência de compras por produto

Por fim, a análise da distribuição das vendas por faixa de preço, apresentada na Figura 49, revela que a maior parte das unidades vendidas se concentra nos intervalos de 0–49 e 50–99, que em conjunto representam mais de 60% das vendas. Este comportamento sugere uma preferência por produtos de menor valor unitário, indicando que o crescimento do volume de vendas é impulsionado por artigos de preço acessível. As faixas de preço mais elevadas, embora com menor peso relativo, podem representar oportunidades para estratégias de *upselling*<sup>4</sup> e diversificação do portfólio.

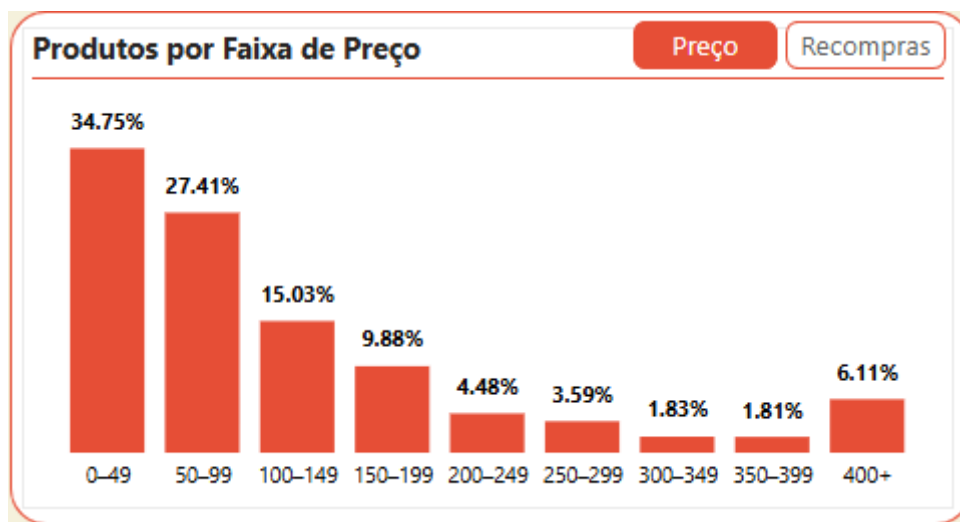


Figura 49. Distribuição das unidades vendidas por faixa de preço

Em síntese, a análise do desempenho dos produtos evidencia que o resultado económico do *e-commerce* é influenciado por um conjunto limitado de categorias e subcategorias, enquanto o crescimento do volume de vendas é impulsionado sobretudo por produtos de baixo valor unitário. Paralelamente, os baixos níveis de recorrência de compra sugerem que o potencial do portfólio não está totalmente explorado ao nível da fidelização.

Estes resultados indicam oportunidades para uma gestão mais estratégica da oferta de produtos, combinando o reforço das categorias mais rentáveis com ações orientadas à recompra e ao aumento do valor médio por cliente.

### 6.7.3. Desempenho Logístico e Cumprimento das Entregas

A análise da página *Logística* permite avaliar de que forma o crescimento do *e-commerce* Olist se reflete no desempenho operacional, em particular no cumprimento dos prazos de

<sup>4</sup> *Upselling*: estratégia de promoção de produtos de maior valor ao cliente no momento da compra.

entrega, na eficiência do processo logístico e nas assimetrias temporais e geográficas associadas à distribuição das encomendas.

De forma global, os indicadores evidenciam um desempenho logístico satisfatório ao longo do período em análise. O tempo médio de entrega situa-se em torno dos 12 dias, enquanto a percentagem de entregas realizadas dentro do prazo ultrapassa os 92%, refletindo um elevado nível de cumprimento operacional. Ainda assim, cerca de 5% a 7% das encomendas registam atrasos, o que, considerando o volume total de pedidos, representa um impacto relevante à escala do negócio.

A análise temporal da taxa de atraso nas entregas, apresentada na Figura 50, revela variações ao longo do tempo, com picos mais acentuados em determinados períodos. Estes aumentos coincidem, em geral, com meses de maior volume de encomendas, sugerindo que o crescimento da procura exerce pressão adicional sobre a capacidade logística. Este comportamento é particularmente evidente em períodos de elevada atividade comercial, reforçando a relação entre sazonalidade, volume transacional e desempenho operacional.

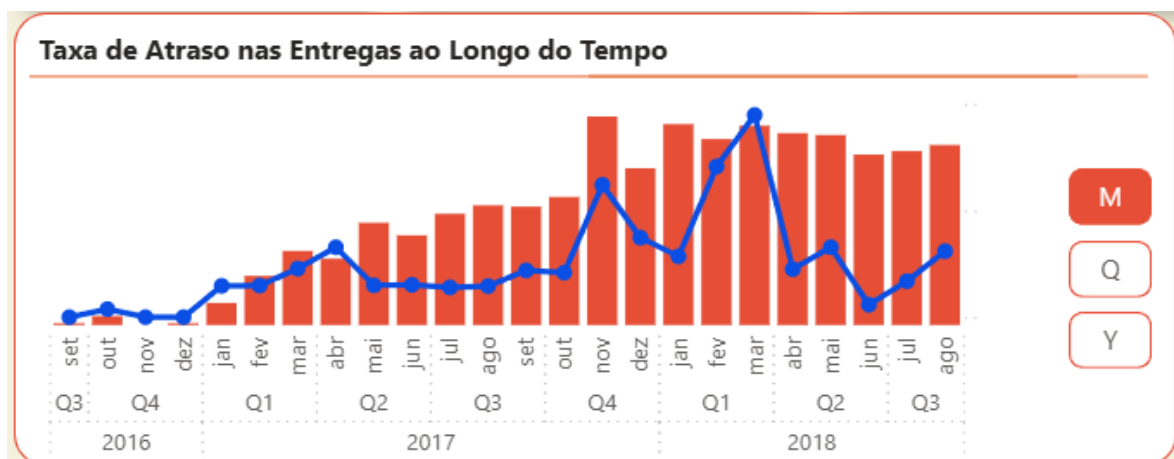


Figura 50. Evolução temporal da taxa de atraso nas entregas

Do ponto de vista geográfico, a análise da taxa de atraso e do número de encomendas por estado, ilustrada na Figura 51, evidencia uma forte heterogeneidade regional. Estados com elevado volume de encomendas, como São Paulo, apresentam taxas de atraso relativamente reduzidas, enquanto outros estados, com menor volume de pedidos, registam percentagens de atraso superiores, como é o caso de Alagoas. Este padrão indica que o desempenho logístico não depende exclusivamente do volume de encomendas, mas também de fatores estruturais como infraestrutura e distância aos centros de distribuição.



Figura 51. Taxa de atraso nas entregas por estado, com destaque para São Paulo e Alagoas

A análise do desempenho logístico por categoria, apresentada na Figura 52, evidencia um comportamento globalmente consistente, embora com algumas diferenças pontuais. Categorias como Casa e Mobiliário e Tecnologia e Eletrônica, que apresentam elevados volumes de encomendas, registam taxas de atraso em torno dos 6,6% a 6,8% e tempos médios de entrega próximos dos 12,4 dias, valores alinhados com a média global. Por outro lado, categorias como Alimentação e Bebidas apresentam tempos médios de entrega ligeiramente inferiores, próximos dos 10 dias, refletindo características operacionais distintas ao longo do processo de entrega.

Estes resultados sugerem que, apesar das diferenças entre categorias, o desempenho logístico mantém-se estável, sendo as variações observadas explicadas sobretudo pelas características operacionais dos produtos e pelo volume de encomendas associado a cada categoria.

| Categoria                | Total de Encomendas | Entregas Atrasadas | Entregas Atrasadas % | PYTD         | YTD vs PYTD  | Tempo Máximo (dias) | Tempo Médio (dias) |
|--------------------------|---------------------|--------------------|----------------------|--------------|--------------|---------------------|--------------------|
| Alimentação e Bebidas    | 1169                | 61                 | 5,22%                | 0,00%        | 0,00%        | 24,10               | 9,93               |
| Beleza e Saúde           | 12013               | 851                | 7,08%                | 0,00%        | 0,00%        | 29,00               | 11,57              |
| Casa e Mobiliário        | 26248               | 1745               | 6,65%                | 0,00%        | 0,00%        | 29,00               | 12,44              |
| Construção e Ferramentas | 4960                | 315                | 6,35%                | 0,00%        | 0,00%        | 28,00               | 12,40              |
| Crianças e Bebês         | 6752                | 468                | 6,93%                | 0,00%        | 0,00%        | 29,00               | 11,68              |
| Desporto e Lazer         | 7720                | 495                | 6,41%                | 0,00%        | 0,00%        | 28,00               | 11,78              |
| Escritório e Papelaria   | 2451                | 154                | 6,28%                | 0,00%        | 0,00%        | 30,00               | 12,21              |
| Livros, Música e Filmes  | 1851                | 124                | 6,70%                | 0,00%        | 0,00%        | 27,00               | 11,28              |
| Moda                     | 3441                | 176                | 5,11%                | 0,00%        | 0,00%        | 27,00               | 11,12              |
| Outros                   | 3701                | 212                | 5,73%                | 0,00%        | 0,00%        | 27,00               | 11,46              |
| Presentes e Brinquedos   | 9421                | 626                | 6,64%                | 0,00%        | 0,00%        | 29,00               | 12,19              |
| Tecnologia e Eletrônica  | 19458               | 1315               | 6,76%                | 0,00%        | 0,00%        | 29,00               | 12,48              |
| <b>Total</b>             | <b>99441</b>        | <b>6534</b>        | <b>6,57%</b>         | <b>0,00%</b> | <b>0,00%</b> | <b>29,00</b>        | <b>12,09</b>       |

Figura 52. Taxa de atraso e tempo médio de entrega por categoria

Em síntese, os resultados mostram que o desempenho logístico do *e-commerce* Olist é globalmente sólido, mas sensível a períodos de maior procura e a limitações regionais. Estes fatores evidenciam a logística como um elemento crítico para a sustentabilidade do crescimento do negócio e para a manutenção de elevados níveis de satisfação do cliente.

