

Shopping Trend Insights

Guilherme Freitas | [GitHub](#) - [Guifreitaas](#)

<https://www.linkedin.com/in/guilhermefreitas-eng/>

Topicos

1. Introdução.
2. Processamento e Análise de Dados.
3. Insights e Conclusão.

1. Introdução

Etapas

- Extração de Dados no Kaggle;
- Criação da tabela no SQL Server;
- SQL Querys ao banco de dados;
- Análise e Storytelling;
- Conclusão;

Fontes

A tabela foi criada no SQL Server com uma versão dos dados fornecidos em [Kaggle](#).

Objetivo

Este estudo tem como principal objetivo analisar os dados contidos na base **shopping_trends** utilizando linguagem **SQL**, através do **SQL Server** para compreender o perfil dos clientes e identificar padrões comportamentais relacionados às compras realizadas. A análise busca explorar aspectos como preferências de itens e categorias. Esses insights podem orientar decisões estratégicas para melhorar a experiência do cliente e otimizar as vendas e marketing

Dados

- Customer ID = Identificador único do cliente no banco de dados.
- Age = Idade do cliente em anos.
- Gender = Gênero do cliente.
- Item Purchased = Item específico comprado pelo cliente.
- Category = Categoria à qual o item comprado pertence.
- Purchase_Amount (USD) = Valor da compra realizada pelo cliente.
- Location = Localização geográfica do cliente.
- Size = Tamanho do produto comprado.
- Color = Cor do item comprado pelo cliente.
- Season = Estação do ano durante a qual a compra foi realizada.
- Review_Rating = Avaliação do cliente sobre o produto ou a experiência de compra.
- Subscription_Status = Status de assinatura do cliente (ativa ou inativa).
- Payment_Method = Método de pagamento utilizado na transação.
- Shipping_Type = Tipo de envio escolhido pelo cliente.
- Discount_Applied = indica se um desconto foi aplicado à compra.
- PromoCode_Used = indica se um código promocional foi utilizado na transação.
- Previous_Purchases = Número de compras anteriores realizadas pelo cliente.
- Preferred_Payment_Method = Método de pagamento preferido do cliente.

- Frequency of Purchases = Frequência com a qual o cliente realiza compras.

2. Processamento e Análise de Dados

Exploração

A primeira fase da análise é a **exploração**. Assim podemos entender melhor os dados que estamos trabalhando com, fazendo perguntas aos dados:

Como visualizar nossa tabela?

Uma consulta simples, que consome menos processamento de dados, pode ser feita usando **SELECT TOP**, seguido do número de linhas que você gostaria de exibir. Isso é útil para obter uma visão geral dos dados da tabela, sem carregar todo o conteúdo.

```
SELECT TOP 10 *
FROM Shopping_Trends;
```

O resultado são as 10 primeiras linhas da tabela, divididas em 2 imagens, mostrando tanto o cabeçalho quanto os dados de cada linha e coluna. Isso permite uma análise preliminar e ajuda a entender os dados disponíveis.

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating	Subscription Status
1	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes
2	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes
3	3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Yes
4	4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Yes
5	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes
6	6	46	Male	Sneakers	Footwear	20	Wyoming	M	White	Summer	2.9	Yes
7	7	63	Male	Shirt	Clothing	85	Montana	M	Gray	Fall	3.2	Yes
8	8	27	Male	Shorts	Clothing	34	Louisiana	L	Charcoal	Winter	3.2	Yes
9	9	26	Male	Coat	Outerwear	97	West Virginia	L	Silver	Summer	2.6	Yes
10	10	57	Male	Handbag	Accessories	31	Missouri	M	Pink	Spring	4.8	Yes

Payment Method	Shipping Type	Discount Applied	Promo Code Used	Previous Purchases	Preferred Payment Method	Frequency of Purchases
Credit Card	Express	Yes	Yes	14	Venmo	Fortnightly
Bank Transfer	Express	Yes	Yes	2	Cash	Fortnightly
Cash	Free Shipping	Yes	Yes	23	Credit Card	Weekly
PayPal	Next Day Air	Yes	Yes	49	PayPal	Weekly
Cash	Free Shipping	Yes	Yes	31	PayPal	Annually
Venmo	Standard	Yes	Yes	14	Venmo	Weekly
Debit Card	Free Shipping	Yes	Yes	49	Cash	Quarterly
Debit Card	Free Shipping	Yes	Yes	19	Credit Card	Weekly
Venmo	Express	Yes	Yes	8	Venmo	Annually
PayPal	2-Day Shipping	Yes	Yes	4	Cash	Quarterly

Que tipo de dados vamos trabalhar?

