TOPIC : Analysing shopping trends using Tableau

Dataset link : https://www.kaggle.com/datasets/iamsouravbanerjee/customer-shopping-trends-dataset/data?select=shopping_trends.csv

Introduction :

The Customer Shopping Preferences Dataset offers valuable insights into consumer behavior and purchasing patterns. Understanding customer preferences and trends is critical for businesses to tailor their products, marketing strategies, and overall customer experience. This dataset captures a wide range of customer attributes including age, gender, purchase history, preferred payment methods, frequency of purchases, and more. Analyzing this data can help businesses make informed decisions, optimize product offerings, and enhance customer satisfaction. The dataset stands as a valuable resource for businesses aiming to align their strategies with customer needs and preferences. It's important to note that this dataset is a Synthetic Dataset Created for Beginners to learn more about Data Analysis and Machine Learning.

Dataset (Column-wise)

Customer ID - Unique identifier for each customer

Age - Age of the customer

Gender - Gender of the customer (Male/Female)

Item Purchased - The item purchased by the customer

Category - Category of the item purchased

Purchase Amount (USD) - The amount of the purchase in USD

Location - Location where the purchase was made

Size - Size of the purchased item

Color - Color of the purchased item

Season - Season during which the purchase was made

Review Rating - Rating given by the customer for the purchased item

Subscription Status - Indicates if the customer has a subscription (Yes/No)

Shipping Type - Type of shipping chosen by the customer

Discount Applied - Indicates if a discount was applied to the purchase (Yes/No)

Promo Code Used - Indicates if a promo code was used for the purchase (Yes/No)

Previous Purchases - The total count of transactions concluded by the customer at the store, excluding the ongoing transaction

Payment Method - Customer's most preferred payment method

Frequency of Purchases - Frequency at which the customer makes purchases (e.g., Weekly, Fortnightly, Monthly)

Data processing of Dataset

The provided Python code is a comprehensive data preprocessing pipeline for a CSV file using the pandas library. It performs essential data processing tasks in a structured manner.

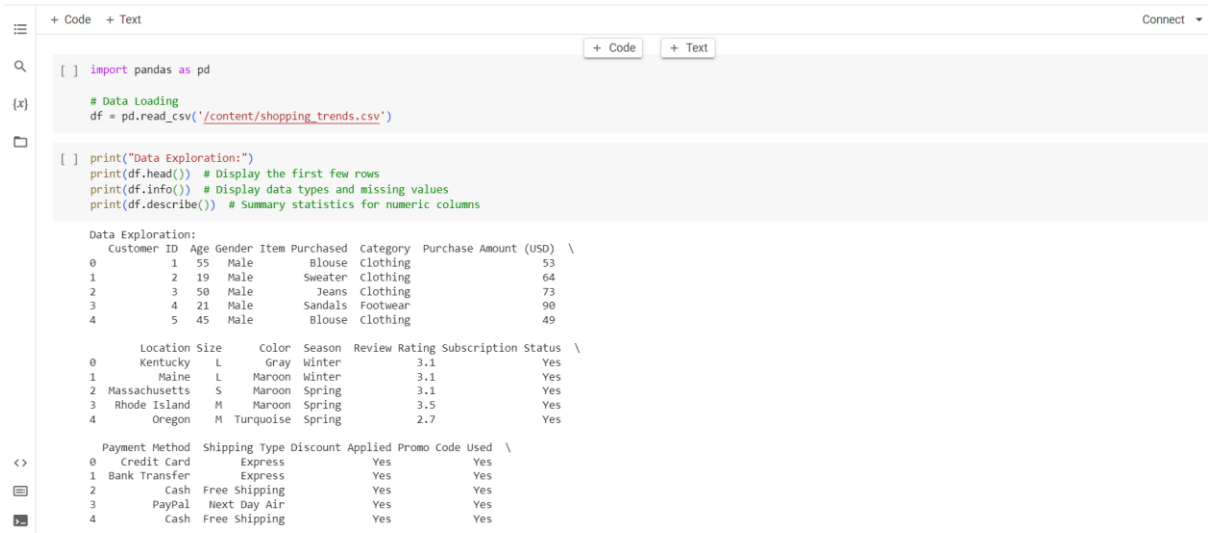
In the first paragraph, the code accomplishes data loading and exploration. It starts by loading a CSV file into a Pandas DataFrame, enabling easy data manipulation and analysis. The "Data Exploration" section prints out the first few rows of the dataset, giving a quick glimpse of the data's structure. Then, it uses the df.info() method to display detailed information about data types and missing values, which is crucial for understanding the quality of the dataset. Finally, it presents summary statistics for numeric columns using df.describe(), providing a deeper insight into the data's central tendencies and distributions. These exploration steps are vital for understanding the dataset's characteristics.

