

Relatório Final

**Dashboard Interativo**

Projeto de Visualização de Dados

**Nome dos Alunos:** Dionísio Miron, Duarte Capinha e Pedro Santos

**Nome do Professor:** Sofia Fernandes

Mestrado em Data Science | 09/06/2025

**Direitos de cópia**

TMS – Talent Management System, Copyright de (*Dionísio Miron)* Universidade Lusófona.

A Escola de Comunicação, Arquitetura, Artes e Tecnologias da Informação (ECATI) e a Universidade Lusófona (UL) têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

# Índice

[1 Índice iii](#_Toc200295087)

[2 Lista de Figuras iv](#_Toc200295088)

[3 Introdução 1](#_Toc200295089)

[4 Planeamento e Fundamentação das Visualizações 2](#_Toc200295090)

[4.1 Gráficos Propostos na Fase de Planeamento (Fase 1) 2](#_Toc200295091)

[5 Problemas e Desafios Encontrados 4](#_Toc200295092)

[6 Conclusão 5](#_Toc200295093)

[7 Anexos 6](#_Toc200295094)

# Lista de Figuras

[Figura 1 - Use Cases 10](#_Toc170868010)

[Figura 2 - Diagrama de atividades do *Login* 11](#_Toc170868011)

[Figura 3 - Diagrama de atividade de Registo 12](#_Toc170868012)

# Introdução

O presente relatório descreve o desenvolvimento de um dashboard interativo para o grupo de mercearias **TemTudo**, com o objetivo de apoiar a tomada de decisões estratégicas através da **visualização de dados de clientes e vendas**. Este trabalho foi desenvolvido no âmbito da unidade curricular de Visualização de Dados, integrada no Mestrado em Data Science da Universidade Lusófona.

A construção do dashboard teve como base um conjunto de dados, que regista todas as transações realizadas no último ano de 2024, contendo informação detalhada sobre produtos, clientes, descontos e lojas. A análise exploratória visou responder às principais preocupações do(a) manager da TemTudo, nomeadamente:

* **Compreender o perfil dos clientes**,
* **Identificar padrões de compra**,
* **Avaliar o impacto das campanhas promocionais**, com especial foco nos **Promo Days** realizados exclusivamente na **Loja 1**.

Para atingir estes objetivos, foi utilizado o **Streamlit** para o desenvolvimento da interface interativa, em conjunto com bibliotecas como **Pandas** (manipulação e limpeza de dados) e **Plotly Express** (criação de gráficos dinâmicos e informativos). As escolhas gráficas e funcionais foram orientadas pelas perguntas definidas em equipa, tendo em conta as necessidades de negócio apresentadas e as limitações dos dados disponíveis.

Este relatório detalha as questões orientadoras do dashboard, os gráficos utilizados, as decisões tomadas ao longo do processo, assim como os desafios encontrados e as adaptações feitas em relação à proposta inicial (fase 1).

# Planeamento e Fundamentação das Visualizações

## Gráficos Propostos na Fase de Planeamento (Fase 1)

Durante a fase inicial do projeto, foi realizado um exercício de antecipação das questões que poderiam ser respondidas com base no conjunto de dados disponível. Essa etapa teve como objetivo orientar a construção de um dashboard funcional, coerente e relevante do ponto de vista analítico.

As perguntas definidas foram organizadas em duas grandes áreas de análise:

**A. Perfil do Cliente**

Estas visualizações procuraram caracterizar os clientes da empresa:

* Qual o poder de compra por faixa etária?
* Clientes com maior agregado familiar tendem a gastar mais?
* Gasto médio dos clientes com pelo menos uma criança?
* Quais as faixas etárias que mais gastam?
* A cidade do cliente influencia o total gasto?
* Existe correlação entre rendimento, presença de filhos e valor gasto?
* Salário vs Preço Médio Gasto?

O objetivo destas análises foi permitir à empresa compreender melhor o seu público-alvo, identificar segmentos mais lucrativos e adaptar campanhas de marketing e fidelização de forma mais eficiente.

**B. Análise do Padrão de Compras**

Aqui analisou-se o comportamento de compra dos clientes ao longo do tempo e em diferentes contextos promocionais e logísticos:

* Quais os produtos onde se aplicam mais cupões?
* Quais os dias da semana com mais faturações?
* Quais os produtos/departamentos mais vendidos?
* Qual a loja que mais fatura?
* Qual a sexta-feira que oferece maiores descontos aos clientes?
* Qual a sexta-feira com maior volume de faturação?
* Quais os departamentos mais frequentados pelos clientes?
* Como evoluem as vendas ao longo do tempo por loja?
* Volume de compras na loja 1 com vs sem promoção?
* A hora de compra influencia o volume de vendas?

