



E-Commerce: Identificar os perfis dos consumidores



Tratamento dos dados

O tratamento da base de dados foi feita de forma majoritariamente fluída, porém alguns dados estranhos foram encontrados, a base de dados contava com IDs de clientes ausentes e itens duplicados com preços diferentes. Para tratar esses dados optei por verificar qual a porcentagem de IDs de clientes ausentes, e por cobrir uma grande parte dos dados, optei por não remover esses valores nulos e tratá-los como “Clientes Anônimos”. Já para tratar itens duplicados com preços diferentes, minha interpretação foi de acreditar que poderiam se tratar de um mesmo item com tamanhos diferentes, ou descontos aplicados na hora do pagamento, porém por esses casos estarem ligados a IDs de clientes ausentes também há a possibilidade de existir um erro na base de dados fornecida.



Análise exploratória

A análise exploratória dos dados foi feita com a apresentação de gráficos, para situar melhor como os dados se apresentam e interagem entre si. Foram essas as métricas utilizadas:

- Gastos médios por grupo de frequência de compras
- Frequência de compras vs Gastos totais
- Gastos médios por categoria de produto
- Distribuição de gastos por categoria de produto
- Receita média da primeira compra por tipo de cliente
- Distribuição da receita de primeira compra por tipo de cliente



Segmentação RFM e clusterização

A segmentação RFM foi aplicada neste projeto para melhorar o entendimento dos perfis dos usuários com base nos scores de RFM, desse modo foi criado um gráfico com três perfis de baixo, médio e alto valor. Para a clusterização dos dados, foi utilizado o “Método do Cotovelo” para encontrar o número de clusters ideal, e depois calculado com base na segmentação RFM.



Teste de hipótese

O teste de hipóteses escolhido foi o teste de Kruskal-Wallis, pois foram trabalhados 4 clusters no total que seguiam variâncias não homogêneas, excluindo dessa forma a possibilidade de utilizar o teste t-student (por apresentar mais de 2 clusters) e o teste ANOVA (por apresentar variâncias não homogêneas). Por fim foi utilizado o teste de Dunn para realizar comparações pareadas entre os clusters, utilizando um mapa de calor para a representação. A hipótese testada foi da diferença na frequência de compras entre os clusters, cujo resultado apresentou que há uma diferença significativa.



Conclusões

A segmentação permitiu agrupar os clientes em diferentes perfis com base em seus comportamentos de compra. Os principais clusters identificados foram:

- Clientes VIP: Compram com alta frequência e gastam valores elevados.
- Clientes Regulares: Compram frequentemente, mas com ticket médio menor.
- Clientes Ocasionais: Compram poucas vezes e gastam pouco.
- Clientes Inativos: Não compram há muito tempo e gastam pouco quando o fazem.



2. Diferença Significativa na Frequência de Compras

O teste estatístico revelou que há diferenças significativas na frequência de compras entre os clusters, reforçando que os clientes realmente se comportam de maneiras distintas. Isso valida a segmentação e justifica estratégias de marketing personalizadas.



3. Estratégias de Marketing Baseadas nos Clusters

Com base nos perfis identificados, recomenda-se as seguintes estratégias para aumentar a retenção e o faturamento:

- VIP: Programas de fidelidade exclusivos e ofertas personalizadas.
- Regulares: Incentivos para aumentar o ticket médio, como descontos progressivos.
- Ocasionais: Campanhas de reativação, como e-mails com ofertas especiais.
- Inativos: Estratégias agressivas de reengajamento ou foco em novos clientes.