

# Relatório Final: Otimizando Estratégias de E-Commerce com Análise de Dados

---

**Nome:** Peterson Alves Pereira

**Data:** 28/03/2025

**Projeto:** Projeto 3 - Análise de Dados para E-commerce

## 1. Introdução

O setor de e-commerce, apesar do rápido crescimento, enfrenta desafios constantes relacionados à retenção de clientes, aumento do valor médio por pedido (ticket médio) e redução de cancelamentos. Para prosperar, as empresas precisam entender profundamente o comportamento de seus consumidores, permitindo a criação de estratégias de marketing mais eficientes e personalizadas.

Este projeto foi desenvolvido para uma loja de e-commerce fictícia com o objetivo de analisar seu histórico de transações, identificar padrões de compra, segmentar clientes de forma eficaz e descobrir fatores que levam a cancelamentos, visando aumentar a lucratividade e a sustentabilidade do negócio.

**Problema de Negócio:** Como o e-commerce pode identificar padrões de compra, segmentar clientes de forma eficiente e reduzir cancelamentos para aumentar a lucratividade?

### Objetivos do Projeto:

- Compreender a dinâmica do negócio e os desafios do e-commerce.
- Analisar o comportamento de compra dos clientes a partir dos dados de transações.
- Identificar padrões de vendas, sazonalidade e produtos populares.
- Analisar os fatores associados aos cancelamentos de pedidos.

- Segmentar a base de clientes utilizando a metodologia RFM (Recência, Frequência, Valor Monetário).
- Gerar insights acionáveis e recomendações estratégicas para a empresa.

## 2. Metodologia

O projeto seguiu as etapas padrão de um processo de análise de dados:

1. **Compreensão do Negócio e dos Dados:** Entendimento dos objetivos e exploração inicial do dataset fornecido ( `data.csv` ).
2. **Limpeza e Preparação dos Dados:** Tratamento de valores ausentes, duplicados, inconsistências e criação de novas variáveis relevantes ( `TotalPrice` , `IsCancelled` , componentes de data).
3. **Análise Exploratória de Dados (EDA):** Investigação de padrões temporais, geográficos, de produtos, cancelamentos e associações de compra através de visualizações e estatísticas descritivas.
4. **Segmentação de Clientes:** Aplicação da técnica RFM para agrupar clientes com base em seu comportamento de compra.
5. **Geração de Insights e Recomendações:** Consolidação dos achados e formulação de estratégias acionáveis.

## 3. Descrição do Dataset e Desafios Iniciais

O dataset fornecido ( `data.csv` ) continha informações detalhadas sobre as transações de e-commerce, incluindo:

- `InvoiceNo` : Número da fatura (identificador da transação).
- `StockCode` : Código do produto.
- `Description` : Descrição do produto.
- `Quantity` : Quantidade de itens comprados/cancelados.
- `InvoiceDate` : Data e hora da transação.
- `UnitPrice` : Preço unitário do produto.
- `CustomerID` : Identificador único do cliente.
- `Country` : País onde a compra foi realizada.

### Desafios Iniciais:

- **Valores Ausentes:** Uma quantidade significativa de valores ausentes na coluna `CustomerID` (cerca de 25%), crucial para a segmentação, e alguns na coluna `Description`.
- **Duplicatas:** Presença de linhas duplicadas que precisavam ser removidas.
- **Tipos de Dados:** Necessidade de converter `InvoiceDate` para formato datetime e `CustomerID` para um tipo adequado após tratar os nulos.
- **Valores Inconsistentes:** Presença de `Quantity` negativa (indicando cancelamentos) e `UnitPrice` igual a zero.

#### 4. Limpeza e Preparação dos Dados

Para garantir a qualidade e a confiabilidade da análise, as seguintes etapas foram executadas:

- Remoção das linhas com `CustomerID` ausente, focando a análise nos clientes identificados.
- Remoção das poucas linhas restantes com `Description` ausente.
- Remoção das linhas duplicadas.
- Conversão da coluna `InvoiceDate` para o formato `datetime`.
- Conversão da coluna `CustomerID` para o tipo `string`.
- Criação da coluna `IsCancelled` (booleana) para identificar transações de cancelamento (baseado em `InvoiceNo` começando com 'C' e `Quantity` negativa).
- Remoção de transações com `UnitPrice` igual a zero, pois não representam vendas monetárias válidas.
- Criação da coluna `TotalPrice` ( $Quantity * UnitPrice$ ).
- Extração de componentes de data (`Year`, `Month`, `Day`, `Hour`, `Weekday`, `MonthYear`) da `InvoiceDate`.

Após a limpeza, o dataset final continha `401.564` linhas de transação (incluindo cancelamentos) e 16 colunas, pronto para análise.