In the second paragraph, the code proceeds with data cleaning and additional data processing. It showcases an example of data cleaning by removing rows with missing values using df.dropna(). This step is critical for handling incomplete data. Lastly, the preprocessed data is saved to a new CSV file, 'preprocessed_data.csv', ensuring that the cleaned and transformed data can be easily accessed and used for further analysis or modeling. The code's modularity and clear structure make it a useful template for data preprocessing tasks, which are often the initial steps in data analysis and machine learning projects.

Code

```
import pandas as pd

# Data Loading

df = pd.read_csv('your_data.csv')


# Data Exploration

print("Data Exploration:")

print(df.head())  # Display the first few rows

print(df.info())  # Display data types and missing values

print(df.describe())  # Summary statistics for numeric columns


# Data Cleaning

# Example: Handling missing values by dropping rows with NaN values

df_cleaned = df.dropna()


# Save the Preprocessed Data to a New CSV File

df_cleaned.to_csv('preprocessed_data.csv', index=False)


print("Preprocessed Data Saved to 'preprocessed_data.csv'")
```

Application of pre-processing using google-collab



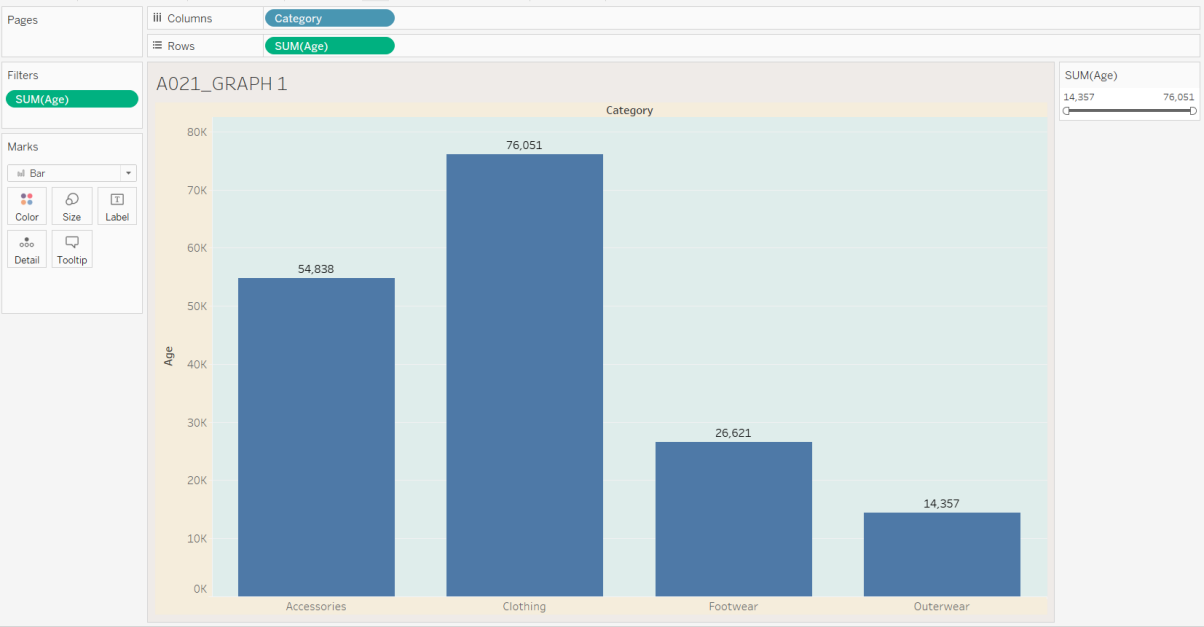
```
+ Code + Text                                                    Connect

[ ] import pandas as pd

    # Data Loading
    df = pd.read_csv('/content/shopping_trends.csv')

[ ] print("Data Exploration:")
    print(df.head())  # Display the first few rows
    print(df.info())  # Display data types and missing values
    print(df.describe())  # Summary statistics for numeric columns

Data Exploration:
   Customer ID  Age Gender Item Purchased  Category  Purchase Amount (USD)  \
0            1   55   Male         Blouse  Clothing                     53
1            2   19   Male        Sweater  Clothing                     64
2            3   50   Male          Jeans  Clothing                     73
3            4   21   Male        Sandals  Footwear                     90
4            5   45   Male         Blouse  Clothing                     49

       Location Size      Color  Season  Review Rating Subscription Status  \