#### **6.7.4. Análise do Comportamento dos Clientes**

A página *Clientes* permite uma leitura integrada do comportamento dos consumidores do *e-commerce* Olist, combinando indicadores de volume, recorrência, perfil demográfico, padrões de compra e métricas de valor. Esta análise contribui para compreender não apenas quem compra, mas também como e com que frequência o faz, enriquecendo a interpretação do desempenho global do negócio.

O volume total de clientes atinge aproximadamente 96 mil, evidenciando uma base de utilizadores significativa e alinhada com o crescimento observado no número de encomendas. No entanto, a taxa de clientes recorrentes, situada em cerca de 3,1%, revela que a maioria dos clientes realiza apenas uma compra ao longo do período analisado. Este padrão indica um modelo de negócio suportado na aquisição de novos clientes, mas com baixo nível de fidelização, sugerindo potencial para o desenvolvimento de estratégias orientadas à retenção e recompra.

A análise da receita média por cliente, que se situa em torno dos R\$ 165, demonstra uma estabilidade ao longo do tempo, mesmo perante variações no volume de clientes e encomendas. Este comportamento sugere que o crescimento da receita é maioritariamente impulsionado pelo aumento da base de clientes, e não por alterações no valor médio gasto por cada consumidor.

A introdução da faixa etária como atributo no processo de transformação dos dados permite aprofundar a análise do perfil dos clientes. Conforme evidenciado na Figura 53, as faixas etárias entre os 26–35 e 36–45 anos concentram a maior proporção de clientes, representando também os segmentos com maior contribuição para a receita média. Por outro lado, faixas etárias mais elevadas apresentam menor peso relativo, tanto em número de clientes como

em valor gerado, o que reforça a importância dos segmentos etários intermédios para o desempenho do negócio.

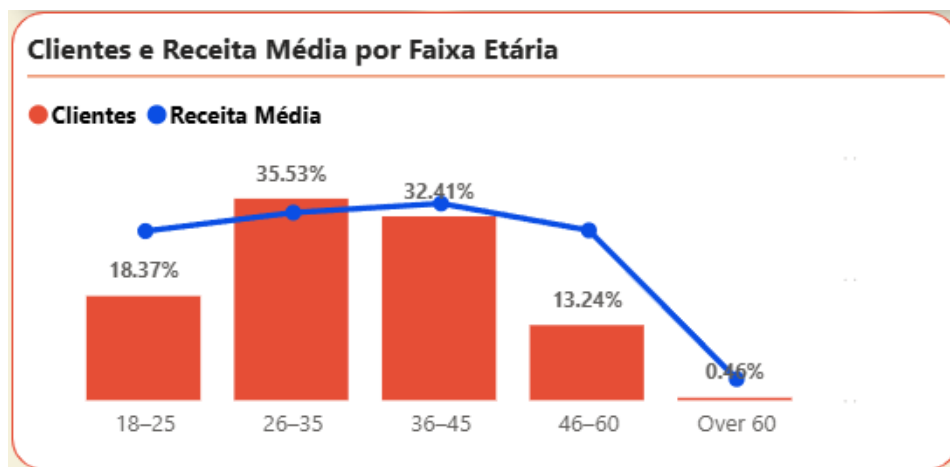


Figura 53. Perfil etário dos clientes e respetiva receita média

No que respeita aos padrões temporais de compra, a análise da média de encomendas por dia do mês evidencia a existência de picos em dias específicos, com maior concentração na segunda quinzena do mês, conforme ilustrado na Figura 54. Este comportamento pode estar associado a fatores como ciclos de rendimento, campanhas promocionais ou datas comerciais, fornecendo informação relevante para o planeamento de ações comerciais e operacionais.

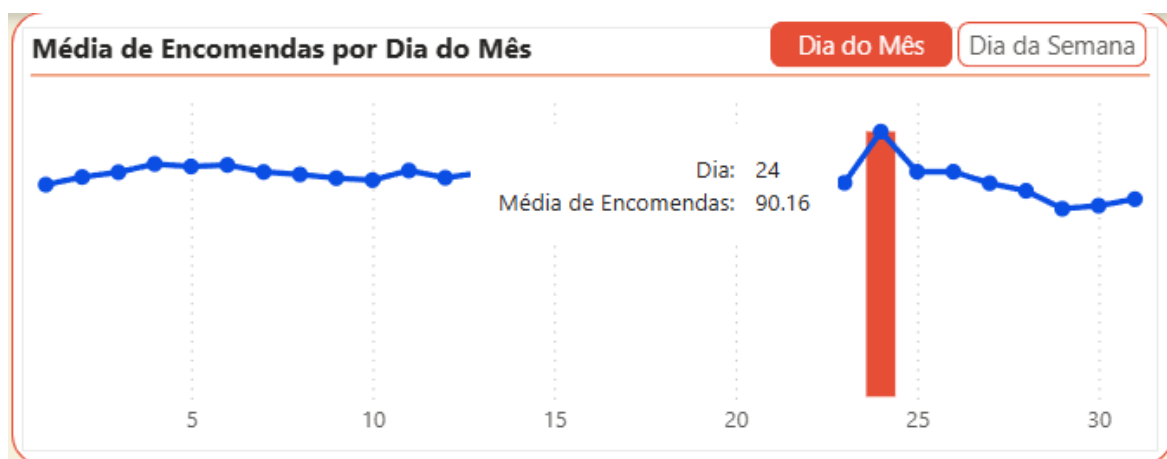


Figura 54. Distribuição da média de encomendas por dia do mês

Complementarmente, a análise das compras por dia da semana e intervalo horário, apresentada na Figura 55, revela que a atividade transacional se concentra predominantemente nos períodos diurnos e de fim de tarde, em particular entre as 13h e as 18h, ao longo de todos os dias da semana. Observa-se ainda uma maior intensidade de

compras nos dias úteis, sobretudo entre segunda e sexta-feira, enquanto os fins de semana apresentam volumes inferiores, ainda que com maior dispersão horária.

| Compras por Dia da Semana e Horário |           |           |           |           |
|-------------------------------------|-----------|-----------|-----------|-----------|
| Dia do Mês                          |           |           |           |           |
| Dia da Semana                       |           |           |           |           |
| Dia da Semana                       | 00h - 05h | 06h - 12h | 13h - 18h | 19h - 23h |
| Segunda-feira                       | 607       | 4575      | 6216      | 4798      |
| Terça-feira                         | 625       | 4601      | 6141      | 4596      |
| Quarta-feira                        | 750       | 4673      | 5921      | 4208      |
| Quinta-feira                        | 695       | 4492      | 5661      | 3913      |
| Sexta-feira                         | 839       | 4365      | 5456      | 3462      |
| Sábado                              | 649       | 2906      | 4213      | 3119      |
| Domingo                             | 575       | 2623      | 4527      | 4235      |

Figura 55. Volume de encomendas por dia da semana e período do dia

Relativamente aos métodos de pagamento, o cartão de crédito destaca-se como o meio de pagamento mais utilizado, representando cerca de 75% das encomendas, conforme apresentado na Figura 56. A distribuição do número de prestações demonstra a preferência dos clientes por pagamentos fracionados de curta duração, evidenciando a importância das condições de pagamento na decisão de compra.

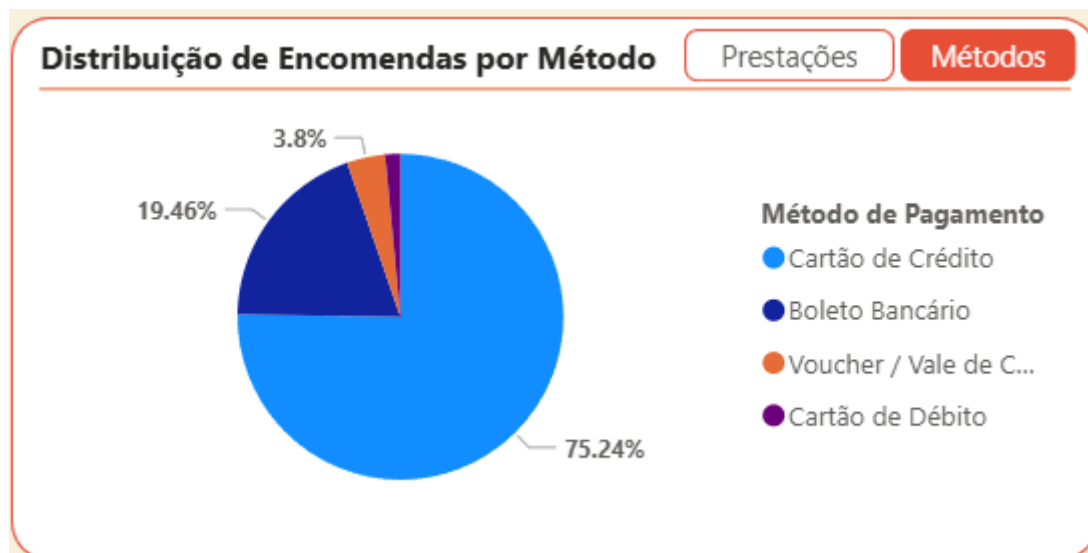


Figura 56. Participação dos métodos de pagamento no total de encomendas

Por fim, a taxa de cancelamento, que se situa em torno de 0,6%, apresenta valores reduzidos e relativamente homogêneos entre categorias de produto. Apesar de algumas variações pontuais, não se identificam categorias com níveis de cancelamento suficientemente

elevados para comprometer o desempenho global, sugerindo um processo de venda e entrega globalmente robusto do ponto de vista do cliente.

