



INSTITUTO POLITÉCNICO
DO CÁVADO E DO AVE
ESCOLA SUPERIOR
DE TECNOLOGIA

Data Processing and Visualization (P02)

DECISION SUPPORT SYSTEMS, 2022-23

Nuno Mendes (2727), Rosário Silva (21138), Tiago Azevedo (21153)

Introduction

No mundo atual, caracterizado pela sua natureza altamente digital e interconectada, a capacidade de processar e visualizar dados tornou-se uma competência essencial em várias disciplinas, incluindo na Engenharia de Sistemas Informáticos. Através da aplicação eficaz de técnicas de processamento de dados, é possível extrair insights valiosos que podem auxiliar na tomada de decisões e promover uma melhor compreensão dos fenómenos em estudo.

Neste contexto, este projeto visa o desenvolvimento de uma solução de software para processamento e visualização de dados, no âmbito da disciplina de Sistemas de Apoio à Decisão (DSS). A solução a ser desenvolvida irá integrar várias fontes de dados, implementar operações de preparação de dados, cálculo de colunas e medidas, apresentação e visualização de dados.

A principal fonte de dados para este projeto será o data mart WWI, complementada por pelo menos uma outra fonte de dados auxiliar. Através da combinação e análise destes dados, procuraremos obter insights que possam contribuir para uma melhor compreensão do contexto em estudo.

No seguimento do trabalho em estudo, foram efetuadas as seguintes “pesquisas” aos dados em tratamento:

1. **Vendas por Dia da Semana:** Existe uma variação significativa nas vendas por dia da semana? Existem dias específicos que registam consistentemente vendas mais altas ou mais baixas?
2. **Média de Vendas por Mês:** Como é que as vendas variam ao longo do ano? Existem meses específicos que registam consistentemente vendas mais altas ou mais baixas?
3. **Vendas por Categoria e Produto:** Quais são as categorias e produtos mais vendidos? Existem categorias ou produtos que registam consistentemente vendas mais altas ou mais baixas?
4. **Vendas por Cliente:** Quais são os clientes que mais compram? Existem clientes que registam consistentemente compras mais altas ou mais baixas?
5. **Novos Clientes por Mês:** Qual é a taxa de aquisição de novos clientes por mês? Existem meses específicos que registam consistentemente uma maior aquisição de novos clientes?
6. **Valor Médio de Vendas por Mês:** Qual é o valor médio de vendas por mês? Como é que este valor varia por estado?

7. **Relação entre a População e as Vendas:** Existe uma correlação entre a população de um estado e as vendas nesse estado? Estados com populações maiores registam vendas mais altas?
8. **Relação entre o Poder de Compra e as Vendas:** Existe uma correlação entre o poder de compra em um estado e as vendas nesse estado? Estados com maior poder de compra registam vendas mais altas?

Data acquisition and preparation

Conexões às Fontes de Dados: A principal fonte de dados para este projeto foi o data mart de vendas, que está armazenado numa base de dados PostgreSQL. Utilizamos o Power BI para estabelecer uma conexão direta a esta base de dados, permitindo-nos importar os dados diretamente para o nosso ambiente de produção. Além disso, planeamos incorporar dados externos sobre a população e o poder de compra por estado, que seriam importados para o Power BI da mesma forma.