Customer ID	Float	Review_Rating	Float
Age	Float	Subscription_Status	String
Gender	String	Payment_Method	String
Item Purchased	String	Shipping_Type	String
Category	String	PromoCode_Used	String
Purchase_Amount (USD)	Float	Previous_Purchases	Float
Location	String	Preferred_Payment_Method	String
Size	String	Frequency of Purchases	String
Color	String	Season	String

Quantos clientes fizeram alguma compra em nosso E-commerce temos cadastrado?

```
SELECT COUNT(DISTINCT [Customer ID]) AS 'NumCustomers'  
FROM Shopping_Trends;
```

	NumCustomers
1	3900

Análise de Dados

Agora que entendemos nossos dados, podemos iniciar a análise. Para isso, faremos as seguintes perguntas:

Quem compra mais? Homem ou Mulher? Em qual época que compram mais? Quais as categorias mais vendidas?

Neste tipo de análise, foi utilizada a função de agregação **SUM** para calcular o faturamento total de um grupo específico. Além disso, a função **GROUP BY** foi empregada para organizar os dados por categorias ou grupos determinados, facilitando a interpretação e a geração de insights.

```
SELECT Gender, SUM([Purchase Amount (USD)])  
FROM Shopping_Trends  
GROUP BY Gender;
```

	Gender	SUM_Purchase
1	Male	2684130
2	Female	1232157

```
SELECT Season, SUM([Purchase Amount (USD)]) AS 'SUM_Purchase'  
FROM Shopping_Trends  
GROUP BY Season  
ORDER BY 'SUM_Purchase' DESC;
```

	Season	SUM_Purchase
1	Fall	1007149
2	Spring	986169
3	Winter	985160
4	Summer	937809

```
SELECT Category, SUM([Purchase Amount (USD)]) AS 'SUM_Purchase'  
FROM Shopping_Trends  
GROUP BY Category  
ORDER BY 'SUM_Purchase' DESC;
```

	Category	SUM_Purchase
1	Clothing	1751719
2	Accessories	1246706
3	Footwear	606396
4	Outerwear	311466

Conclusões com base na análise:

- Homens são responsáveis por mais do que o dobro do faturamento em comparação às mulheres uma diferença total de U\$1.451.973.
- O outono se destaca como a estação mais lucrativa, apresentando um pico significativo no consumo.
- Roupas representam o principal motor de vendas, seguidas por acessórios.

Insights e Recomendações:

Foque em estratégias de marketing direcionadas ao público masculino como por exemplo, contato com influenciadores e anúncios direcionados, que domina a maior parte do faturamento.

- Aproveite o aumento de consumo no outono com ações promocionais e campanhas específicas para maximizar o retorno.
- Explore estratégias para impulsionar categorias com menor desempenho, como calçados e roupas de inverno, por meio de promoções, reposicionamento de produtos ou lançamento de novas coleções.

Quais são as localizações que mais vendem?

Para classificar as cidades no topo, foi utilizada a função **RANK** junto com a função de agregação **SUM**.

```
SELECT Location,
       [Revenue by Region],
       [Ranking by Location]
FROM (SELECT Location,
       SUM([Purchase Amount (USD)]) AS 'Revenue by Region',
       RANK() OVER (ORDER BY SUM([Purchase Amount (USD)]) DESC) AS 'Ranking by Location'
FROM Shopping_Trends
GROUP BY Location
) AS RankedData
WHERE [Ranking by Location] <= 10;
```

Depois, para selecionar apenas o **top 10** e em seguida as **últimas posições**, foi criada uma subquery.

Top 10

	Location	Revenue by Region	Ranking by Location
1	Montana	96948	1
2	Illinois	94446	2
3	California	94367	3
4	Idaho	94023	4
5	Nevada	92631	5
6	Alabama	88410	6
7	New York	88392	7
8	North Dakota	87705	8
9	West Virginia	87233	9
10	Nebraska	86968	10

Top 40 - 50

	Location	Revenue by Region	Ranking by Location
1	South Dakota	71186	40
2	Iowa	70951	41
3	Colorado	70872	42
4	New Hampshire	70692	43
5	Wisconsin	70551	44
6	Connecticut	70513	45
7	Rhode Island	64974	46
8	Florida	63790	47
9	New Jersey	63740	48
10	Hawaii	63431	49
11	Kansas	57275	50

Conclusões com base na análise:

- Distribuição geográfica não segue padrões econômicos tradicionais.
- Estados menores ou com menos diversidade econômica, como Kansas ou Havai, podem ter economias mais concentradas em setores específicos, o que pode limitar o crescimento comparado a estados maiores.