Em síntese, a análise do comportamento dos clientes revela que o crescimento do *e-commerce* Olist é impulsionado sobretudo pelo aumento da base de clientes, mantendo níveis reduzidos de fidelização. Os padrões temporais de compra, o perfil etário dominante e a preferência por pagamentos fracionados evidenciam oportunidades para estratégias mais direcionadas, orientadas à retenção, à recorrência de compra e ao aumento do valor gerado por cliente.

#### **6.7.5. Análise dos Vendedores**

A página dedicada aos *Vendedores* permite analisar como o desempenho económico do *marketplace* da Olist se distribui pelos diferentes vendedores, evidenciando padrões de concentração e heterogeneidade no contributo para os resultados globais. Apesar do elevado número de vendedores ativos, responsáveis por um volume significativo de encomendas e de lucro, o desempenho económico não se encontra distribuído de forma uniforme.

A análise do *Top 15 Empresas por Lucro*, apresentada na Figura 57, evidencia uma clara concentração do desempenho económico num número reduzido de vendedores. O conjunto das 15 empresas com maior lucro totaliza aproximadamente R\$ 1,64 milhões, o que representa cerca de 18% do lucro total do *marketplace*, estimado em R\$ 9,25 milhões, apesar de a plataforma contar com mais de 3 000 vendedores ativos. Destaca-se a Empresa 2618 como o vendedor com maior contributo individual para o lucro. Este padrão mostra uma distribuição desigual do desempenho entre os vendedores, sendo que uma minoria concentra uma parte relevante dos resultados globais.

| Empresa      | Lucro       | Lucro YTD | Lucro PYTD | YTD vs PYTD | Quota de Mercado YTD | Quota de Mercado PYTD | Crescimento de Quota Mercado |
|--------------|-------------|-----------|------------|-------------|----------------------|-----------------------|------------------------------|
| Empresa 2618 | R\$ 158.89K | R\$ 0.00  | R\$ 0.00   | 0.0% →      | 0.00%                | 0.00%                 | 0.00% →                      |
| Empresa 902  | R\$ 152.56K | R\$ 0.00  | R\$ 0.00   | 0.0% →      | 0.00%                | 0.00%                 | 0.00% →                      |
| Empresa 2464 | R\$ 137.82K | R\$ 0.00  | R\$ 0.00   | 0.0% →      | 0.00%                | 0.00%                 | 0.00% →                      |
| Empresa 558  | R\$ 133.64K | R\$ 0.00  | R\$ 0.00   | 0.0% →      | 0.00%                | 0.00%                 | 0.00% →                      |
| Empresa 1183 | R\$ 130.60K | R\$ 0.00  | R\$ 0.00   | 0.0% →      | 0.00%                | 0.00%                 | 0.00% →                      |
| Empresa 69   | R\$ 116.19K | R\$ 0.00  | R\$ 0.00   | 0.0% →      | 0.00%                | 0.00%                 | 0.00% →                      |
| Empresa 1874 | R\$ 111.87K | R\$ 0.00  | R\$ 0.00   | 0.0% →      | 0.00%                | 0.00%                 | 0.00% →                      |
| Empresa 2208 | R\$ 97.76K  | R\$ 0.00  | R\$ 0.00   | 0.0% →      | 0.00%                | 0.00%                 | 0.00% →                      |
| Empresa 2721 | R\$ 96.75K  | R\$ 0.00  | R\$ 0.00   | 0.0% →      | 0.00%                | 0.00%                 | 0.00% →                      |
| Empresa 391  | R\$ 92.29K  | R\$ 0.00  | R\$ 0.00   | 0.0% →      | 0.00%                | 0.00%                 | 0.00% →                      |
| Empresa 2956 | R\$ 85.97K  | R\$ 0.00  | R\$ 0.00   | 0.0% →      | 0.00%                | 0.00%                 | 0.00% →                      |
| Empresa 798  | R\$ 84.49K  | R\$ 0.00  | R\$ 0.00   | 0.0% →      | 0.00%                | 0.00%                 | 0.00% →                      |
| Empresa 1894 | R\$ 78.72K  | R\$ 0.00  | R\$ 0.00   | 0.0% →      | 0.00%                | 0.00%                 | 0.00% →                      |
| Empresa 243  | R\$ 78.71K  | R\$ 0.00  | R\$ 0.00   | 0.0% →      | 0.00%                | 0.00%                 | 0.00% →                      |
| Empresa 2341 | R\$ 77.79K  | R\$ 0.00  | R\$ 0.00   | 0.0% →      | 0.00%                | 0.00%                 | 0.00% →                      |

Figura 57. Top 15 empresas por lucro

A evolução temporal do lucro dos principais vendedores, apresentada na Figura 58, revela um comportamento heterogéneo ao longo do tempo, mesmo entre as empresas com maior contributo global para o lucro. Observam-se oscilações relevantes e picos concentrados em períodos específicos para diferentes vendedores, indicando que o desempenho não é constante nem distribuído de forma homogénea ao longo do ano. Alguns vendedores evidenciam aumentos acentuados do lucro em momentos pontuais, seguidos de períodos com valores significativamente inferiores ou nulos. Como exemplo, destaca-se a Empresa 902, representada a rosa, que apresenta um pico expressivo em setembro de 2017, enquanto noutros meses o seu contributo é reduzido. Este padrão repete-se, ainda que em momentos distintos, para outros vendedores, sugerindo a existência de eventos sazonais, contratos pontuais ou operações extraordinárias que impactam significativamente os resultados mensais.

No que respeita à distribuição dos vendedores por categoria, observa-se que Casa e Mobiliário é a categoria com o maior número de vendedores ativos, concentrando 1080 vendedores. Este resultado indica uma elevada atratividade desta categoria dentro do *marketplace*, refletindo uma maior oferta de vendedores associada a este tipo de produto.

Em conjunto, estes resultados evidenciam que o desempenho do *marketplace* assenta numa estrutura assimétrica, marcada pela concentração do lucro em poucos vendedores e por uma elevada variabilidade temporal do desempenho individual. Este padrão sugere a importância de estratégias orientadas para a diversificação do desempenho económico, a redução da

dependência de um número limitado de vendedores e o reforço do equilíbrio do ecossistema da plataforma.

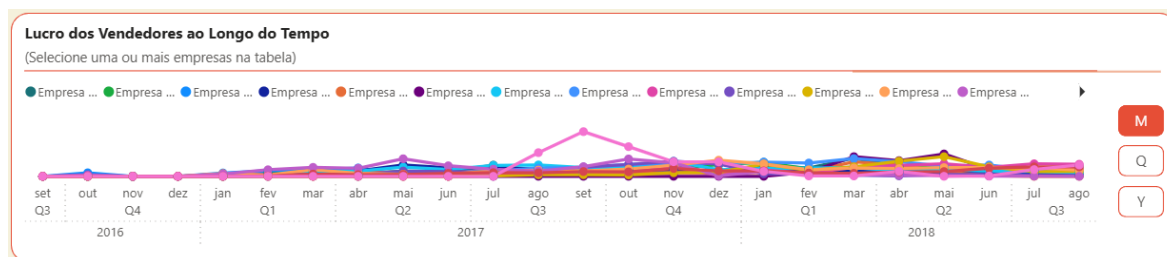


Figura 58. Evolução temporal do lucro dos principais vendedores

### 6.7.6. Avaliações dos Clientes e Qualidade do Serviço

A página de *Avaliações* permite analisar a percepção dos clientes relativamente à sua experiência de compra, integrando informação sobre classificações atribuídas, evolução temporal da satisfação, impacto dos atrasos nas entregas e diferenças entre categorias de produto.

De forma global, observa-se uma evolução positiva da avaliação média ao longo do período analisado. Em 2016, a avaliação média situa-se em torno de 3,53, evidenciando um nível de satisfação mais moderado. A partir de 2017, observa-se uma melhoria, com a avaliação média a subir para cerca de 4,09, valor que se mantém em 2018. Apesar de pequenas variações, a tendência nos anos mais recentes é estável, indicando uma melhoria da experiência de compra e maior consistência na qualidade do serviço.

A análise das avaliações por encomenda mostra um predomínio de classificações elevadas, sobretudo de 4 e 5 estrelas, que representam a maioria das respostas. Em contraste, as avaliações de 1 e 2 estrelas têm pouca expressão, reforçando a percepção globalmente positiva dos clientes em relação à plataforma.