0      Kentucky    L       Gray  Winter            3.1                  Yes
1         Maine    L     Maroon  Winter            3.1                  Yes
2  Massachusetts    S     Maroon  Spring            3.1                  Yes
3   Rhode Island    M     Maroon  Spring            3.5                  Yes
4        Oregon    M  Turquoise  Spring            2.7                  Yes

  Payment Method  Shipping Type Discount Applied Promo Code Used  \
0    Credit Card        Express              Yes              Yes
1  Bank Transfer        Express              Yes              Yes
2           Cash  Free Shipping              Yes              Yes
3         PayPal   Next Day Air              Yes              Yes
4           Cash  Free Shipping              Yes              Yes
```

```
+ Code + Text                                                    Connect

 17  Preferred Payment Method  3900 non-null   object
 18  Frequency of Purchases    3900 non-null   object
dtypes: float64(1), int64(4), object(14)
memory usage: 579.0+ KB
None
       Customer ID          Age  Purchase Amount (USD)  Review Rating  \
count  3900.000000  3900.000000            3900.000000    3900.000000
mean   1950.500000    44.068462              59.764359       3.749949
std    1125.977353    15.207589              23.685392       0.716223
min       1.000000    18.000000              20.000000       2.500000
25%     975.750000    31.000000              39.000000       3.100000
50%    1950.500000    44.000000              60.000000       3.700000
75%    2925.250000    57.000000              81.000000       4.400000
max    3900.000000    70.000000             100.000000       5.000000

       Previous Purchases
count         3900.000000
mean            25.351538
std             14.447125
min              1.000000
25%             13.000000

[ ] # Data Cleaning
    # Example: Handling missing values by dropping rows with NaN values
    df_cleaned = df.dropna()

[ ] # Save the Preprocessed Data to a New CSV File
    df_cleaned.to_csv('preprocessed_data.csv', index=False)

    print("Preprocessed Data Saved to 'preprocessed_data.csv'")

    Preprocessed Data Saved to 'preprocessed_data.csv'
```