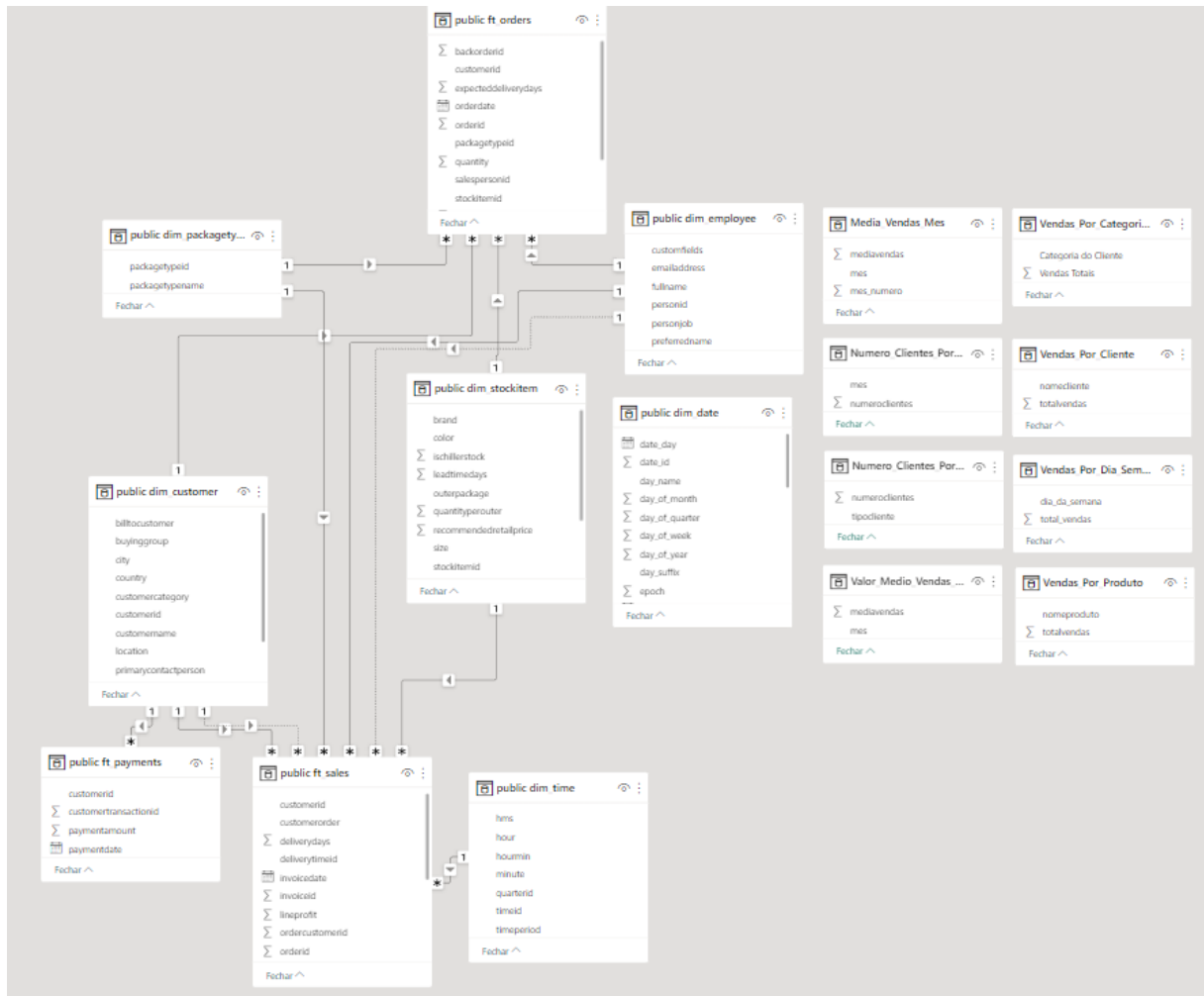
Limpeza de Dados: A limpeza de dados foi realizada principalmente através de consultas SQL. Estas consultas foram usadas para extrair os dados relevantes do data mart e para transformar os dados de uma forma que fosse adequada para análise. Por exemplo, usamos funções SQL para extrair partes específicas de datas (como o dia da semana ou o mês) e para calcular agregações (como a soma total ou a média de vendas).

Preparação de Dados: A preparação de dados envolveu a criação de consultas SQL que combinam dados de várias tabelas do data mart. Por exemplo, criamos consultas que juntam dados de vendas com informações sobre clientes e produtos. Além disso, planeamos usar o Power Query Editor no Power BI para, eventualmente, juntar os nossos dados internos e externos.

Transformação de Dados: As consultas SQL também foram usadas para transformar os dados de uma forma que fosse mais útil para a análise. Por exemplo, calculamos a média de vendas por mês e o número de clientes por mês. Além disso, planeamos usar o Power BI para criar visualizações que nos permitisse ver os dados de uma maneira mais visual e intuitiva.

No geral, o objetivo dessas tarefas foi garantir que os dados estivessem limpos, consistentes e prontos para análise.