A relação entre o cumprimento dos prazos de entrega e a avaliação atribuída pelo cliente evidencia um padrão consistente, conforme ilustrado na Figura 59. Observa-se que as encomendas entregues dentro do prazo estão fortemente associadas a avaliações mais elevadas, enquanto as entregas atrasadas concentram uma maior proporção de avaliações baixas. Este comportamento confirma o papel do desempenho logístico na satisfação do cliente e na percepção global da qualidade do serviço.

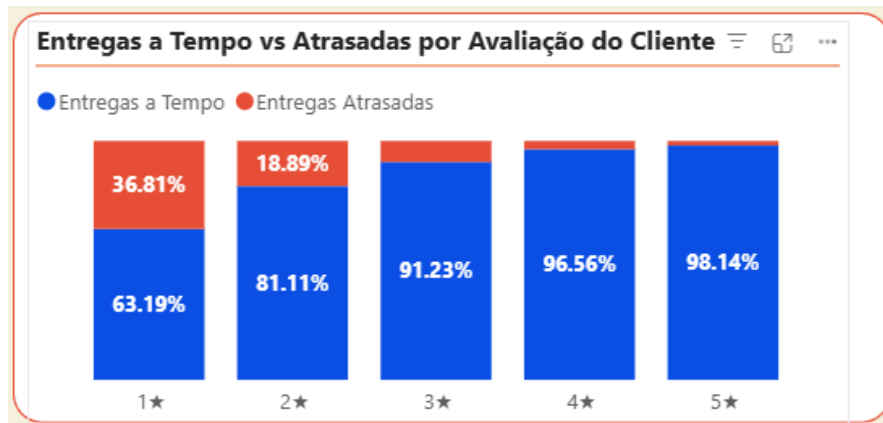


Figura 59. Relação entre o cumprimento do prazo de entrega e a avaliação atribuída pelo cliente

Adicionalmente, a análise por subcategoria de produto revela diferenças relevantes tanto na percentagem de avaliações com comentário textual como na avaliação média atribuída. Algumas subcategorias apresentam níveis de satisfação superiores e maior propensão para feedback detalhado, enquanto outras evidenciam avaliações médias mais baixas, sugerindo oportunidades de melhoria específicas associadas ao produto ou ao serviço prestado.

Em síntese, os resultados desta análise evidenciam que a satisfação do cliente no *e-commerce* Olist está fortemente associada à fiabilidade do serviço logístico e à consistência da experiência de compra. A monitorização sistemática das avaliações permite não apenas acompanhar a percepção dos clientes, mas também identificar pontos críticos de melhoria, reforçando o papel das avaliações como um indicador estratégico da qualidade do serviço.

### 6.7.7. Síntese da Análise Global

A análise integrada dos *dashboards* desenvolvidos permite construir uma leitura coerente e articulada do desempenho do *e-commerce* Olist, evidenciando a forma como crescimento comercial, eficiência operacional, comportamento dos clientes e qualidade do serviço se influenciam mutuamente ao longo do período em estudo.

De forma global, os resultados mostram que o crescimento do negócio entre 2016 e 2018 é motivado sobretudo pelo aumento do volume de encomendas e pela expansão da base de clientes, sendo fortemente condicionado por efeitos sazonais, em particular em períodos de elevada intensidade comercial. Este padrão indica um modelo de crescimento extensivo, assente no aumento da procura, mais do que na variação do valor médio das transações.

A análise do desempenho dos produtos mostra que o resultado económico está concentrado num número limitado de categorias e subcategorias, indicando que uma parte reduzida da oferta gera grande parte do lucro total. Em paralelo, os baixos níveis de recorrência de compra sugerem que o potencial económico do portefólio não se encontra totalmente explorado, apontando para oportunidades ao nível da fidelização e do estímulo à recompra.

Do ponto de vista operacional, o desempenho logístico revela-se globalmente consistente, mas sensível a picos de procura e a assimetrias regionais. Os períodos de maior volume de encomendas exercem pressão adicional sobre a capacidade logística, originando aumentos pontuais nos atrasos de entrega. Esta relação evidencia a logística como um elemento crítico não apenas para a eficiência operacional, mas também para a sustentabilidade do crescimento do negócio.

A análise do comportamento dos clientes reforça esta leitura, ao demonstrar que o crescimento do *e-commerce* assenta maioritariamente na aquisição de novos consumidores, mantendo níveis reduzidos de fidelização. A estabilidade da receita média por cliente confirma que o aumento da receita resulta essencialmente do crescimento da base de clientes, enquanto os padrões temporais de compra, o perfil etário dominante e a preferência por pagamentos fracionados fornecem informação relevante para o desenho de estratégias comerciais mais direcionadas.

No contexto do *marketplace*, a análise dos vendedores evidencia uma estrutura assimétrica, marcada por uma forte concentração do lucro num número reduzido de vendedores e por uma elevada variabilidade temporal do desempenho individual. Este padrão sugere riscos associados à dependência excessiva de alguns vendedores-chave, mas também oportunidades para promover um maior equilíbrio e diversificação do ecossistema da plataforma.

Por fim, a análise das avaliações dos clientes confirma de forma consistente a relação entre o cumprimento dos prazos de entrega e a satisfação percebida. As encomendas entregues dentro do prazo estão associadas a níveis de avaliação mais elevados, reforçando o papel central da fiabilidade logística na qualidade da experiência do cliente e na perceção global do serviço.

Em conjunto, estes resultados demonstram que os *dashboards* desenvolvidos constituem uma ferramenta eficaz de apoio à tomada de decisão, ao permitir transformar dados operacionais em *insights* claros, contextualizados e acionáveis. A solução proposta proporciona uma visão integrada do negócio, apoiando a definição de estratégias orientadas para a eficiência operacional, a melhoria da experiência do cliente, o reforço da fidelização e a sustentabilidade do *e-commerce*.

De forma a sintetizar o processo analítico desenvolvido ao longo do projeto, a Figura 60 apresenta uma representação esquemática das principais fases da solução de BI aplicada ao contexto do *e-commerce*. O diagrama ilustra o percurso desde os *datasets* de origem até à comunicação dos *insights* aos utilizadores finais, incluindo as etapas de planeamento analítico, preparação e transformação dos dados, modelação dimensional, desenvolvimento do sistema analítico e construção da narrativa baseada em dados.

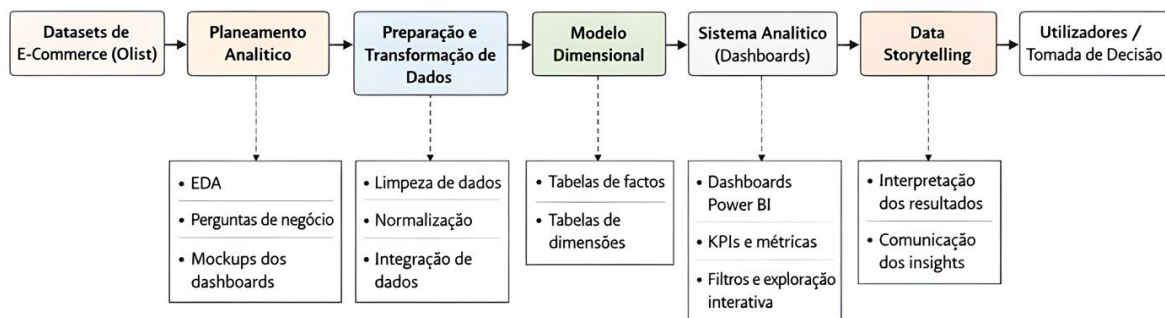


Figura 60. Esquema do processo analítico desenvolvido no projeto

## 7. Conclusão e Trabalho Futuro

O presente trabalho teve como propósito desenvolver uma solução de BI aplicada ao contexto do *e-commerce*, demonstrando como a organização estruturada dos dados, a definição consistente de indicadores de desempenho e a construção de uma narrativa analítica podem reforçar o suporte à decisão em ambientes digitais.

A aplicação desta abordagem ao *marketplace* Olist permitiu identificar padrões relevantes no comportamento do negócio, evidenciando a influência da sazonalidade no volume de encomendas, a relação entre o cumprimento dos prazos de entrega e os níveis de satisfação dos clientes, bem como a concentração do desempenho económico em determinadas categorias de produto e num conjunto restrito de vendedores. Estes resultados confirmam que o desempenho de um *marketplace* depende da articulação entre fatores comerciais, operacionais e relacionais.

A estruturação dos *dashboards* segundo uma lógica narrativa revelou-se uma mais-valia no contexto do *e-commerce*, ao permitir transformar métricas isoladas em interpretações organizadas e orientadas à ação. A integração entre BI e *data storytelling* facilitou a identificação dos principais motores de crescimento, a avaliação da eficiência logística, a análise do desempenho dos melhores vendedores e a compreensão do impacto da experiência de entrega na satisfação do cliente. Desta forma, o trabalho demonstra que a combinação entre definição estruturada de KPIs e organização narrativa da informação reforça o papel dos sistemas de BI enquanto instrumentos estratégicos de apoio à decisão.

Relativamente a trabalho futuro, considera-se pertinente aprofundar a exploração da componente qualitativa dos dados, nomeadamente através da análise dos comentários textuais associados às avaliações dos clientes, permitindo complementar os indicadores quantitativos de satisfação com uma interpretação mais detalhada da experiência de compra.

Adicionalmente, poderá ser relevante aplicar a metodologia desenvolvida a períodos temporais mais recentes ou a outros *marketplaces* digitais, com o objetivo de validar a consistência da abordagem em diferentes contextos e reforçar o seu potencial de generalização.