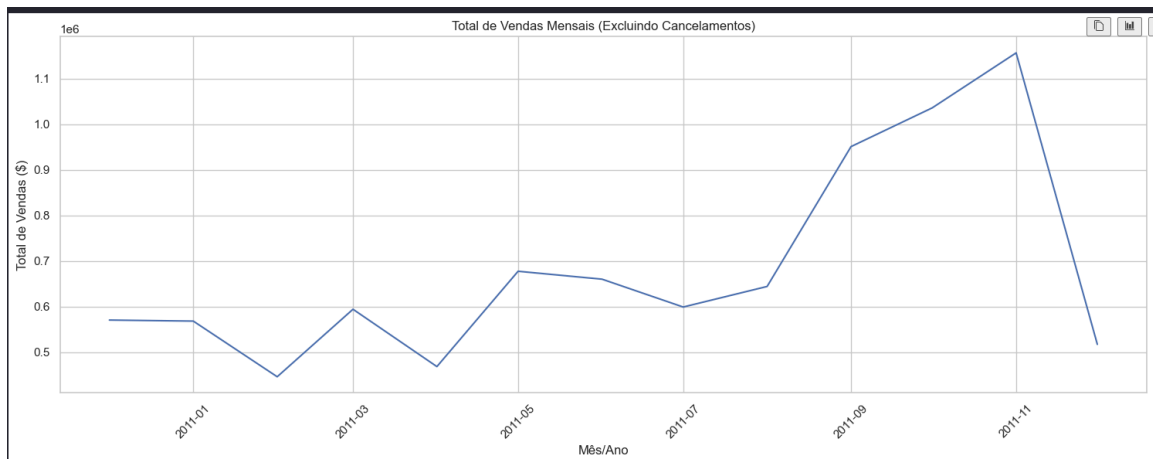
#### 5. Análise Exploratória de Dados (EDA)

##### a) Análise Temporal:

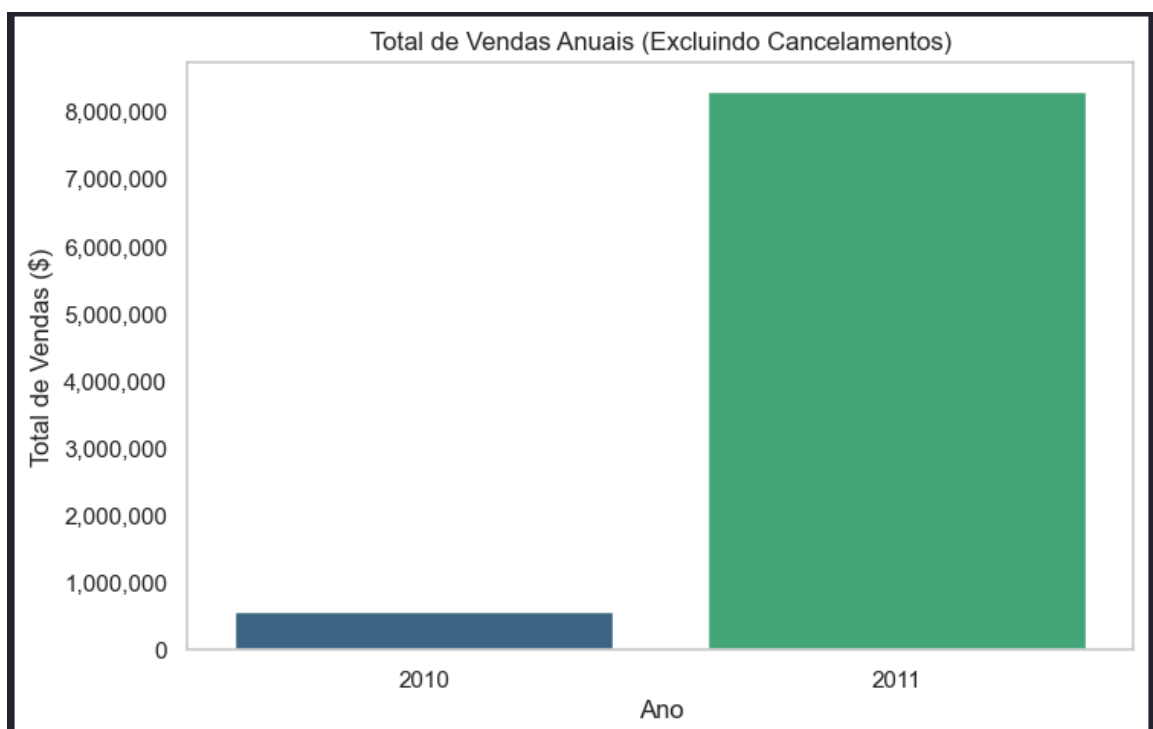
- **Tendência:** Observou-se uma tendência geral de crescimento nas vendas totais e no número de pedidos ao longo de 2011.

- **Sazonalidade:** Há uma forte sazonalidade, com picos de vendas expressivos em Outubro e Novembro, indicando a importância do período de fim de ano. Os meses de Janeiro e Fevereiro apresentaram menor volume de vendas.