Data modelling and processing



Data visualization

Objetivos dos Dashboards: Os dashboards que estamos a criar têm como objetivo fornecer insights sobre as vendas da empresa. Estamos interessados em entender padrões de vendas ao longo do tempo, como as vendas variam por dia da semana e por mês, e como as vendas se relacionam com diferentes categorias de clientes e produtos. Além disso, queremos explorar a relação entre as vendas e dados externos, como a população e o poder de compra por estado.

Perfil dos Utilizadores Finais: Os dashboards serão utilizados principalmente por gestores e analistas de vendas na empresa. Estes utilizadores estão interessados em entender as tendências de vendas para tomar decisões informadas sobre estratégias de marketing e vendas. Eles precisam de informações claras e acessíveis que possam ajudá-los a identificar oportunidades e desafios.

Descrição dos Dashboards: Os dashboards incluirão uma variedade de visualizações, incluindo gráficos de barras, gráficos de linhas e tabelas. Por exemplo, um gráfico de barras pode mostrar as vendas totais por dia da semana, enquanto um gráfico de linhas pode mostrar a média de vendas por mês ao longo do tempo. As tabelas podem ser usadas para apresentar dados detalhados sobre vendas por categoria de cliente ou produto.

Elementos Relevantes: Os dashboards incluirão vários elementos interativos, como slicers, que permitem aos utilizadores filtrar os dados com base em critérios específicos. Por exemplo, um slicer pode permitir que os utilizadores vejam dados apenas para um determinado estado ou mês. Além disso, os dashboards incluirão KPIs (Indicadores Chave de Desempenho) que fornecem uma visão rápida do desempenho das vendas, como vendas totais ou média de vendas. Também planeamos explorar o uso de influenciadores no Power BI, que podem ajudar a identificar fatores que estão a ter um impacto significativo nas vendas.