## Bibliografia

- Ackoff, R. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, 16(1), 3–9.
- Ahmed, H., Jilani, T., Haider, W., Asad, M., Nand, S., & Kamran, S. (2017). Establishing standard rules for choosing best KPIs for an e-commerce business based on google analytics and machine learning technique. *International Journal of Advanced Computer Science and Applications*, 8(5).  
<https://doi.org/10.14569/ijacsa.2017.080570>
- Anwar, D., Faizanuddin, M., Rahman, F., & Dayal, R. (2025). Analyzing Consumer Behavior in E-Commerce: Insights from Data-Driven Approaches. *Management (Montevideo)*, 3, 127–127. <https://doi.org/10.62486/agma2025127>
- Anzar, M. (2020). Benefits and Challenges Of E-commerce. *Studies in Indian Place Names (SIPN)*, 40(33), 11–12.
- Aparicio, M., Org, M. A., & Costa, C. J. (2015). Data Visualization. *Communication Design Quarterly*, 3(1), 7–11.
- Armbrust, M., Ghodsi, A., Xin, R., Zaharia, M., & Berkeley, U. (2021). Lakehouse: a new generation of open platforms that unify data warehousing and advanced analytics. *In Proceedings of CIDR*, 8, 28.
- Bach, B., Freeman, E., Abdul-Rahman, A., Turkay, C., Khan, S., Fan, Y., & Chen, M. (2023). Dashboard Design Patterns. *IEEE Transactions on Visualization and Computer Graphics*, 29(1), 342–352. <https://doi.org/10.1109/TVCG.2022.3209448>
- Bhargava, S., Kumar Lalwani, M., Jain, A., & Arya, S. (2023). Data Visualization on E-Commerce Data. *PRATIBODH, (NCDSNS)*.
- Bodschwinna, C., & Philipp, F. (2024). *Dashboard KPI for OST* [Doctoral dissertation]. OST Ostschweizer Fachhochschule.
- Boldosova, V. (2020). Telling stories that sell: The role of storytelling and big data analytics in smart service sales. *Industrial Marketing Management*, 86, 122–134.  
<https://doi.org/10.1016/j.indmarman.2019.12.004>
- Boldosova, V., & Luoto, S. (2020). Storytelling, business analytics and big data interpretation: Literature review and theoretical propositions. *Management Research Review*, 43(2), 204–222. <https://doi.org/10.1108/MRR-03-2019-0106>
- Boogaard, K. (2023). *How to write SMART goals*.  
<https://www.atlassian.com/blog/productivity/how-to-write-smart-goals>

- Boy, J., Rensink, R. A., Bertini, E., & Fekete, J. D. (2014). A principled way of assessing visualization literacy. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 1963–1972. <https://doi.org/10.1109/TVCG.2014.2346984>
- Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8), 88–98. <https://doi.org/10.1145/1978542.1978562>
- Chee, C.-H., Yeoh, W., & Gao, S. (2011). Enhancing business intelligence traceability through an integrated metadata framework. *ACIS 2011 Proceedings*. <https://aisel.aisnet.org/acis2011/11/>
- Cleveland, W. S., & McGill, R. (1984). Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, 79(387), 531–554.
- Daradkeh, M., & Atalla, S. (2023). From data to narrative: Automating and engineering the art of data storytelling. *2023 International Conference on Information Technology (ICIT)*, 550–555. <https://doi.org/10.1109/ICIT58056.2023.10225929>
- Dykes, B. (2019). *Effective data storytelling: How to drive change with data, narrative and visuals*. John Wiley & Sons.
- Echeverria, V., Martinez-Maldonado, R., Granda, R., Chiluíza, K., Conati, C., & Shum, S. B. (2018). Driving data storytelling from learning design. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, 131–140. <https://doi.org/10.1145/3170358.3170380>
- El-Sappagh, S. H. A., Hendawi, A. M. A., & El Bastawissy, A. H. (2011). A proposed model for data warehouse ETL processes. *Journal of King Saud University - Computer and Information Sciences*, 23(2), 91–104. <https://doi.org/10.1016/j.jksuci.2011.05.005>
- Ferreira, T., Pedrosa, I., & Bernardino, J. (2019). Integration of Business Intelligence with e-commerce. *2019 14th Iberian Conference on Information Systems and Technologies (CISTI)*, 1–7.
- Foshay, N., Taylor, A., & Mukherjee, A. (2014). Winning the hearts and minds of business intelligence users: The role of metadata. *Information Systems Management*, 31(2), 167–180. <https://doi.org/10.1080/10580530.2014.890444>
- Fricke, M. (2019). The knowledge pyramid: A critique of the DIKW hierarchy. *Journal of Information Science*, 46(1), 131–142. <https://doi.org/10.5771/0943-7444-2019-1-33>
- Ganeshan, A., Macari, E., O'Brien, J., Schlegel, K., & Long, C. (2025, June 16). *2025 Gartner® Magic Quadrant™ for Analytics and Business Intelligence Platforms*. <https://www.qlik.com/us/gartner-magic-quadrant-business-intelligence>

- Garani, G., Chernov, A. V., Savvas, I. K., & Butakova, M. A. (2019). A Data Warehouse Approach for Business Intelligence. *2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, 70–75. <https://doi.org/10.1109/WETICE.2019.00022>
- Ghandour, A. (2015). Big data driven e-commerce architecture. *International Journal of Economics, Commerce and Management*, 3(5), 940–947.
- Green, E., Hattery, M., & Watt, C. (2023). *Using tables, charts, maps, and other visuals: Guidance for better management reports and papers*. <https://dyson.cornell.edu/outreach/extension-bulletins/>
- Hayward, E. (2024). *4 Types of Dashboards: Operational, Analytical, Strategic & Tactical*. <https://www.klipfolio.com/blog/starter-guide-to-dashboards>
- Herschel, R. T., & Clements, N. (2017). The importance of storytelling in business intelligence. *International Journal of Business Intelligence Research*, 8(1), 26–39. <https://doi.org/10.4018/ijbir.2017010102>
- Howson, A. C., Richardson, J., Sallam, R., & Kronz, A. (2019). *Magic Quadrant for Analytics and Business Intelligence Platforms*. <https://www.gartner.com/doc/reprints?id=1-68720FP&ct=190213&st=sb>
- Hullman, J., & Diakopoulos, N. (2011). Visualization rhetoric: Framing effects in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), 2231–2240.
- Ishaq Bhatti, M., & Awan, H. M. (2014). The key performance indicators (KPIs) and their impact on overall organizational performance. *Quality & Quantity*, 48(6), 3127–3143. <https://doi.org/10.1007/s11135-013-9945-y>
- Ishwarappa, & Anuradha, J. (2015). A brief introduction on Big Data 5Vs characteristics and Hadoop technology. *Procedia Computer Science*, 48, 319–324. <https://doi.org/10.1016/j.procs.2015.04.188>
- Islam, M., & Jin, S. (2019). An overview of data visualization. *2019 International Conference on Information Science and Communications Technologies (ICISCT)*, 1–7.
- Jain, V., Malviya, B., & Arya, S. (2021). An overview of electronic commerce (e-Commerce). *Journal of Contemporary Issues in Business and Government*, 27(3). <https://doi.org/10.47750/cibg.2021.27.03.090>
- Jiang, L. (2022). Visual design elements based on digital visualization. *Soft Computing – A Fusion of Foundations, Methodologies & Applications*, 26(16), 7855–7863. <https://doi.org/10.1007/s00500-022-06811-8>

- Jorgensen, M., Svarre, T., & Christensen, T. (2021). The role of end users in efficient business intelligence solutions: A preliminary study. *Proceedings of the 2021 4th International Conference on Computers in Management and Business*, 1–4.
- Joshi, P., & Dumbre, G. M. (2017). Basic concept of e-commerce. *International Research Journal of Multidisciplinary Studies*, 3(3), 2454–8499.
- Kale, K. K. (2022). A study on benefits and challenges in an emerging economy. *International Journal of Research in Social Sciences & Informatics Studies*, 8(3), 92–99.
- Karampournioti, E., & Wiedmann, K. P. (2021). Storytelling in online shops: the impacts on explicit and implicit user experience, brand perceptions and behavioral intention. *Internet Research*, 32(7), 228–259. <https://doi.org/10.1108/INTR-09-2019-0377>
- Keong Choong, K., & Bacon, L. (2024). A metric and indicator performance measurement system for e-commerce organizations: A consensus analysis of their usefulness. *Electronic Commerce Research and Applications*, 67. <https://doi.org/10.1016/j.elerap.2024.101420>
- Khalajzadeh, H., Simmons, A. J., & Abdelrazek, M. (2020). End-user-oriented tool support for modeling data analytics requirements. *2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 4(CSCW1). <https://doi.org/10.1145/3392826>
- Lopes, J., Guimarães, T., & Santos, M. F. (2020). Adaptive business intelligence: A new architectural approach. *Procedia Computer Science*, 177, 540–545. <https://doi.org/10.1016/j.procs.2020.10.075>
- Lundqvist, A., Liljander, V., Gummerus, J., & Van Riel, A. (2013). The impact of storytelling on the consumer brand experience: The case of a firm-originated story. *Journal of Brand Management*, 20(4), 283–297. <https://doi.org/10.1057/bm.2012.15>
- Luo, X., Yuan, Y., Zhang, K., Xia, J., Zhou, Z., Chang, L., & Gu, T. (2019). Enhancing statistical charts: toward better data visualization and analysis. *Journal of Visualization*, 22(4), 819–832. <https://doi.org/10.1007/s12650-019-00569-2>
- M Kirmani, M. (2017). Dimensional modeling using star schema for data warehouse creation. *Oriental Journal of Computer Science and Technology*, 10(4), 745–754. <https://doi.org/10.13005/ojcs/10.04.07>
- Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*, 5(2), 110–141.
- Matei, S. A., & Hunter, L. (2023). The data lakehouse: Data warehousing and more. In *The Information Society* (Vol. 37, Number 5). Routledge. <https://doi.org/10.1080/01972243.2021.1951415>

