Shopping Trend Insights

Data Analyst - Guilherme Freitas

Topics

- Introduction.
- Data Processing and Analysis.
- Insights and Conclusion.

1. Introduction

Steps

- Data Extraction from Kaggle.
- Table Creation in SQL Server.
- SQL Queries to the Database.
- Analysis and Storytelling.
- Conclusion.

Sources

The table was created in SQL Server with a version of the data provided on Kaggle.

Objective

The main objective of this study is to analyze the data contained in the **shopping_trends** database using **SQL** language, through **SQL Server**, to understand the customer profile and identify behavioral patterns related to the purchases made. The analysis seeks to explore aspects such as item and category preferences. These insights can guide strategic decisions to enhance the customer experience and optimize sales and marketing.

Data

- Customer ID = Unique identifier for the customer in the database.
- Age = Customer's age in years.
- Gender = Customer's gender.
- Item Purchased = Specific item purchased by the customer.
- Category = Category to which the purchased item belongs.
- Purchase_Amount (USD) = Purchase amount made by the customer.
- Location = Customer's geographical location.
- Size = Size of the product purchased.
- Color = Color of the item purchased by the customer.
- Season = Season of the year during which the purchase was made.
- Review_Rating = Customer's rating of the product or shopping experience.
- Subscription_Status = Customer's subscription status (active or inactive).
- Payment_Method = Payment method used in the transaction.
- Shipping_Type = Shipping method chosen by the customer.
- Discount_Applied = Indicates whether a discount was applied to the purchase.
- PromoCode_Used = Indicates whether a promotional code was used in the transaction.
- Previous_Purchases = Number of previous purchases made by the customer.
- Preferred_Payment_Method = Customer's preferred payment method.
- Frequency of Purchases = Frequency with which the customer makes purchases.

2. Data Processing and Analysis

Exploration

The first phase of the analysis is exploration. This allows us to better understand the data we are working with by asking questions to the data:

How can we visualize our table?

A simple query, which requires less data processing, can be done using **SELECT TOP**, followed by the number of rows you would like to display. This is useful for getting an overview of the table data without loading the entire content.

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

The result is the first 10 rows of the table, split into 2 images, showing both the header and the data of each row and column. This allows for a preliminary analysis and helps to understand the available data.

	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

What type of data will we work with?

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

How many customers have made a purchase in our registered e-commerce?

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

	NumCustomers
1	3900

Data Analysis

Now that we understand our data, we can begin the analysis. To do this, we will ask the following questions:

Who buys more? Men or women? What time of the year do they buy the most? What are the top-selling categories?

In this type of analysis, the **SUM** aggregation function was used to calculate the total revenue of a specific group. Additionally, the **GROUP BY** function was employed to organize the data by categories or defined groups, making interpretation and insight generation easier.

```
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

Conclusions based on the analysis:

- Men are responsible for more than double the revenue compared to women, with a total difference of $1,451,973.
- Autumn stands out as the most profitable season, showing a significant peak in consumption.
- Clothing represents the primary driver of sales, followed by accessories.

Insights and Recommendations:

- Focus on marketing strategies targeted at the male audience, such as influencer collaborations and targeted ads, as they dominate the majority of the revenue.
- Focus on marketing strategies targeted at the male audience, such as influencer collaborations and targeted ads, as they dominate the majority of the revenue.
- Explore strategies to boost underperforming categories, such as footwear and winter clothing, through promotions, product repositioning, or launching new collections.

What are the top-selling locations?

To rank the cities at the top, the **RANK** function was used along with the **SUM** aggregation function. Then, to select only the top 10 and subsequently the lowest positions, a subquery was created.

```
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;
```

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

Concluion based on the analysis:

- The geographical distribution does not follow traditional economic patterns.
- Smaller states or those with less economic diversity, such as Kansas or Hawaii, may have economies more concentrated in specific sectors, which could limit growth compared to larger states.

Insights e Recommendations:

- Focus on regions with high revenues and growth potential: States like Illinois, California, and Nevada are the strongest and should be areas of focus for broader initiatives.
- Develop specific strategies for underperforming markets.

How is the e-commerce recurrence?

For the recurrence study, the **CASE** function was used, as we already had the number of purchases for each customer in the store. We created a table where, if the customer bought between 1 and 12 times, their recurrence was classified as low. Purchases between 38 and 50 times are classified as very high.

```
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

Conclusion on based on the analysis:

- We can observe how customers are distributed across different purchase ranges.
- The majority of customers are in the low range, indicating that many customers buy sporadically.
- Customers who purchase between 38 and 50 times represent the top of the loyalty pyramid. This group is extremely valuable to the e-commerce.

Insights and Recommendations:

- Encourage recurrence from "Low" customers through promotions and reactivation strategies.
- Maximize the value of "Medium" and "High" customers with additional sales strategies and loyalty programs.
- Maintain and reward "Very High" customers, who are your top revenue generators, with exclusive benefits and personalized experiences.

3. Conclusion

This is a case study in data analysis, where specific parts were selected for a deeper analysis. The reference source used is listed above. Below is a summary of the main insights found, focusing on areas with the greatest impact on the development of marketing and sales strategies.

Summary of Conclusions and Insights:

- Revenue by Gender: Men generate more than double the revenue compared to women.
- Season of the Year: Autumn is the most profitable season, with a consumption peak.
- Products: Clothing is the main driver of sales, followed by accessories.
- Geographical Distribution: Smaller states, like Kansas and Hawaii, have economies concentrated in specific sectors, without following traditional economic patterns.
- Customer Loyalty: The majority of customers are sporadic ("Low"), while the most loyal (frequent buyers) are highly valuable.

Recommendations:

- Focus on High-Potential Regions: States like Illinois, California, and Nevada should receive special attention.
- Male Marketing: Target strategies toward the male audience.
- Autumn Promotions: Take advantage of the consumption peak in autumn with specific campaigns.
- Improvement in Underperforming Categories: Boost sales of footwear and winter clothing.
- Loyalty: Engage sporadic customers with promotions, and reward loyal ones with exclusive benefits.

To maximize growth, it is essential for the company to capitalize on consumption peak moments, such as autumn, and focus on personalized strategies for the most loyal customers, who generate the bulk of the revenue. This will ensure not only increased sales but also the loyalty of a valuable and growing customer base.