Soma de mediavendas por mes

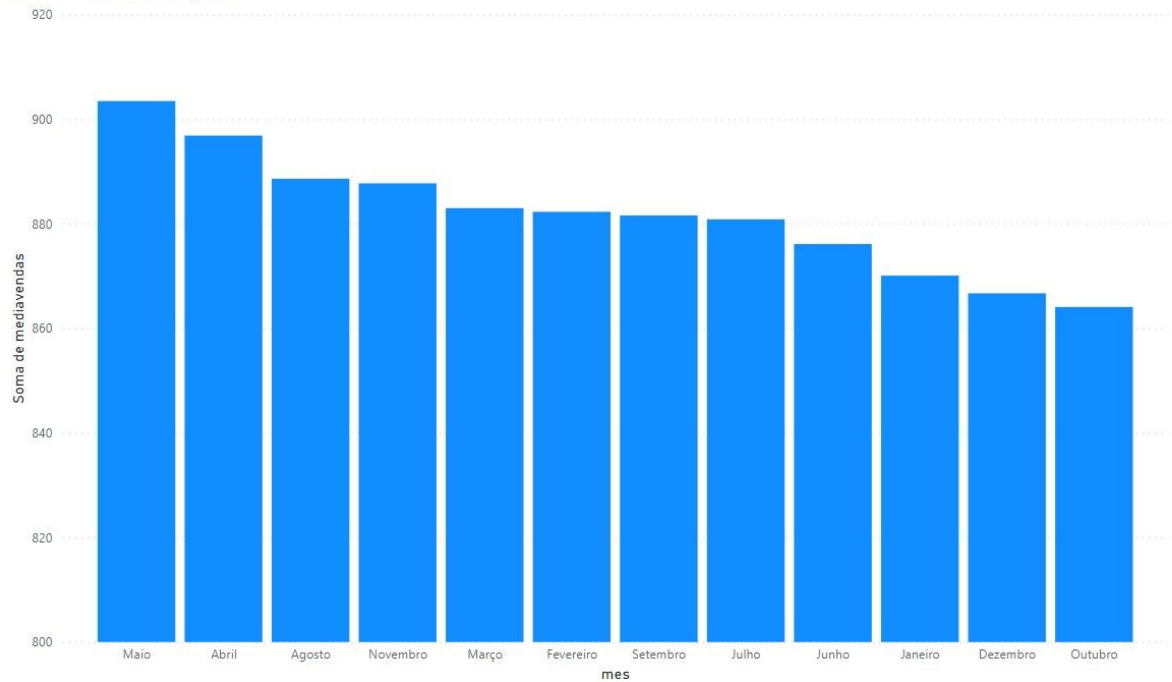


Figure 1 - Média de Vendas por Mês

Soma de Vendas Totais por Categoria do Cliente

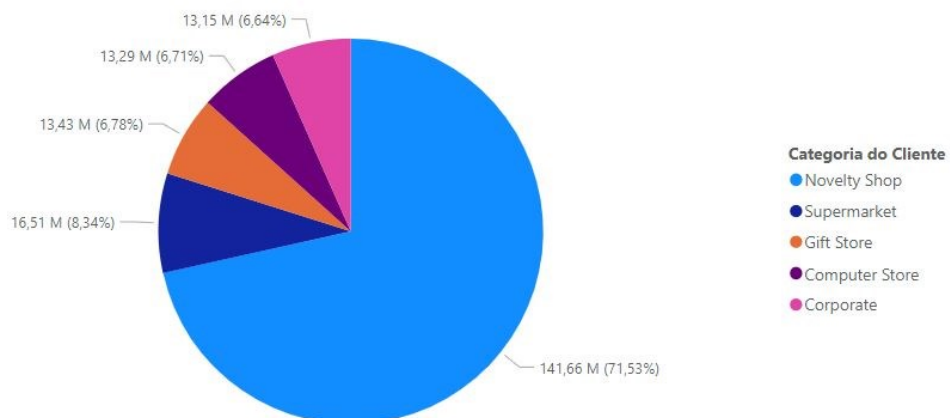


Figure 2 - Vendas por Categoria do Cliente

| nomecliente | Soma de totalvendas |
|------------------------------------|-----------------------|
| Tailspin Toys (Inguadona, MN) | 441.919,01 |
| Tailspin Toys (Minidoka, ID) | 436.459,73 |
| Mauno Laurila | 433.675,99 |
| Wingtip Toys (Sarversville, PA) | 427.962,53 |
| Ingrida Zeltina | 423.257,08 |
| Tailspin Toys (Long Meadow, MD) | 422.244,73 |
| Nasrin Omidzadeh | 421.690,63 |
| Wingtip Toys (Cuyamungue, NM) | 420.623,45 |
| Wingtip Toys (San Jacinto, CA) | 420.068,20 |
| Wingtip Toys (Morrison Bluff, AR) | 414.566,09 |
| Wingtip Toys (West Frostproof, FL) | 414.393,90 |
| Wingtip Toys (Grabill, IN) | 413.186,15 |
| Camille Authier | 412.445,67 |
| Daniel Martensson | 412.300,47 |
| Laszlo Gardenier | 411.042,74 |
| Dinh Mai | 410.758,51 |
| Total | 204.222.315,38 |