- Mazumdar, D., Hughes, J., & Onofre, J. (2023). *The Data Lakehouse: Data Warehousing and More*. <http://arxiv.org/abs/2310.08697>
- McDowell, K. (2021). Storytelling wisdom: Story, information, and DIKW. *Journal of the Association for Information Science and Technology*, 72(10), 1223–1233. <https://doi.org/10.1002/asi.24466>
- MUNTEAN, M. I., TARNAVEANU, D., & ION, A. R. (2016). E-commerce performance. Shopping cart key performance indicators. *Informatica Economica*, 20(1), 62–71. <https://doi.org/10.12948/issn14531305/20.1.2016.06>
- Negash, S., & Gray, P. (2008). *Business intelligence* (pp. 175–193). Springer.
- Nepal, A. (2017). *Key performance indicators for e-commerce*.
- Ntumba, C., Aguayo, S., & Maina, K. (2023). Revolutionizing Retail: A Mini Review of E-commerce Evolution. *Journal of Digital Marketing and Communication*, 3(2), 100–110. <https://doi.org/10.53623/jdmc.v3i2.365>
- Ong, I., Siew, P., & Wong, S. (2011). A Five-Layered Business Intelligence Architecture. *Communications of the IBIMA*, 1–11. <https://doi.org/10.5171/2011.695619>
- Pan, C. L., Bai, X., Li, F., Zhang, D., Chen, H., & Lai, Q. (2021). How business intelligence enables e-commerce: Breaking traditional e-commerce mode and driving the transformation of digital economy. *2021 2nd International Conference on E-Commerce and Internet Technology (ECIT)*, 26–30. <https://doi.org/10.1109/ECIT52743.2021.00013>
- Pandey, A. V., Rall, K., Satterthwaite, M. L., Nov, O., & Bertini, E. (2015). How deceptive are deceptive visualizations? An empirical analysis of common distortion techniques. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI 2015), 2015-April*, 1469–1478. <https://doi.org/10.1145/2702123.2702608>
- Plaisant, C. (2004). The challenge of information visualization evaluation. *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI 2004)*, 109–116.
- Praful Bharadiya, J. (2023). A comparative study of business intelligence and artificial intelligence with big data analytics. *American Journal of Artificial Intelligence*, 7(1). <https://doi.org/10.11648/j.ajai.20230701.14>
- Qin, X., Luo, Y., Tang, N., & Li, G. (2020). Making data visualization more efficient and effective: A survey. *The VLDB Journal*, 29(1), 93–117. <https://doi.org/10.1007/s00778-019-00588-3>
- Quenum, V. J., Mosconi, E., & Gagnon, E. (2025). Enhancing Data Storytelling: A Conceptual Framework. *Proceedings of the 58th Hawaii International Conference on System Sciences (HICSS 2025)*.

- Ranjan, J. (2009). Business intelligence: Concepts, components, techniques and benefits. *Journal of Theoretical and Applied Information Technology*, 9(1), 60–70.
- Rebrova, A. (2024, August 15). *Deliver Projects On Time: Mastering Functional and Non-Functional Requirements for Success*. <https://maddevs.io/blog/functional-vs-non-functional-requirements/>
- Ren, P., Wang, Y., & Zhao, F. (2023). Re-understanding of data storytelling tools from a narrative perspective. *Visual Intelligence*, 1(1). <https://doi.org/10.1007/s44267-023-00011-0>
- Riwayat, A. A. P., Susilawati, A. D., & Naqiah, Z. (2024). Purchasing patterns analysis in e-commerce: A big data-driven approach and methodological. *International Journal of Software Engineering and Computer Science*, 4(1), 148–164. <https://doi.org/10.35870/ijsecs.v4i1.2384>
- Rosli, W., & Cabrera, A. (2015). Gestalt principles in multimodal data representation. *IEEE Computer Graphics and Applications*, 35(2), 80–87.
- Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, 33(2), 163–180. <https://doi.org/10.1177/0165551506070706>
- Shadare, A. E., Musa, S. M., Akujuobi, C., Sadiku, M. N. O., Akujuobi, C. M., & Perry, R. G. (2016). Data visualization. *International Journal of Engineering Research and Advanced Technology*, 2(12), 11–16.
- Sharma, A. (2010). Improve performance of extract, transform and load (ETL) in data warehouse. *Article in International Journal on Computer Science and Engineering*, 2(3), 786–789. <https://www.researchgate.net/publication/49618608>
- Sharma, M. (2019). *A study of narrative engagement in e-commerce* [Master's]. MICA (Mudra Institute of Communications), Ahmedabad, India.
- Shneiderman, B. (2003). The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization* (pp. 364–371). <https://doi.org/10.1016/B978-1-55860-915-0.50046-9>
- Sionek, A. (2018). *Brazilian E-Commerce Public Dataset by Olist*. <https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce>
- Slutsky, D. (2014). The effective use of graphs. *Journal of Wrist Surgery*, 3(2), 67–68. <https://doi.org/10.1055/s-0034-1375704>
- Tejero -Dakay, I., Ferolin, R. J., & Ceniza -Canillo, A. (2023). *End-user business intelligence tools adoption in higher education*.
- Tewari, A. (2019, August 11). *Art of storytelling in Data Science*. <https://www.linkedin.com/pulse/art-storytelling-data-science-anshul-tewari/>

- Tsai, Y. C., & Cheng, Y. T. (2012). Analyzing key performance indicators (KPIs) for e-commerce and Internet marketing of elderly products: A review. *Archives of Gerontology and Geriatrics*, 55(1), 126–132. <https://doi.org/10.1016/j.archger.2011.05.024>
- Tucakov, J. (2025, March 28). *50+ Different Types of Graphs and Charts to Make Your Data Speak*. <https://wpdatatables.com/types-of-charts/>
- Tufte, E. R. (1983). *The Visual Display of Quantitative Information* (9th ed., Vol. 2). Graphics Press.
- Unwin, A. (2020). Why is data visualization important? What is important in data visualization. *Harvard Data Science Review*, 2(1). <https://doi.org/10.1162/99608f92.8ae4d525>
- Vamsi, P., Rajesh, R., & Lakshman, M. (2025). E-Commerce Dashboard. *International Journal of Innovative Science and Research Technology*, 10(3), 1397–1405. <https://doi.org/10.38124/ijisrt/25mar897>
- Vanderplas, S., Cook, D., & Hofmann, H. (2020). Testing statistical charts: What makes a good graph? *Annual Review of Statistics and Its Application*, 7(1), 61–88. <https://doi.org/10.1146/annurev-statistics-031219>
- Velimirovića, D., Velimirović, M., & Stanković, R. (2010). Role and importance of key performance indicators measurement. *Serbian Journal of Management*, 6(1), 63–72.
- Ware, C. (2019). *Information visualization: Perception for design*. Morgan Kaufmann.
- Zimmer, M., Baars, H., & Kemper, H. G. (2012). The impact of requirements on business intelligence architectures. *2012 45th Hawaii International Conference on System Sciences (HICSS)*, 4189–4198. <https://doi.org/10.1109/HICSS.2012.567>

## Anexos

### Anexo A – Estrutura dos Conjuntos de Dados

#### Clientes

| Coluna                          | Tipo de Dado | Descrição   |
|---------------------------------|--------------|---|
| <b>customer_id</b>              | object       | Identificador único atribuído a cada cliente                    |
| <b>customer_unique_id</b>       | object       | Identificador persistente do cliente, mesmo em caso de recompra |
| <b>customer_zip_code_prefix</b> | int64        | Primeiros dígitos do código postal do cliente                   |
| <b>customer_city</b>            | object       | Cidade de residência do cliente                                 |
| <b>customer_state</b>           | object       | Estado de residência do cliente                                 |

#### Geolocalização

| Coluna                             | Tipo de Dado | Descrição  |
|------------------------------------|--------------|--|
| <b>geolocation_zip_code_prefix</b> | int64        | Prefixo do código postal associado à localização |
| <b>geolocation_lat</b>             | float64      | Latitude geográfica                              |
| <b>geolocation_lng</b>             | float64      | Longitude geográfica                             |
| <b>geolocation_city</b>            | object       | Cidade correspondente à coordenada               |
| <b>geolocation_state</b>           | object       | Estado correspondente à coordenada               |

### Encomendas

| Coluna                               | Tipo de Dado | Descrição   |
|--------------------------------------|--------------|---|
| <b>order_id</b>                      | object       | Identificador único da encomenda                          |
| <b>customer_id</b>                   | object       | Referência ao cliente que efetuou a encomenda             |
| <b>order_status</b>                  | object       | Estado atual da encomenda                                 |
| <b>order_purchase_timestamp</b>      | object       | Data e hora da realização da compra                       |
| <b>order_approved_at</b>             | object       | Data de aprovação do pagamento                            |
| <b>order_delivered_carrier_date</b>  | object       | Data em que a encomenda foi recolhida pela transportadora |
| <b>order_delivered_customer_date</b> | object       | Data em que a encomenda foi entregue ao cliente           |
| <b>order_estimated_delivery_date</b> | object       | Data estimada de entrega                                  |