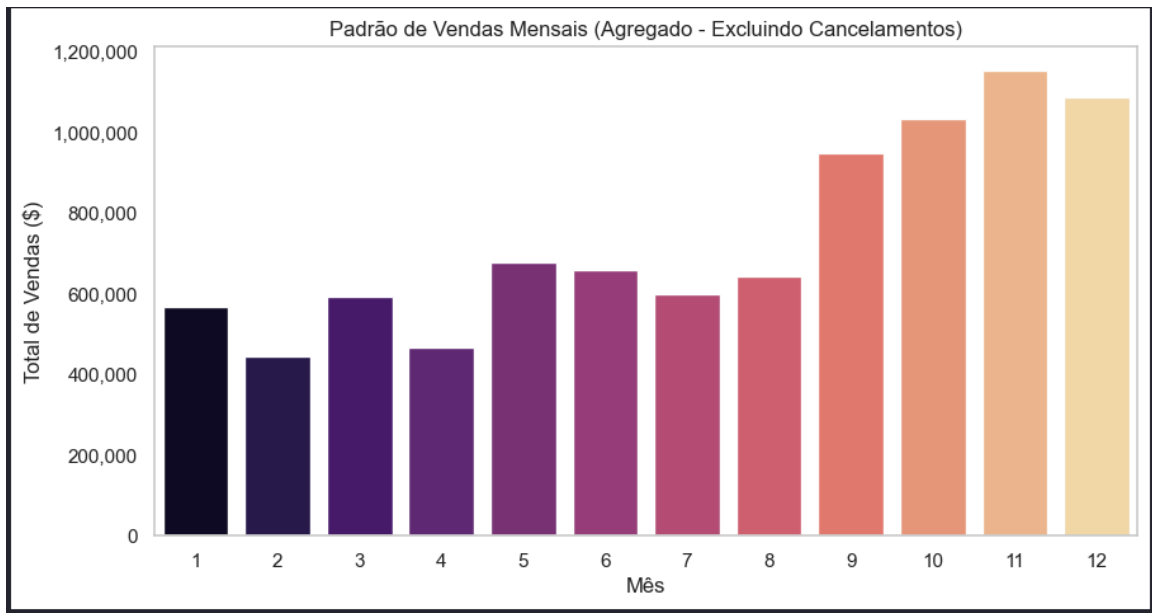
- **Total de Vendas Mensais**



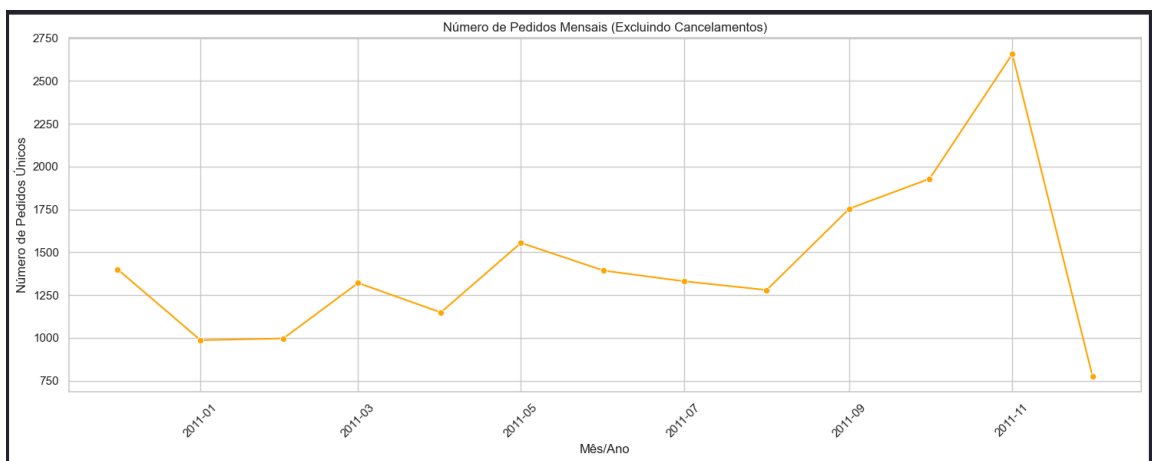
- **Totais de vendas anuais**



- **Padrão de Vendas Mensais Agregado**



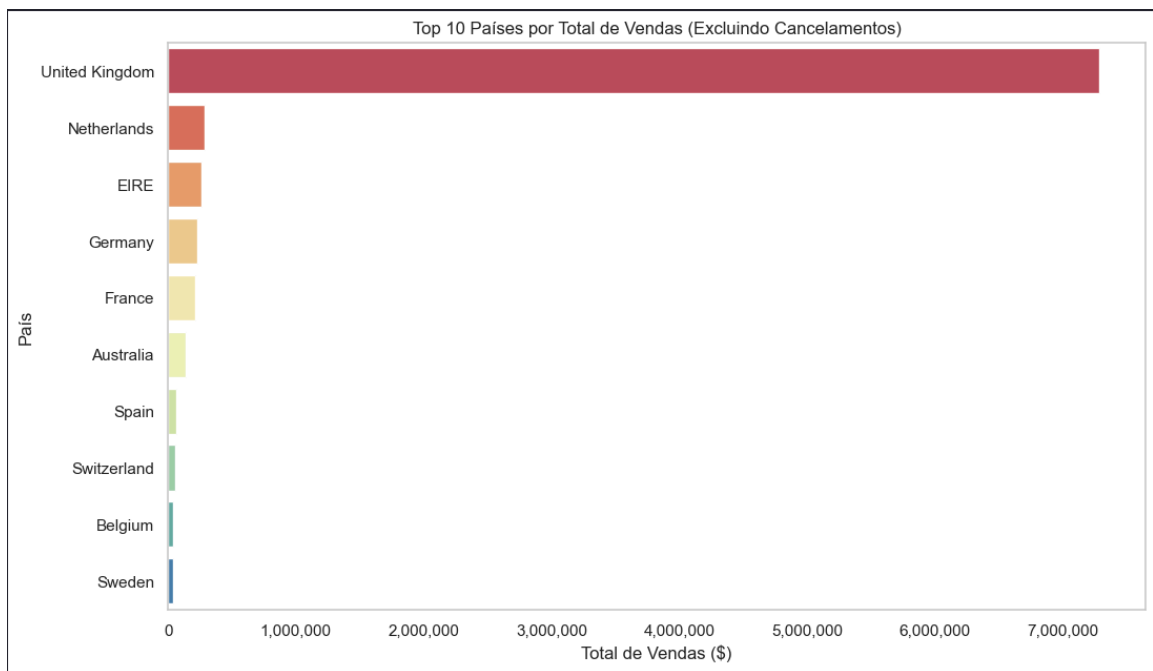