Insights e Recomendações:

- Foco nas regiões com altas receitas e potenciais de crescimento: Estados como Illinois, Califórnia, e Nevada são os mais fortes e devem ser áreas de foco para iniciativas mais amplas.
- Desenvolver estratégias específicas para mercados underperforming.

Como está a recorrência do E-commerce?

Para o estudo de recorrência, foi utilizada a função CASE, pois já tínhamos o número de compras de cada cliente na loja. Criamos uma tabela em que, se o cliente comprou entre 1 e 12 vezes, sua recorrência é classificada como baixa. Já as compras entre 38 e 50 vezes são classificadas como muito alta.

```
SELECT [Customer ID], [Previous Purchases],  
  
CASE  
WHEN [Previous Purchases] BETWEEN 1 AND 12 THEN 'Low'  
WHEN [Previous Purchases] BETWEEN 13 AND 25 THEN 'Medium'  
WHEN [Previous Purchases] BETWEEN 26 AND 37 THEN 'High'  
WHEN [Previous Purchases] BETWEEN 38 AND 50 THEN 'Very High'  
END AS 'Purchase_Frequency'  
  
FROM Shopping_trends
```

	Customer ID	Previous Purchases	Purchase_Frequency
1	1	14	Medium
2	2	2	Low
3	3	23	Medium
4	4	49	Very High
5	5	31	High
6	6	14	Medium
7	7	49	Very High
8	8	19	Medium
9	9	8	Low
10	10	4	Low
11	11	26	High
12	12	10	Low
13	13	37	High
14	14	31	High
15	15	34	High
16	16	8	Low
17	17	44	Very High
18	18	36	High

1 -12	Low
13 - 25	Medium
26 - 37	High
38 - 50	Very High

Conclusões com base na análise:

- Podemos observar como os clientes estão distribuídos nas diferentes faixas de compra.
- A maior parte dos clientes se encontra na faixa low, indicando que muitos clientes compram esporadicamente.
- Os clientes que compram entre 38 e 50 vezes representam o topo da pirâmide de clientes leais e fiéis. Este grupo é extremamente valioso para o e-commerce.

Insights e Recomendações:

- Incentivar a recorrência de clientes "Low" com promoções e estratégias de reativação.
- Maximizar o valor de clientes "Medium" e "High" com estratégias de vendas adicionais e programas de fidelidade.
- Manter e recompensar os clientes "Very High", que são seus principais geradores de receita, com benefícios exclusivos e experiências personalizadas.

3. Conclusão

Este é um estudo de caso de análise de dados, no qual foram selecionadas algumas partes específicas para aprofundar a análise. A fonte de referência utilizada encontra-se acima. A seguir, apresentarei um resumo dos principais insights encontrados, com foco nas áreas de maior impacto para o desenvolvimento de estratégias de marketing e vendas.

Resumo das Conclusões e Insights:

- Faturamento por Gênero: Homens geram mais do que o dobro do faturamento em comparação às mulheres.
- Estação do Ano: O outono é a estação mais lucrativa, com pico no consumo.
- Produtos: Roupas são o principal motor de vendas, seguidas por acessórios.
- Distribuição Geográfica: Estados menores, como Kansas e Havaí, têm economias concentradas em setores específicos, sem seguir padrões econômicos tradicionais.
- Fidelização de Clientes: A maioria dos clientes é esporádica ("Low"), enquanto os mais leais (compras frequentes) são altamente valiosos.

Recomendações:

- Foco em Regiões de Alto Potencial: Estados como Illinois, Califórnia e Nevada devem receber atenção especial.
- Marketing Masculino: Direcionar estratégias para o público masculino.
- Promoções no Outono: Aproveitar o pico de consumo da estação com campanhas específicas.
- Melhoria nas Categorias de Menor Desempenho: Impulsionar vendas de calçados e roupas de inverno.
- Fidelização: Engajar clientes esporádicos com promoções, e recompensar os leais com benefícios exclusivos.

Para maximizar o crescimento, é essencial que a empresa capitalize sobre os momentos de pico de consumo, como o outono, e foque em estratégias personalizadas para os clientes mais fiéis, que geram a maior parte da receita. Isso garantirá não apenas o aumento das vendas, mas também a fidelização de uma base de clientes valiosa e crescente.