Figure 3 - Vendas por Cliente

Soma de total_vendas por dia_da_semana

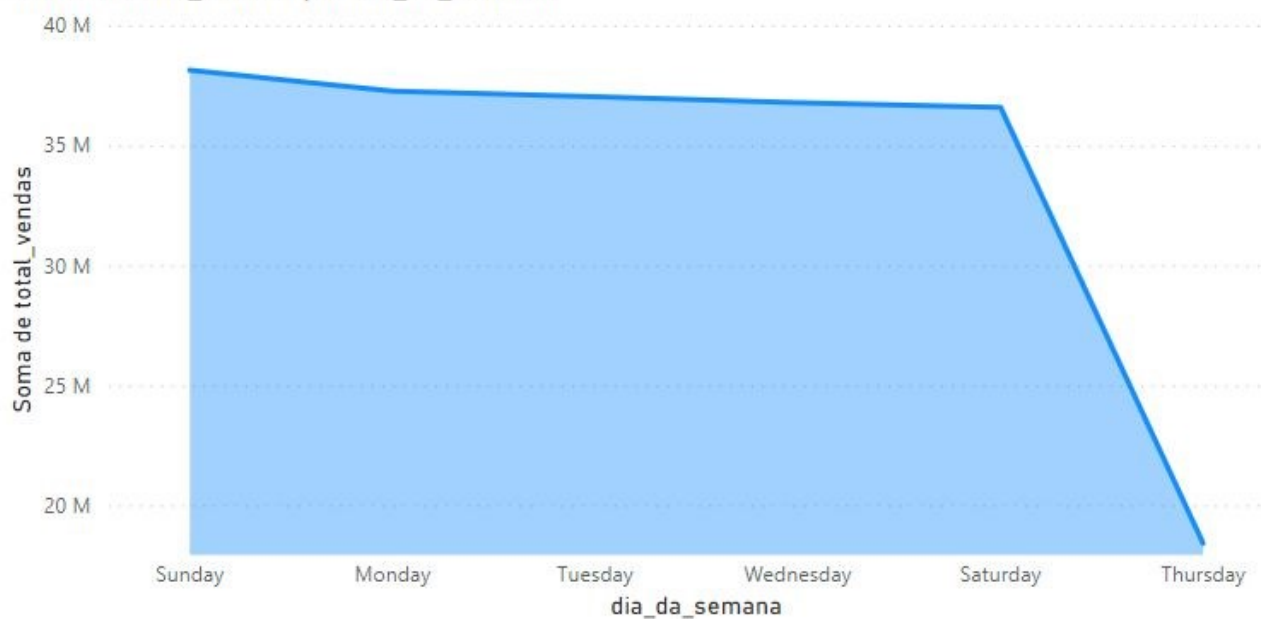


Figure 4 - Vendas por dia da Semana

| nomeproduto | Soma de totalvendas |
|--|-----------------------|
| Air cushion machine (Blue) | 12.773.338,65 |
| 32 mm Anti static bubble wrap (Blue) 50m | 7.341.600,00 |
| 10 mm Anti static bubble wrap (Blue) 50m | 7.278.430,50 |
| 20 mm Double sided bubble wrap 50m | 7.146.468,00 |
| 32 mm Double sided bubble wrap 50m | 7.118.776,00 |
| 10 mm Double sided bubble wrap 50m | 6.834.450,00 |
| 20 mm Anti static bubble wrap (Blue) 50m | 6.664.986,00 |
| 32 mm Anti static bubble wrap (Blue) 20m | 3.335.184,00 |
| Void fill 400 L bag (White) 400L | 3.302.225,00 |
| 20 mm Anti static bubble wrap (Blue) 20m | 2.839.005,00 |
| 10 mm Anti static bubble wrap (Blue) 20m | 2.766.141,00 |
| Total | 204.222.315,38 |

Figure 5 - Vendas por Produto

Conclusion

Ao longo deste projeto, conseguimos realizar uma análise de dados robusta e significativa usando o Power BI e SQL. O processo começou com a conexão ao nosso data mart em PostgreSQL, onde realizamos várias consultas SQL para extrair os dados necessários para a nossa análise. Durante este processo, realizamos várias tarefas de limpeza e preparação de dados para garantir que os dados estavam prontos para análise.

Através do Power BI, conseguimos criar dashboards interativos que nos permitiram visualizar os nossos dados de várias maneiras. Estes dashboards nos deram insights valiosos sobre as vendas da empresa, incluindo padrões de vendas ao longo do tempo e diferenças nas vendas com base em categorias de clientes e produtos. Além disso, conseguimos incorporar dados externos sobre a população e o poder de compra, o que nos permitiu explorar a relação entre esses fatores e as vendas.

O Power BI provou ser uma ferramenta poderosa para este projeto, permitindo-nos manipular e visualizar os nossos dados de maneira eficaz. A sua interface intuitiva e capacidades de visualização robustas tornaram a análise de dados uma tarefa mais simples e acessível. Além disso, a capacidade de integrar SQL diretamente no Power BI permitiu-nos aproveitar a nossa familiaridade com SQL enquanto trabalhávamos com os nossos dados.

Em geral, este projeto demonstrou o valor da análise de dados na tomada de decisões informadas. Com as ferramentas certas e uma abordagem estruturada, podemos transformar grandes volumes de dados em insights acionáveis que podem ajudar a impulsionar o sucesso de uma empresa.