- **Número de Pedidos Mensais**



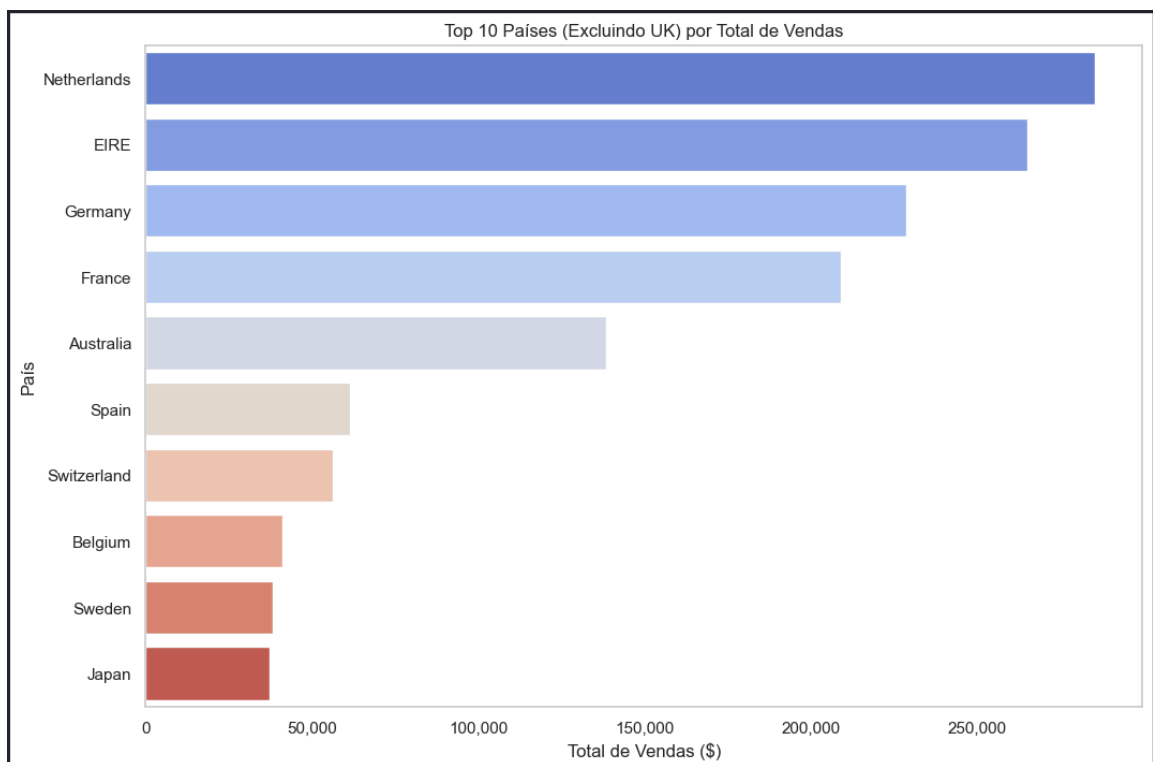
- **Observação:** A análise temporal é impactada pela disponibilidade parcial dos dados para Dezembro de 2010 e Dezembro de 2011.