Analysis using Tableau

A021_BIRVASHAH

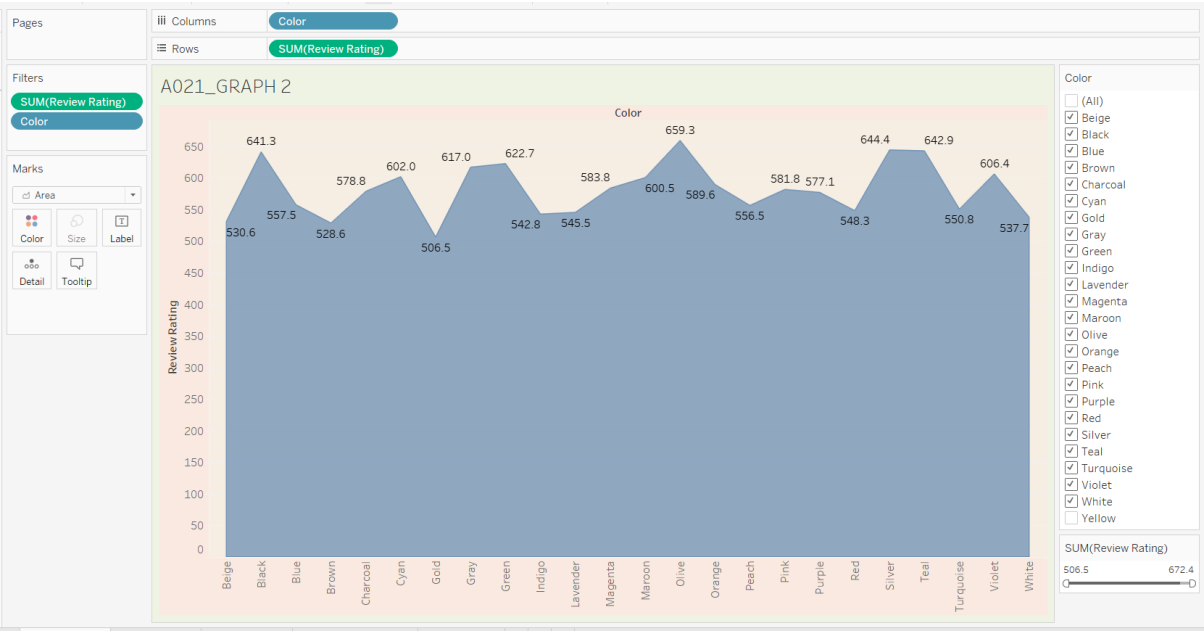
Graph 1



To analyze the different product categories (accessories, outwear, clothing, and footwear) in relation to the sum of ages, a bar graph was created where the categories were plotted on the x-axis, and the sum of ages for each category was represented on the y-axis. This graph provided a clear visual representation of how age is distributed across these product categories.

The bar graph revealed valuable insights into the demographic aspects of customer preferences. By examining the heights of the bars, it became apparent which product category had the highest sum of ages, indicating its popularity among a certain age group. This information can be used for targeted marketing and product development strategies. For instance, if the outwear category had the highest sum of ages, it suggests that a specific age group is more inclined to purchase outwear items. Businesses can use this insight to tailor their advertising campaigns or product offerings to cater more effectively to this demographic.