### Itens da Encomenda

| Coluna                     | Tipo de Dado | Descrição                                     |
|----------------------------|--------------|---|
| <b>order_id</b>            | object       | Identificador da encomenda                    |
| <b>order_item_id</b>       | int64        | Número sequencial do item dentro da encomenda |
| <b>product_id</b>          | object       | Identificador do produto                      |
| <b>seller_id</b>           | object       | Identificador do vendedor                     |
| <b>shipping_limit_date</b> | object       | Prazo máximo para expedição                   |
| <b>price</b>               | float64      | Preço do produto                              |
| <b>freight_value</b>       | float64      | Valor do custo de transporte                  |

### Pagamentos

| Coluna                      | Tipo de Dado | Descrição                      |
|-----------------------------|--------------|--------------------------------|
| <b>order_id</b>             | object       | Identificador da encomenda     |
| <b>payment_sequential</b>   | int64        | Número sequencial do pagamento |
| <b>payment_type</b>         | object       | Tipo de pagamento utilizado    |
| <b>payment_installments</b> | int64        | Número de prestações           |
| <b>payment_value</b>        | float64      | Valor total pago               |

### Avaliações

| Coluna                         | Tipo de Dado | Descrição   |
|--------------------------------|--------------|---|
| <b>review_id</b>               | object       | Identificador único da avaliação                  |
| <b>order_id</b>                | int64        | Referência à encomenda avaliada                   |
| <b>review_score</b>            | object       | Pontuação atribuída pelo cliente (1 a 5)          |
| <b>review_comment_title</b>    | int64        | Título do comentário                              |
| <b>review_comment_message</b>  | float64      | Texto do comentário                               |
| <b>review_creation_date</b>    | object       | Data em que a pesquisa de satisfação foi enviada  |
| <b>review_answer_timestamp</b> | object       | Data e hora em que o cliente respondeu à pesquisa |

### Produtos

| Coluna                            | Tipo de Dado | Descrição                               |
|-----------------------------------|--------------|---|
| <b>product_id</b>                 | object       | Identificador do produto                |
| <b>product_category_name</b>      | object       | Categoria do produto em português       |
| <b>product_name_length</b>        | float64      | Número de caracteres no nome do produto |
| <b>product_description_length</b> | float64      | Número de caracteres na descrição       |
| <b>product_photos_qty</b>         | float64      | Quantidade de fotografias do produto    |
| <b>product_weight_g</b>           | float64      | Peso do produto (gramas)                |

|                          |         |                             |
|--------------------------|---------|-----------------------------|
| <b>product_length_cm</b> | float64 | Comprimento do produto (cm) |
| <b>product_height_cm</b> | float64 | Altura do produto (cm)      |
| <b>product_width_cm</b>  | float64 | Largura do produto (cm)     |

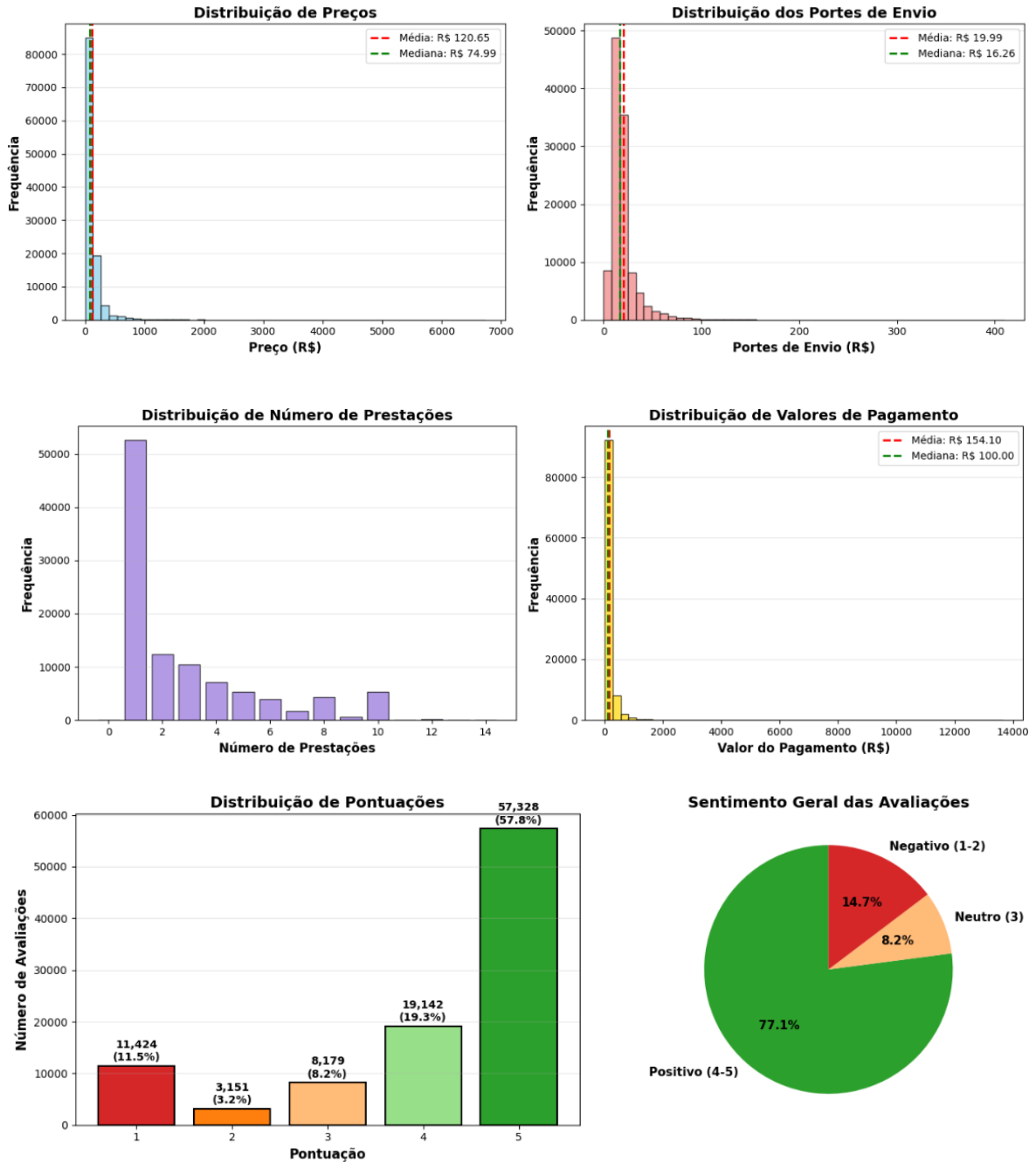
### Tradução de Categorias

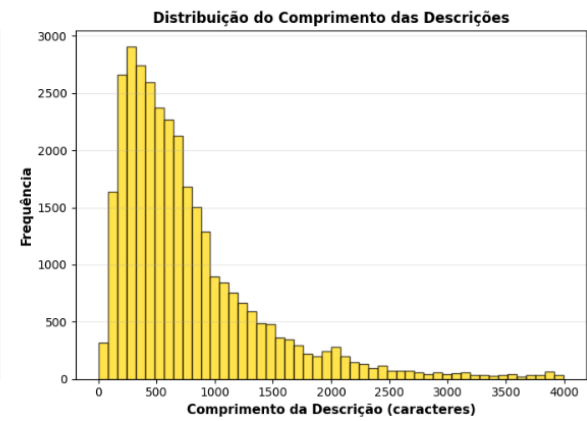
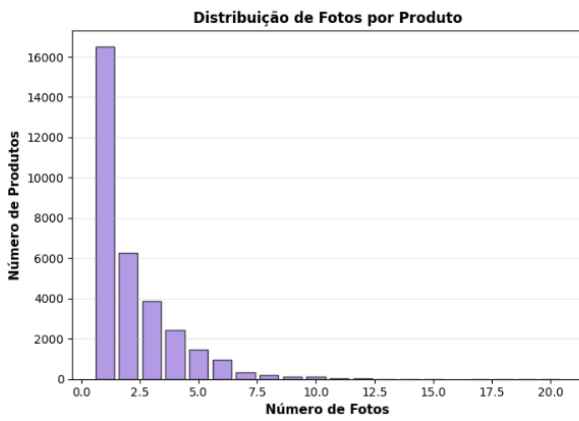
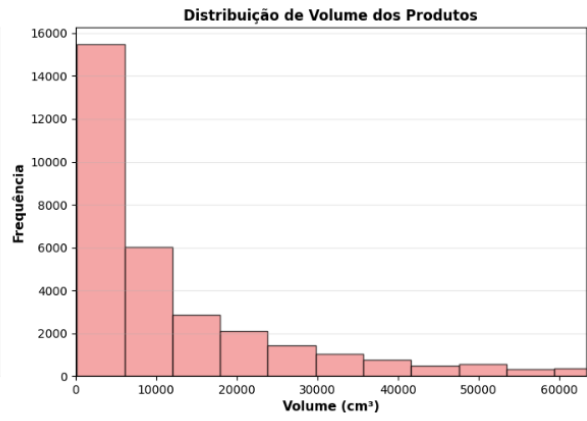
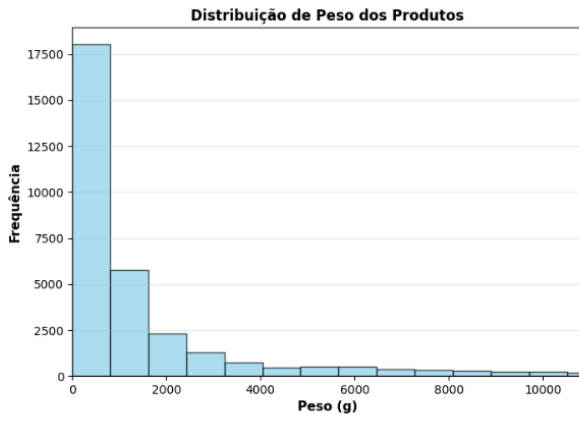
| Coluna                               | Tipo de Dado | Descrição                                 |
|--------------------------------------|--------------|---|
| <b>product_category_name</b>         | object       | Nome da categoria do produto em português |
| <b>product_category_name_english</b> | object       | Tradução do nome da categoria para inglês |