**b) Análise Geográfica:**

- **Concentração:** O Reino Unido (United Kingdom) é, de forma esmagadora, o principal mercado, respondendo pela grande maioria das vendas.
- **Mercados Secundários:** Países como Holanda (Netherlands), Irlanda (EIRE), Alemanha (Germany) e França (France) representam os mercados internacionais mais significativos, mas com volumes muito inferiores ao UK.
- **Top 10 Países por Total de Vendas**

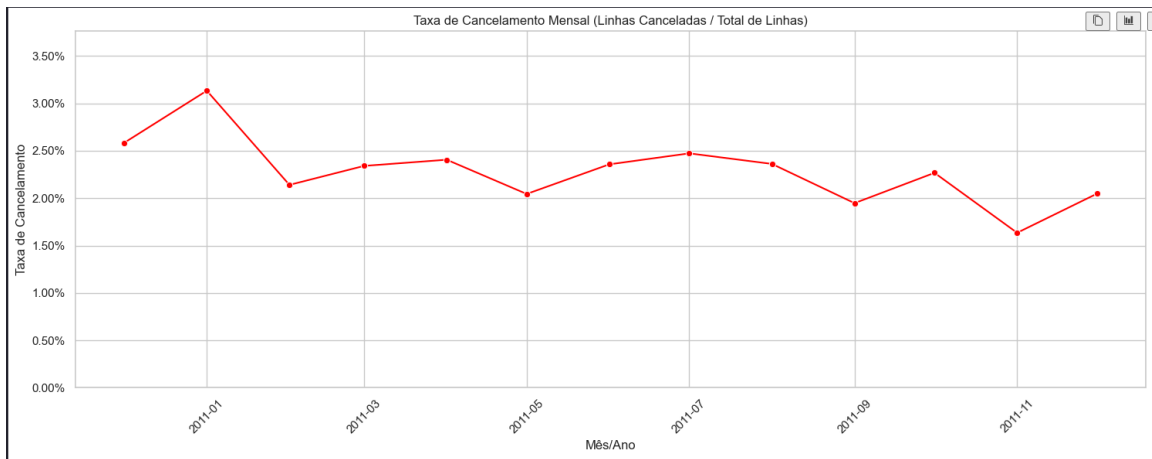


- **Top 10 Países (Excluindo UK)**

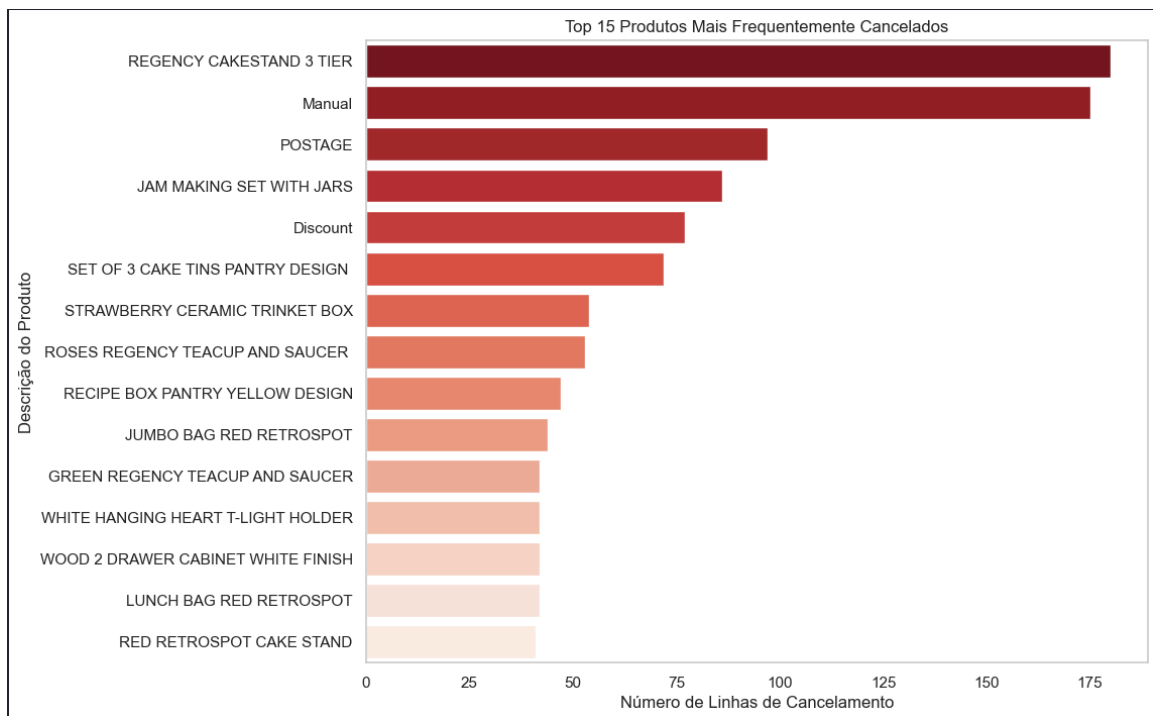


**c) Análise de Cancelamentos:**

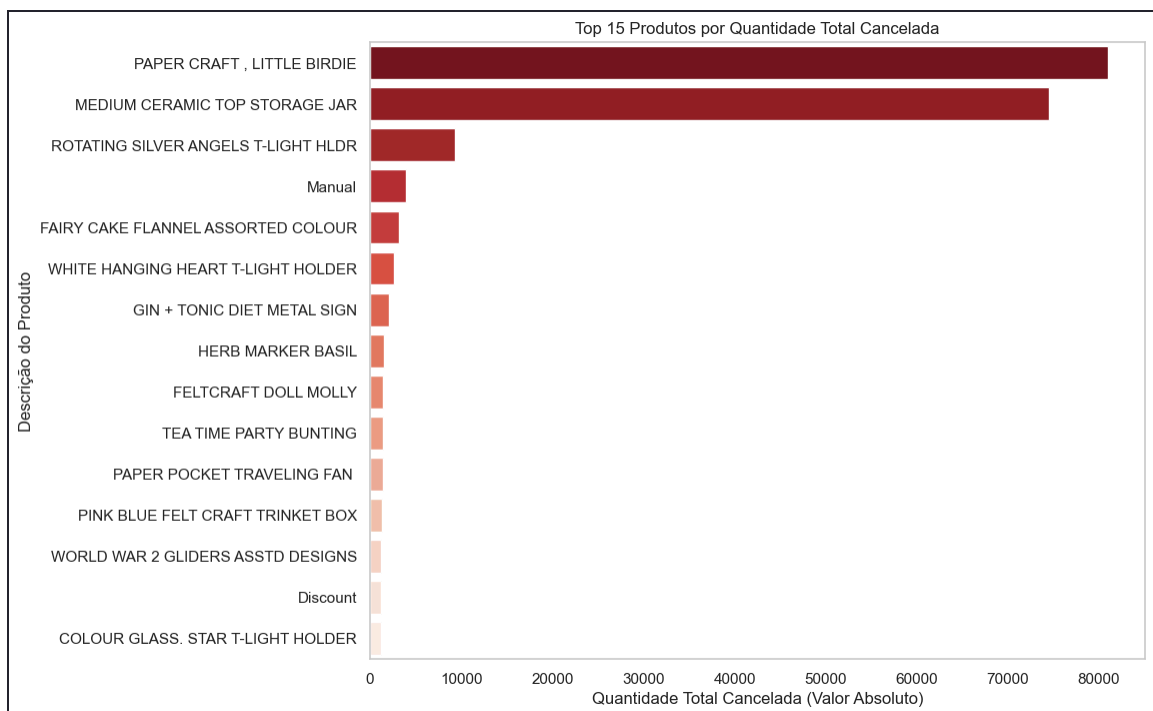
- **Taxa:** A taxa de linhas de transação correspondentes a cancelamentos flutuou, na maioria dos meses, entre 2% e 3%.
  - **Taxa de Cancelamento Mensal**



- **Itens Operacionais:** Uma parcela relevante dos cancelamentos (em frequência e valor) está associada a códigos como "Manual", "Postage", "Discount", sugerindo ajustes manuais ou processos operacionais que precisam ser investigados.
- **Produtos Problemáticos:**
  - "PAPER CRAFT , LITTLE BIRDIE": Apresentou um volume e valor de cancelamento extremamente altos (outlier), necessitando investigação urgente.
  - Outros produtos como "REGENCY CAKESTAND 3 TIER" e "MEDIUM CERAMIC TOP STORAGE JAR" também se destacaram por cancelamentos frequentes ou de alto volume/valor.
  - **Top 15 Produtos Mais Frequentemente Cancelados**