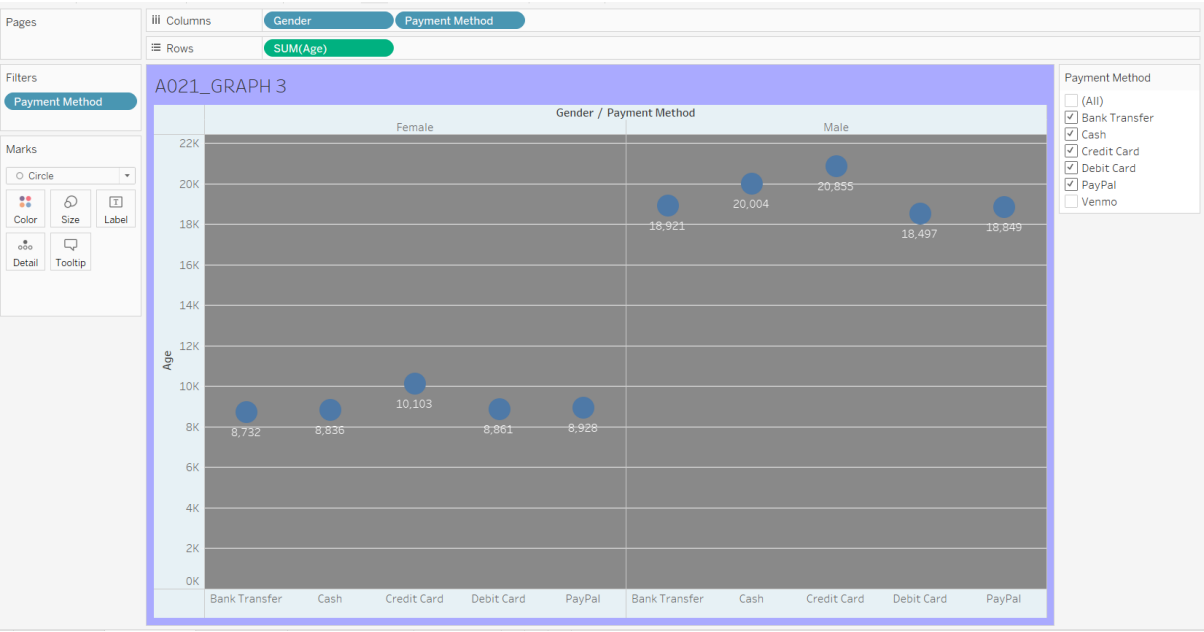
GRAPH 2



An area graph was created to visually represent data using "GENDER" as the row category and "REVIEW RATING" as the column measure. This graphical representation effectively conveyed the distribution of review ratings across different gender categories. The "GENDER" variable was placed on the horizontal axis, while the "REVIEW RATING" values served as the vertical axis. The area graph's shaded areas illustrated the density or count of each review rating within each gender category.

This visual depiction not only facilitated a quick and intuitive understanding of how review ratings were distributed among different genders but also allowed for easy comparisons between these categories.

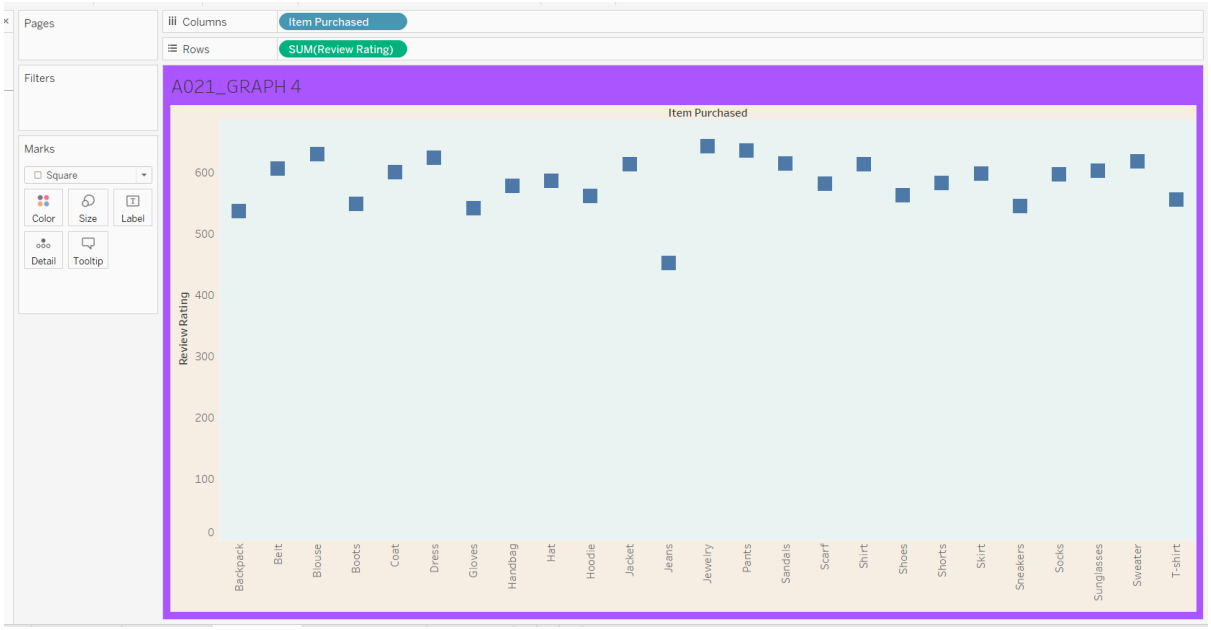
GRAPH 3



A circular graph was created to visually represent the distribution of ages among two categories, namely "GENDER" and "PAYMENT METHOD." The age values served as the measure for this visualization. Each sector of the circle graph corresponded to a specific age range, allowing for a quick and intuitive comparison between different groups.