Estas perguntas funcionaram como guia para a construção das visualizações, sendo que algumas se mantiveram tal como foram concebidas inicialmente, enquanto outras foram adaptadas ou refinadas após a fase de exploração dos dados. A escolha final de gráficos teve também em consideração a clareza da visualização, a disponibilidade e qualidade dos dados, e a capacidade de resposta do dashboard em tempo real.

## Estratégia de Visualização

Durante o desenvolvimento do dashboard, optámos por uma estrutura dividida entre **Perfil do Cliente** e **Análise do Padrão de Compras**, com o objetivo de garantir clareza na navegação e uma segmentação lógica da informação. Inicialmente, estas secções foram desenvolvidas em páginas distintas, mas à medida que o projeto evoluiu, decidimos integrá-las num único ficheiro principal (*main.py*), tornando o dashboard mais dinâmico e fácil de utilizar.

Durante a pesquisa e análise de exemplos de dashboards interativos, observámos que muitos seguem uma lógica de apresentação simultânea de múltiplos gráficos, acompanhados de filtros laterais ou superiores. Apesar de reconhecermos o valor dessa abordagem pela riqueza informativa e interatividade, optámos por não a seguir devido a constrangimentos de tempo e à complexidade adicional que implicaria. Para garantir que o nosso dashboard permanecia claro, funcional e compreensível, decidimos simplificar a apresentação, mostrando um gráfico de cada vez com filtros específicos e adaptados, de forma a não comprometer a qualidade da visualização e a experiência do utilizador. Os filtros também se ajustam dinamicamente conforme o gráfico selecionado, promovendo uma experiência mais fluida e intuitiva.

## Gráficos Selecionados e Justificação das Escolhas

Relativamente aos gráficos selecionados, tentámos escolher aqueles que melhor respondiam às perguntas definidas e que ofereciam uma visualização clara, objetiva e relevante para os utilizadores do dashboard.

**A. Perfil do Cliente**

Para caracterizar o público-alvo da empresa, optámos pelos seguintes gráficos:

* **client\_age\_hist** (Histograma da idade dos clientes): Permite visualizar a distribuição etária dos clientes, facilitando a identificação de segmentos com maior representatividade.
* **client\_fam\_num\_bar** (Gráfico de barras do número de elementos no agregado familiar): Ajuda a perceber o impacto do tamanho do agregado familiar no comportamento de compra.
* **client\_with\_kids\_or\_not\_pie** (Gráfico de pizza sobre presença de filhos): Representa a proporção de clientes com e sem filhos, elemento relevante para segmentação de mercado.
* **client\_salary\_hist** (Histograma do salário dos clientes): Fundamental para analisar o poder de compra e a relação entre rendimento e consumo.
* **client\_num\_per\_city\_bar** (Gráfico de barras do número de clientes por cidade): Permite identificar as cidades com maior concentração de clientes.

**B. Análise do Padrão de Compras**

Para compreender o comportamento de compra e a evolução das vendas, foram selecionados os seguintes gráficos:

* **depart\_sales\_num\_bar** (Gráfico de barras das vendas por departamento): Facilita a identificação dos departamentos mais vendidos.
* **sales\_evolution\_by\_store** (Evolução das vendas por loja): Apresenta a dinâmica das vendas ao longo do tempo para cada ponto de venda, crucial para análise temporal e sazonal.
* **top\_selling\_store** (Loja com maiores vendas): Destaca a loja com melhor performance comercial.
* **product\_share\_by\_category** (Distribuição das vendas por categoria de produto): Ajuda a entender a contribuição de cada categoria para o total de vendas.
* **vendas\_boxplot\_promocao** (Boxplot das vendas em função da promoção): Permite analisar o impacto das promoções no volume de vendas, evidenciando diferenças entre compras com e sem desconto.

A escolha destes gráficos justifica-se pela sua capacidade de fornecer insights estratégicos sobre o comportamento dos consumidores e o desempenho comercial, enquanto são visuais simples e de fácil interpretação.

**Exclusões da Fase 1**

Embora na fase inicial do planeamento tenham sido propostas várias visualizações, algumas não foram incluídas na versão final do dashboard por várias razões, nomeadamente a relevância analítica, o tempo disponível para implementação e a complexidade técnica envolvida.