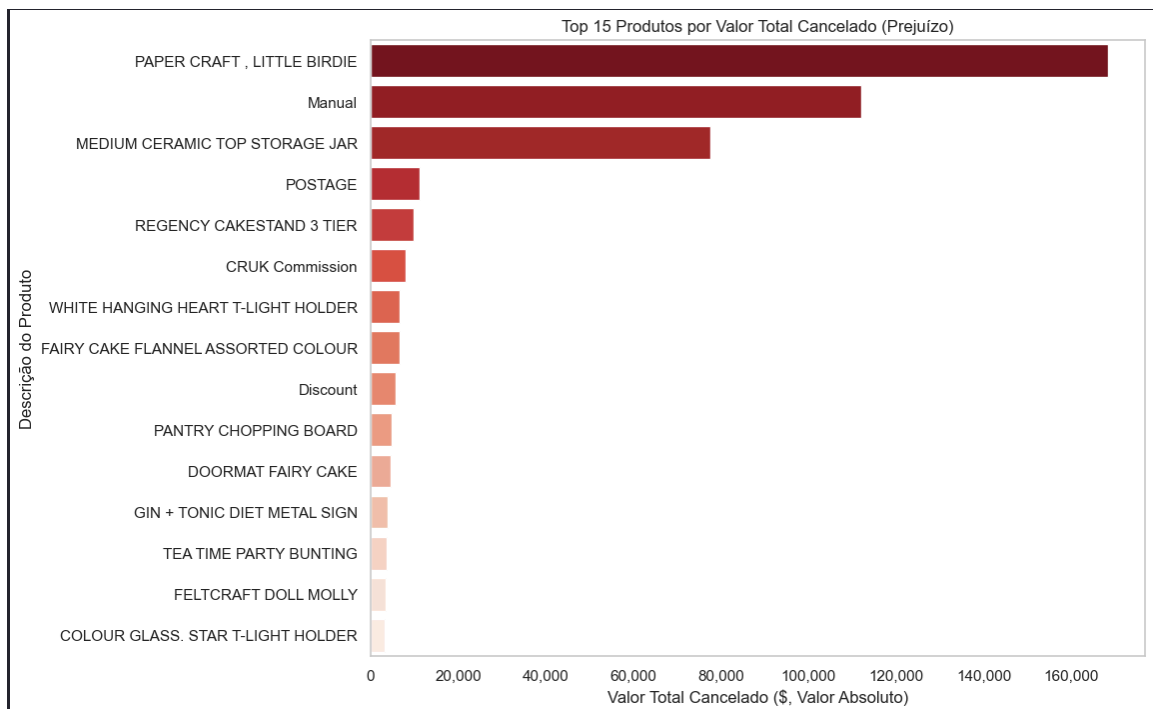


### ○ Top 15 Produtos por Quantidade Total Cancelada



### ○ Top 15 Produtos por Valor Total Cancelado (Prejuízo)





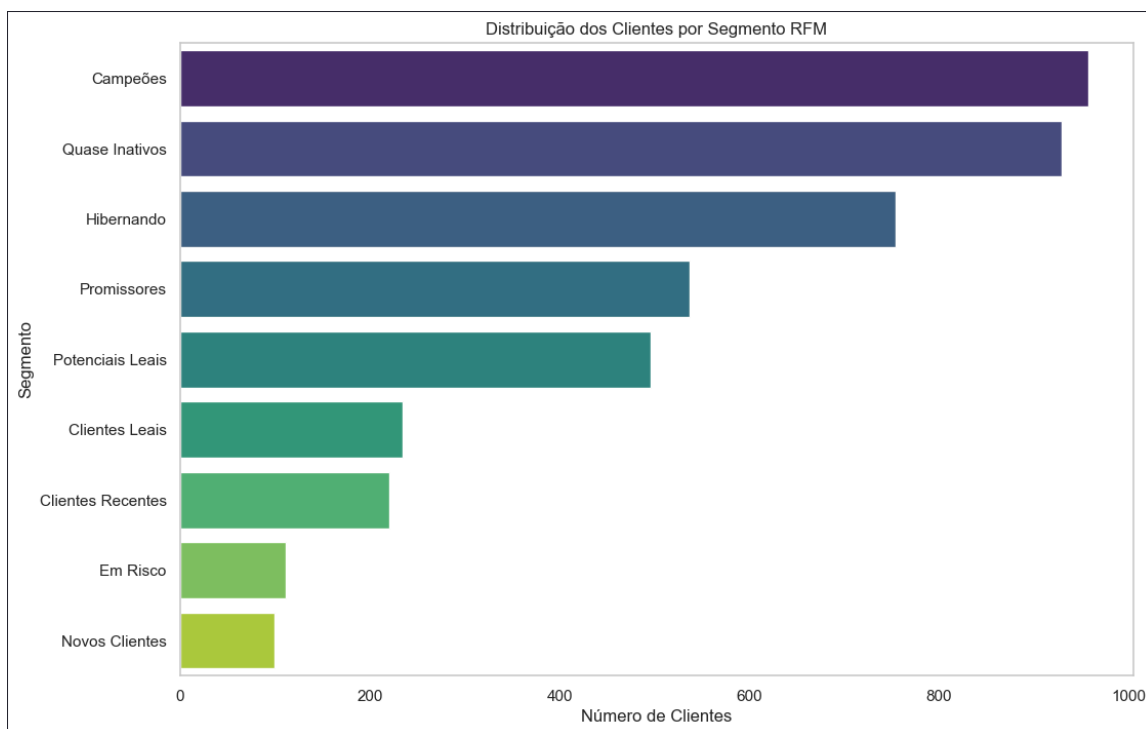
#### d) Análise de Cesta de Compras (Market Basket Analysis):

- Foram identificadas regras de associação fortes entre determinados produtos, especialmente no mercado do Reino Unido.
- Associações Notáveis:** Conjuntos como ( 22697 , 22698 , 22699 ) e pares como ( 23300 , 23301 ) e ( 22726 , 22727 ) são comprados juntos com uma frequência muito acima do esperado pelo acaso (alto **lift** e **confidence** ). Isso indica oportunidades claras para estratégias de cross-selling e bundling.