The circular graph effectively displayed the age distribution for each gender and payment method, offering an immediate insight into any variations or patterns in the data. It presented a concise and easy-to-interpret overview of how age was distributed across these categories, facilitating the identification of any trends or differences in the dataset

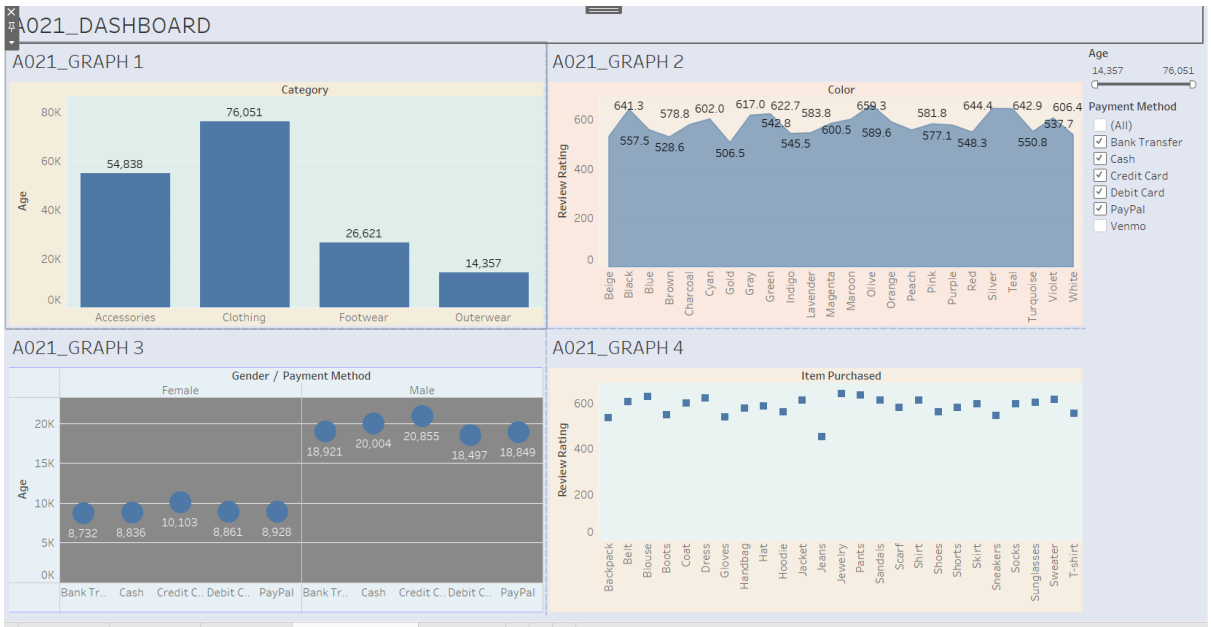
GRAPH 4



I created a square graph with "Shipping Type Method" as the rows and "Review Rating" as the columns, utilizing "Age" as the measure. This graph was designed to visually represent the relationship between these variables. The purpose of this graph was to analyze how shipping method, review ratings, and age were interrelated. Each cell in the graph represented a combination of a specific shipping method, review rating, and age measure. By examining the contents of each cell, it became possible to draw conclusions and identify any patterns or correlations between these variables.

The square graph was a valuable tool for understanding the dataset as it provided a compact and organized way to present the data

