

Análise do Comportamento de Compra do Cliente

1. Visão Geral do Projeto

Este projeto analisa o comportamento de compra do cliente usando dados transacionais de 3.900 compras em várias categorias de produtos. O objetivo é descobrir insights sobre padrões de gastos, segmentos de clientes, preferências de produtos e comportamento de assinatura para orientar decisões estratégicas de negócios.

2. Resumo do Conjunto de Dados

- **Linhas:** 3.900
- **Colunas:** 18
- **Principais Características:**
 - **Demografia do Cliente:** Idade, Gênero, Localização, Status de Assinatura
 - **Detalhes da Compra:** Item Comprado, Categoria, Valor da Compra, Estação, Tamanho, Cor
 - **Comportamento de Compra:** Desconto Aplicado, Código Promocional Usado, Compras Anteriores, Frequência de Compras, Avaliação (Rating), Tipo de Envio.
- **Dados Faltantes:** 37 valores na coluna de Avaliação (Review Rating)

3. Análise Exploratória de Dados usando Python

Começamos com a preparação e limpeza dos dados em Python:

- **Carregamento de Dados:** Importação do conjunto de dados usando pandas.
- **Exploração Inicial:** Uso de df.info() para verificar a estrutura e describe() para estatísticas resumidas.

Estatísticas Descritivas (Resumo dos Dados)

Estatística	Customer ID	Age	Purchase Amount (USD)	Review Rating	Previous Purchases
Contagem	3900.0	3900.0	3900.0	3863.0	3900.0
Média	1950.5	44.07	59.76	3.75	25.35
Desvio Padrão	1125.9	15.21	23.69	0.72	14.45
Mínimo	1.0	18.0	20.0	2.5	1.0
25%	975.75	31.0	39.0	3.1	13.0
50% (Mediana)	1950.5	44.0	60.0	3.8	25.0
75%	2925.25	57.0	81.0	4.4	38.0
Máximo	3900.0	70.0	100.0	5.0	50.0

Dados Categóricos (Top Valores):

- **Gênero:** Masculino (2652)
 - **Item Comprado:** Blusa (171)
 - **Categoria:** Roupas (1737)
 - **Localização:** Montana (96)
 - **Tamanho:** M (1755)
 - **Cor:** Oliva (177)
 - **Estação:** Primavera (999)
 - **Assinatura:** Não (2847)
 - **Tipo de Envio:** Frete Grátis (675)
 - **Método de Pagamento:** PayPal (677)
 - **Frequência:** A cada 3 meses (584)
-
- **Tratamento de Dados Faltantes:** Verificação de valores nulos e imputação dos valores faltantes na coluna *Review Rating* usando a mediana da avaliação de cada categoria de produto.

- **Padronização de Colunas:** Colunas renomeadas para `snake_case` para melhor legibilidade e documentação.
- **Engenharia de Atributos (Feature Engineering):**
 - Criação da coluna `age_group` (faixa etária) agrupando as idades dos clientes.
 - Criação da coluna `purchase_frequency_days` a partir dos dados de compra.
- **Verificação de Consistência dos Dados:** Verificado se `discount_applied` e `promo_code_used` eram redundantes; `promo_code_used` foi descartado.
- **Integração com Banco de Dados:** Conexão do script Python ao PostgreSQL e carregamento do DataFrame limpo no banco de dados para análise SQL.

4. Análise de Dados usando SQL (Transações de Negócios)

Realizamos análises estruturadas no PostgreSQL para responder a perguntas-chave de negócios:

1. Receita por Gênero - Comparação da receita total gerada por clientes masculinos vs. femininos.

	gender text 	revenue numeric 
1	Female	75191
2	Male	157890

2. Usuários de Desconto com Alto Gasto - Identificação de clientes que usaram descontos, mas ainda assim gastaram acima da média de compra.

	customer_id	purchase_amount
	bigint	bigint
1	2	64
2	3	73
3	4	90
4	7	85
5	9	97
6	12	68
7	13	72
8	16	81
9	20	90
10	22	62
11	24	88

Total rows: 839 Query complete 00:00

3. Top 5 Produtos por Avaliação - Produtos com as maiores médias de avaliação.

	item_purchased	Average Product Rating
	text	numeric
1	Gloves	3.86
2	Sandals	3.84
3	Boots	3.82
4	Hat	3.80
5	Skirt	3.78

4. Comparação de Tipo de Envio - Comparação dos valores médios de compra entre envio Standard e Express.

	shipping_type	round
	text	numeric
1	Standard	58.46
2	Express	60.48

5. Assinantes vs. Não Assinantes - Comparação do gasto médio e receita total por status

de assinatura.

	subscription_status	total_customers	avg_spend	total_revenue
1	Yes	1053	59.49	62645.00
2	No	2847	59.87	170436.00

6. Produtos Dependentes de Desconto - Identificação dos 5 produtos com a maior porcentagem de compras com desconto.

	item_purchased	discount_rate
1	Hat	50.00
2	Sneakers	49.66
3	Coat	49.07
4	Sweater	48.17
5	Pants	47.37

7. Segmentação de Clientes - Classificação dos clientes em segmentos "Novos", "Recorrentes" e "Leais" com base no histórico de compras.

	customer_segment text	Number of Customers bigint
1	Loyal	3116
2	New	83
3	Returning	701

8. Top 3 Produtos por Categoria - Listagem dos produtos mais comprados dentro de cada categoria.

	item_rank bigint	category text	item_purchased text	total_orders bigint
1	1	Accessories	Jewelry	171
2	2	Accessories	Sunglasses	161
3	3	Accessories	Belt	161
4	1	Clothing	Blouse	171
5	2	Clothing	Pants	171
6	3	Clothing	Shirt	169
7	1	Footwear	Sandals	160
8	2	Footwear	Shoes	150
9	3	Footwear	Sneakers	145
10	1	Outerwear	Jacket	163
11	2	Outerwear	Coat	161

9. Compradores Recorrentes e Assinaturas - Verificação se clientes com mais de 5 compras são mais propensos a assinar.

	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

10. Receita por Faixa Etária - Cálculo da contribuição da receita total de cada grupo etário.

	age_group	total_revenue
	text	numeric
1	Young Adult	62143
2	Middle-aged	59197
3	Adult	55978
4	Senior	55763

5. Dashboard no Power BI

Finalmente, construímos um dashboard interativo no Power BI para apresentar os insights visualmente.



6. Recomendações de Negócios

- Impulsionar Assinaturas:** Promover benefícios exclusivos para assinantes.
- Programas de Fidelidade do Cliente:** Recompensar compradores recorrentes para mover-los para o segmento "Leal".
- Revisar Política de Descontos:** Equilibrar o aumento de vendas com o controle de margem.
- Posicionamento de Produto:** Destacar produtos bem avaliados e mais vendidos em campanhas.
- Marketing Direcionado:** Focar esforços em grupos etários de alta receita e usuários de envio expresso.