Por exemplo, optámos por não incluir o gráfico do **gasto médio dos clientes com pelo menos uma criança**, pois considerámos que essa informação, apesar de interessante, não acrescentaria valor significativo à análise geral do dashboard.

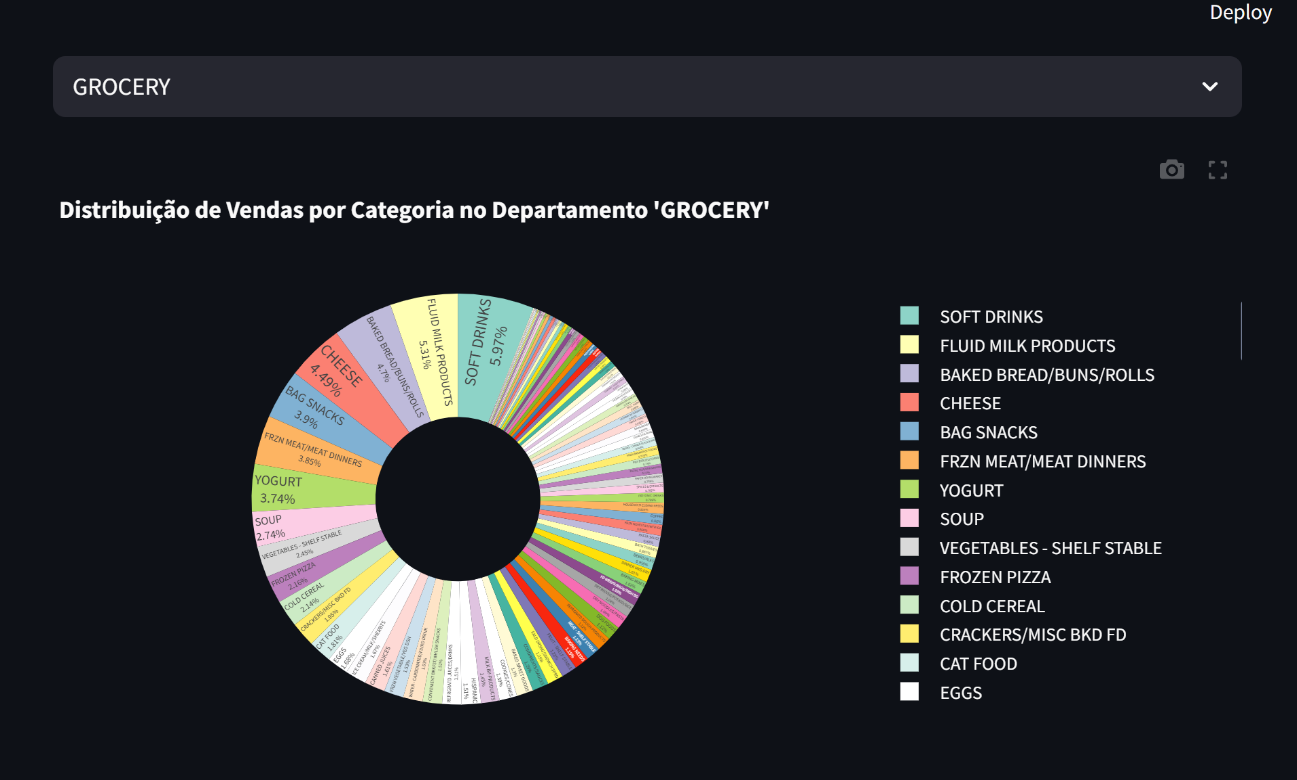
Também decidimos não implementar as análises relacionadas com **os dias da semana com mais faturações**, **a sexta-feira com maiores descontos** e **a sexta-feira com maior volume de faturação**, bem como a influência da **hora da compra no volume de vendas**. Estas visualizações implicariam manipulações mais complexas dos dados temporais, como a extração de dias da semana específicos e o tratamento detalhado dos timestamps, tarefas que, dado o tempo limitado disponível para o projeto, não se justificavam face ao impacto esperado na utilidade do dashboard.

Desta forma, priorizámos a criação de gráficos mais diretos e de fácil interpretação, garantindo uma boa experiência do utilizador e a qualidade das análises apresentadas.