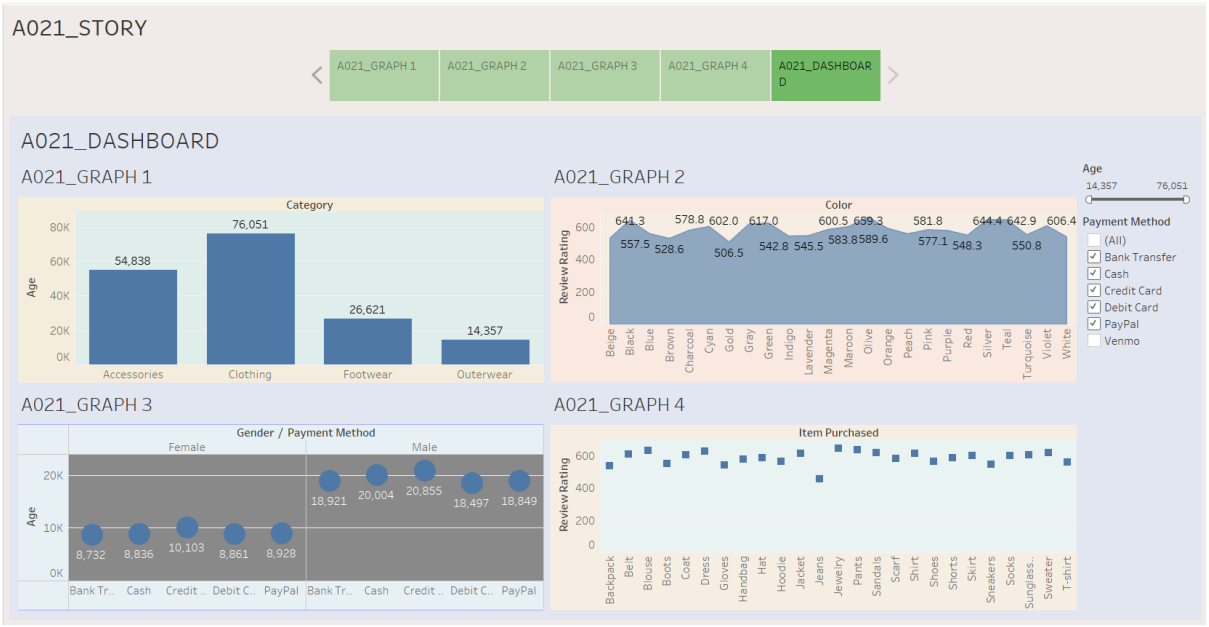
DASHBOARD



I created a dashboard featuring four distinct graphs aimed at analyzing shopping trends. These graphs collectively offered a comprehensive view of various aspects of shopping behavior and preferences. One graph provided insights into the distribution of age groups among shoppers, highlighting the demographic composition of the customer base. Another graph focused on product categories, showcasing which types of items were most popular among shoppers. A four graph sdelved into the frequency of purchases, offering an understanding of how often customers were making transactions.

The dashboard served as a powerful tool for decision-makers, enabling them to swiftly grasp and interpret essential shopping trends. It provided a holistic perspective on customer demographics, product preferences, purchase frequencies, and payment habits, which are pivotal for businesses to tailor their strategies, optimize inventory, and enhance customer experiences.