#### 6. Segmentação de Clientes (RFM)

Utilizando a metodologia RFM (Recência, Frequência, Valor Monetário), os clientes foram agrupados em segmentos distintos com base em seu comportamento de compra:

- Distribuição:** Os maiores segmentos identificados foram "Campeões" (clientes mais valiosos e engajados), "Quase Inativos" e "Hibernando" (clientes que não compram há um tempo considerável). Isso destaca tanto um núcleo forte de clientes quanto um desafio significativo de retenção.
  - Distribuição dos Clientes por Segmento RFM**



- **Características dos Segmentos:** Cada segmento demonstrou características médias de Recência, Frequência e Valor Monetário consistentes com sua definição, validando a segmentação.

Análise das Médias de R, F, M e Contagem por Segmento:

	Recency_Mean	Frequency_Mean	MonetaryValue_Mean	Nu mClientes
Segment				
Campeões	12.8	11.1	6051.9	957
Quase Inativos	110.2	2.6	875.2	929
Hibernando	273.7	1.2	454.6	754
Promissores	53.4	1.6	606.3	537
Potenciais Leais	16.2	3.0	1077.7	496
Clientes Leais	50.4	6.4	3061.3	235
Clientes Recentes	23.7	1.2	435.5	220
Em Risco	233.9	3.5	1937.1	111
Novos Clientes	7.1	1.3	499.1	99

## 7. Principais Insights Consolidados

1. **Forte Sazonalidade:** O negócio é altamente dependente das vendas de fim de ano (Q4).

2. **Dependência do UK:** O mercado doméstico é vital, mas há potencial nos mercados europeus secundários.
3. **Oportunidades de Venda Cruzada:** Existem produtos com fortíssima associação de compra conjunta.
4. **Problemas Operacionais em Cancelamentos:** Itens não-produto ("Manual", etc.) representam uma parcela significativa dos cancelamentos e precisam ser investigados.
5. **Produtos de Alto Cancelamento:** Produtos específicos ("PAPER CRAFT", "CAKESTAND", "STORAGE JAR") necessitam de atenção quanto à qualidade, descrição ou embalagem.
6. **Base de Clientes Polarizada:** Um grande grupo de "Campeões" e um grupo ainda maior de clientes inativos ("Quase Inativos", "Hibernando"), indicando a necessidade de estratégias focadas tanto em nutrição quanto em reativação/retenção.

## 8. Recomendações Estratégicas

Com base nos insights gerados, recomendamos as seguintes ações:

- **Marketing e Promoções:**
  - Intensificar campanhas no Q4 e criar ações específicas para a baixa temporada (Q1).
  - Implementar campanhas de marketing direcionadas para cada segmento RFM (nutrição para Campeões, reativação para Inativos, incentivos para Potenciais Leais, etc.).
- **Vendas (Cross-sell/Up-sell):**
  - Identificar os produtos exatos das associações fortes ( 22697 , etc.).
  - Criar ofertas de pacotes (bundles) para esses itens.
  - Implementar recomendações de "comprados juntos" nas páginas de produto e carrinho.
- **Operações (Cancelamentos):**
  - Auditar e otimizar os processos que geram cancelamentos "Manuais", "Postage", etc.
  - Investigar urgentemente o caso do "PAPER CRAFT , LITTLE BIRDIE".

- Revisar a qualidade, descrição e embalagem dos produtos com alta taxa de cancelamento.
- **Retenção e Fidelização:**
  - Desenvolver programas de fidelidade focados nos "Campeões" e "Clientes Leais".
  - Monitorar a Recência e acionar campanhas de retenção antes que clientes valiosos se tornem inativos.
  - Gerenciar o esforço de reativação, focando nos segmentos com maior potencial de retorno.
- **Gestão de Produto e Estoque:**
  - Garantir estoque adequado dos produtos populares e associados, especialmente no Q4.
- **Estratégia Geográfica:**
  - Manter o foco no UK e desenvolver planos de crescimento direcionados para Holanda, Irlanda, Alemanha e França.

## 9. Conclusão

Esta análise de dados forneceu uma visão detalhada do comportamento dos clientes e das operações do e-commerce. Os insights gerados sobre padrões de compra, sazonalidade, cancelamentos e a segmentação RFM oferecem uma base sólida para a tomada de decisões estratégicas. A implementação das recomendações propostas tem o potencial de otimizar as campanhas de marketing, melhorar a experiência do cliente, reduzir custos associados a cancelamentos, aumentar a fidelização e, conseqüentemente, impulsionar a lucratividade do negócio.

## 10. Próximos Passos Sugeridos

- Implementar um dashboard para monitoramento contínuo dos KPIs e segmentos.
  - Aprofundar a análise de cancelamentos "Manuais".
  - Explorar outras técnicas de segmentação (ex: K-Means baseado em RFM ou outros atributos).
  - Incorporar dados de margem de lucro para uma análise de rentabilidade por produto/cliente.
-