**Considerações adicionais**

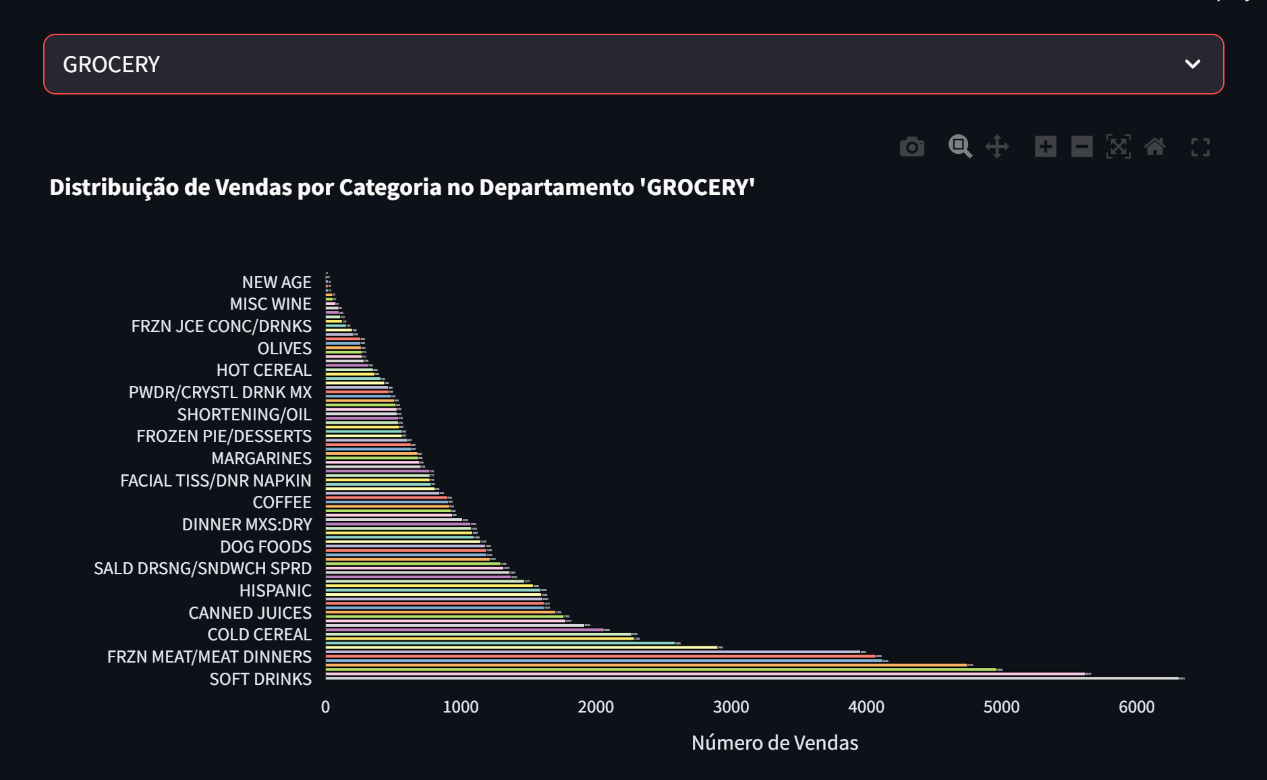
Reconhecemos que, apesar do esforço dedicado à construção do dashboard, existe uma limitação importante na análise apresentada. Durante o desenvolvimento, focámo-nos maioritariamente no total de produtos vendidos, sem explorar suficientemente o valor gasto pelos clientes. Esta abordagem não considera que os produtos têm preços variados, o que pode afetar a precisão da avaliação do comportamento de compra.

Esta situação resultou sobretudo da gestão do tempo, uma vez que dedicámos grande parte do período disponível à construção e otimização das visualizações selecionadas. Só nos lembrámos desta oportunidade de análise mais aprofundada do valor gasto numa fase avançada do projeto, não tendo sido possível implementá-la devido à falta de tempo.

 Futuramente, a exploração mais detalhada do valor gasto por cliente, ajustado ao preço dos produtos, poderá trazer insights mais ricos e relevantes para a empresa.

Um dos gráficos que considerámos incluir foi a proporção de vendas por categoria e por departamento. Contudo, como ilustrado na imagem abaixo, existem tantas categorias dentro dos departamentos que a visualização ficaria demasiado complexa e difícil de interpretar. Por essa razão, decidimos excluir este gráfico, pois acreditámos que não acrescentaria valor significativo ao dashboard.

Em alternativa, optámos por criar um gráfico de barras horizontais que compara as categorias dentro de cada departamento de forma mais clara e simples, facilitando a análise e a compreensão dos dados.