### Vendedores

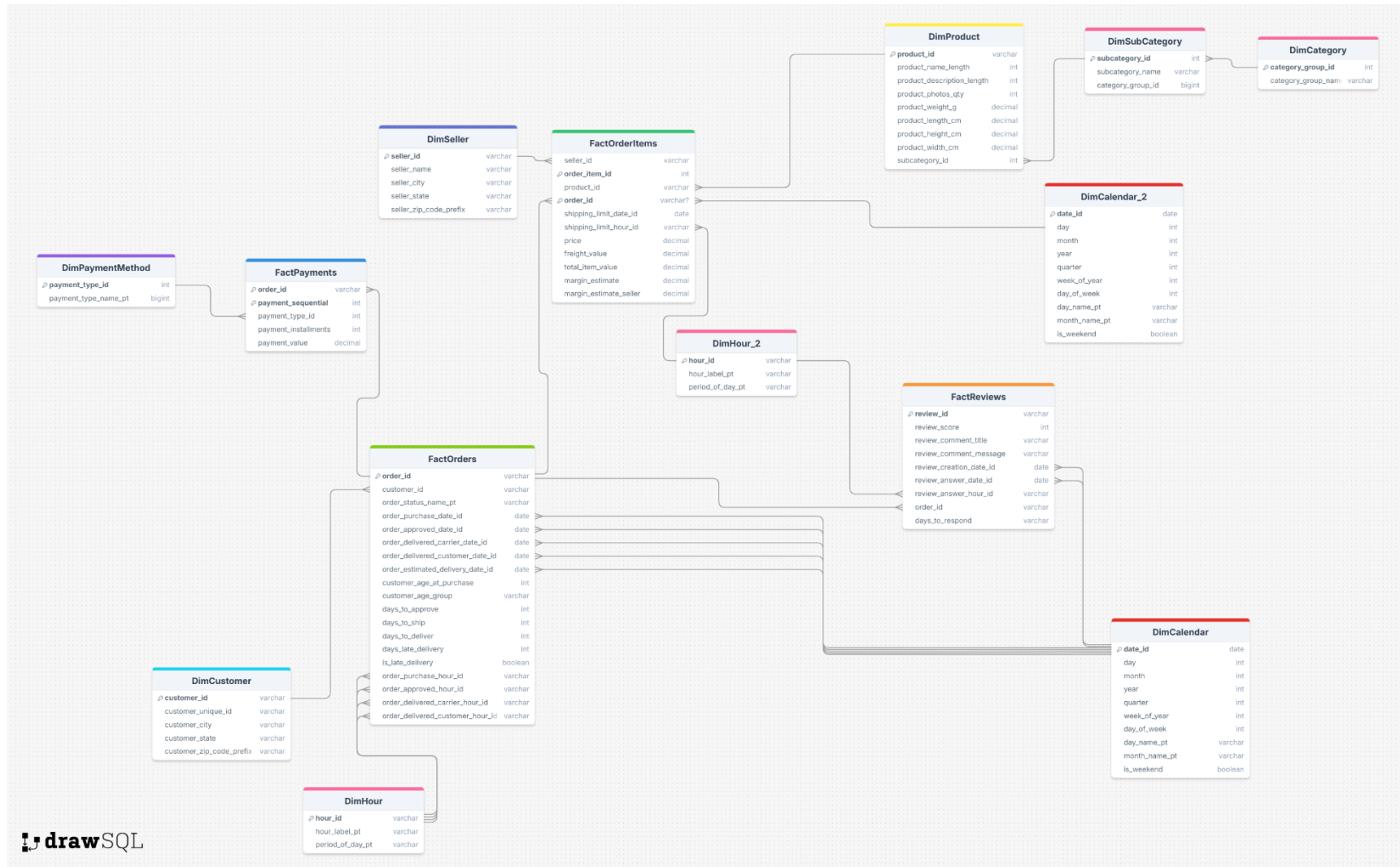
| Coluna                        | Tipo de Dado | Descrição                                    |
|-------------------------------|--------------|--|
| <b>seller_id</b>              | object       | Identificador único atribuído a cada cliente |
| <b>seller_zip_code_prefix</b> | int64        | Prefixo do código postal do vendedor         |
| <b>seller_city</b>            | object       | Cidade do vendedor                           |
| <b>seller_state</b>           | object       | Estado do vendedor                           |

## Anexo B – Análise Gráfica das Variáveis Numéricas



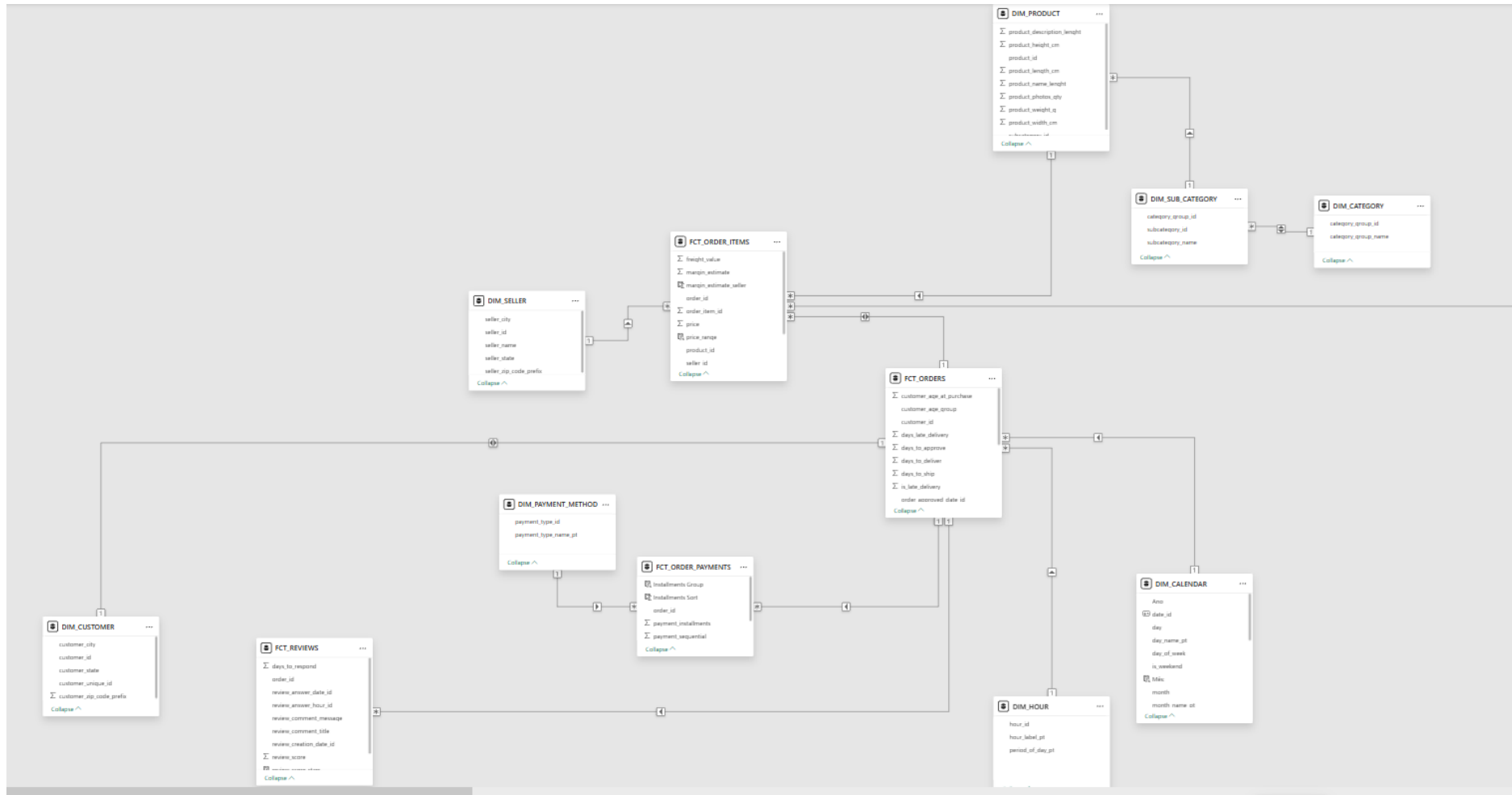


## Anexo C – Modelo Dimensional



drawSQL

## Anexo D – Modelo de Dados Implementado no Power BI



## Anexo E – Mockups dos Dashboards