Como ilustrado na imagem acima, existem tantas categorias dentro dos departamentos que a visualização ficaria demasiado complexa e difícil de interpretar. Este gráfico em particular, que mostra a distribuição de vendas no departamento 'GROCERY', reforçou essa conclusão: a grande quantidade de categorias com nomes extensos e a sobreposição de barras tornaram a leitura confusa. Além disso, muitas categorias tinham volumes de vendas muito baixos, o que dispersava a atenção das mais relevantes. Por essa razão, decidimos excluir este gráfico e simplificar a abordagem, restringindo a análise apenas ao top 10 de categorias com mais vendas por departamento. Desta forma, garantimos uma visualização mais clara e focada nos dados que realmente impactam a análise.

# Problemas e Desafios Encontrados

Durante o desenvolvimento do dashboard, surgiram diversos desafios, tanto a nível técnico como ao nível da organização e apresentação da informação.

Um dos primeiros obstáculos prendeu-se com a presença de registos duplicados no dataset. Por exemplo, o mesmo cliente podia surgir várias vezes, o que comprometia a fidelidade de gráficos como o histograma da distribuição da idade dos clientes. Isto era algo que passaria facilmente despercebido, mas posteriormente, ao analisarmos o número de clientes e compararmos com os valores dos gráficos, apercebemo-nos disso. Foi necessário aplicar alguns métodos sobre a coluna client\_id, como o método .nunique, e adotar uma lógica semelhante em outros momentos para garantir que a análise fosse feita com base em dados únicos e representativos.

A construção do gráfico circular para representar a percentagem de clientes com e sem filhos revelou-se inicialmente mais complexa do que o necessário. A coluna respetiva continha valores booleanos em formato não numérico, o que dificultava a agregação e visualização direta dos dados. A solução acabou por ser bastante simples: utilizar o método value\_counts(normalize=True), mas a dificuldade inicial esteve mais relacionada com o excesso de complexidade nas abordagens testadas.

A implementação de filtros interativos por cidade ou loja, apesar de parecer inicialmente intimidante, foi grandemente facilitada pelas funcionalidades intuitivas do Streamlit. A principal aprendizagem foi perceber que a lógica de filtragem precisava de ser bem planeada para não afetar gráficos que deveriam manter uma visão global, como o número total de clientes por cidade.

No que toca à visualização, houve desafios específicos em gráficos com elevado número de categorias distintas, como o gráfico de barras com as categorias de produto mais vendidas. A solução passou por restringir a visualização às 10 categorias com maior número de vendas e ajustar os rótulos dos eixos, incluindo a rotação dos mesmos, para manter a legibilidade e evitar sobrecarga visual.

Por fim, uma das maiores dificuldades prendeu-se com a organização visual e estrutural do dashboard. A decisão entre mostrar muitos gráficos num único ecrã ou construir um ecrã com vários gráficos dinâmicos, dependentes de filtros, exigiu bastante reflexão. Além disso, a personalização do layout com CSS inline foi necessária para melhorar a estética dos botões e seções, o que implicou resolver certas incompatibilidades das bibliotecas usadas.

# Conclusão

O desenvolvimento do dashboard para o grupo TemTudo permitiu consolidar uma visão estruturada e funcional sobre o perfil dos clientes e os padrões de compra registados durante 2024. Através da combinação de ferramentas como Streamlit, Pandas e Plotly Express, conseguimos construir uma solução interativa que responde de forma clara e objetiva às principais questões de negócio definidas no início do projeto.

Apesar das limitações encontradas, nomeadamente ao nível do tempo disponível e da complexidade técnica de algumas análises mais avançadas, o dashboard final apresenta-se como um instrumento útil para apoiar a tomada de decisões estratégicas na empresa. A opção por uma estrutura simples, que privilegia a clareza e a facilidade de uso, mostrou-se eficaz para garantir que os utilizadores consigam interpretar facilmente os dados apresentados.

Reconhecemos que algumas análises poderiam ser aprofundadas, em particular a exploração do valor gasto ajustado aos preços dos produtos, que teria enriquecido os insights obtidos. Este aspeto fica como uma recomendação para trabalhos futuros, onde a gestão do tempo e o foco analítico poderão ser mais bem equilibrados.

Por fim, a experiência adquirida no projeto reforçou a importância de um planeamento cuidadoso das visualizações, bem como a necessidade de adaptar constantemente as soluções às restrições reais do projeto, seja em termos de dados, tempo ou objetivos de negócio. Este trabalho representa assim uma base sólida para futuras melhorias e explorações no âmbito da visualização de dados para apoio à decisão.

**Repositório GitHub:**  
<https://github.com/duartecapinha/projeto_visualizacao.git>