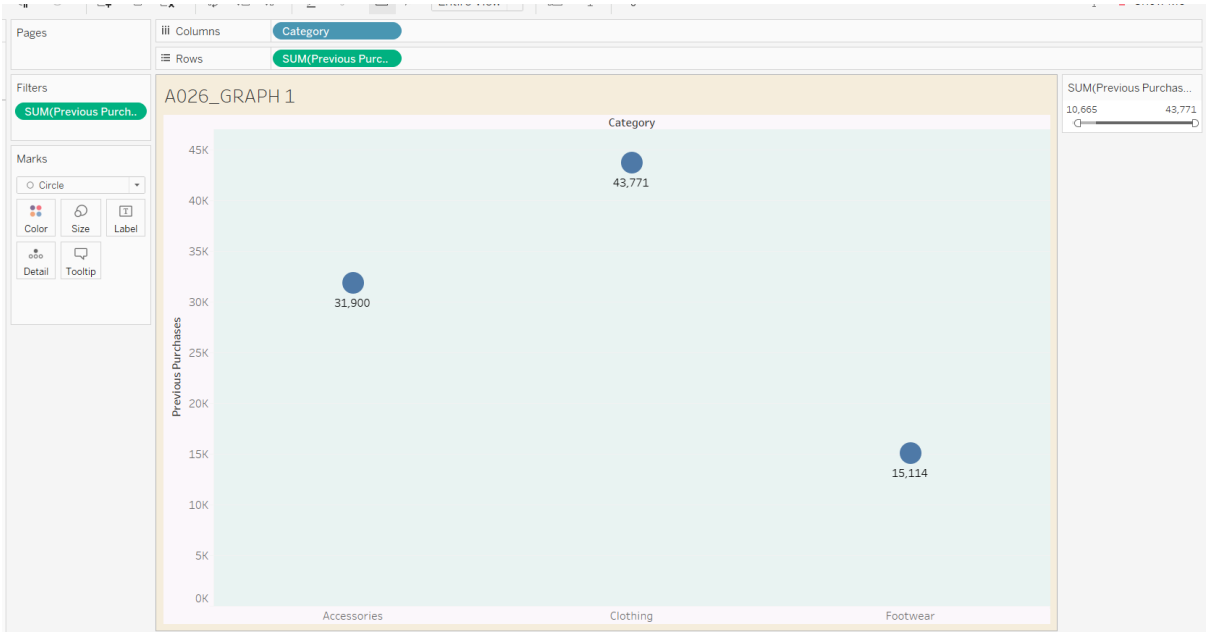
STORY



Created a story which contains all the four graphs and dashboard so that a visual presentation of shopping trends can be easily analysed . With the click on different graphs and dashboard we can also see each graph individually so that we get a clear a perspective of visualization of shopping trends.

A026_VISHWA JOBALIA

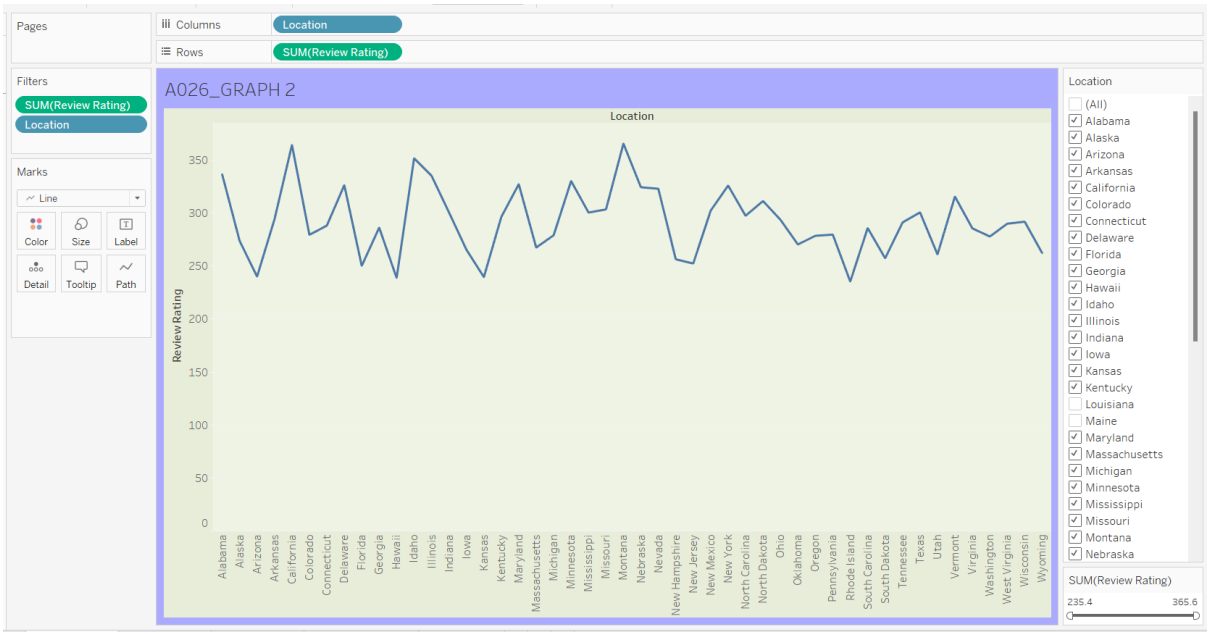
GRAPH 1



Created a circular graph that visualizes the relationship between product categories and previous purchases, we utilized the "Previous Purchases" as the measure, and "Category" as the rows. This graph was generated to gain insights into the distribution of purchases across various product categories. Each category is represented as a segment in the circle, with the size of each segment corresponding to the number of previous purchases for that specific category.

The circular graph effectively conveys the distribution of purchases, allowing us to identify which product categories have the highest and lowest purchase counts. It offers a visually intuitive representation of the data, making it easier to identify patterns and trends in customer behaviour.

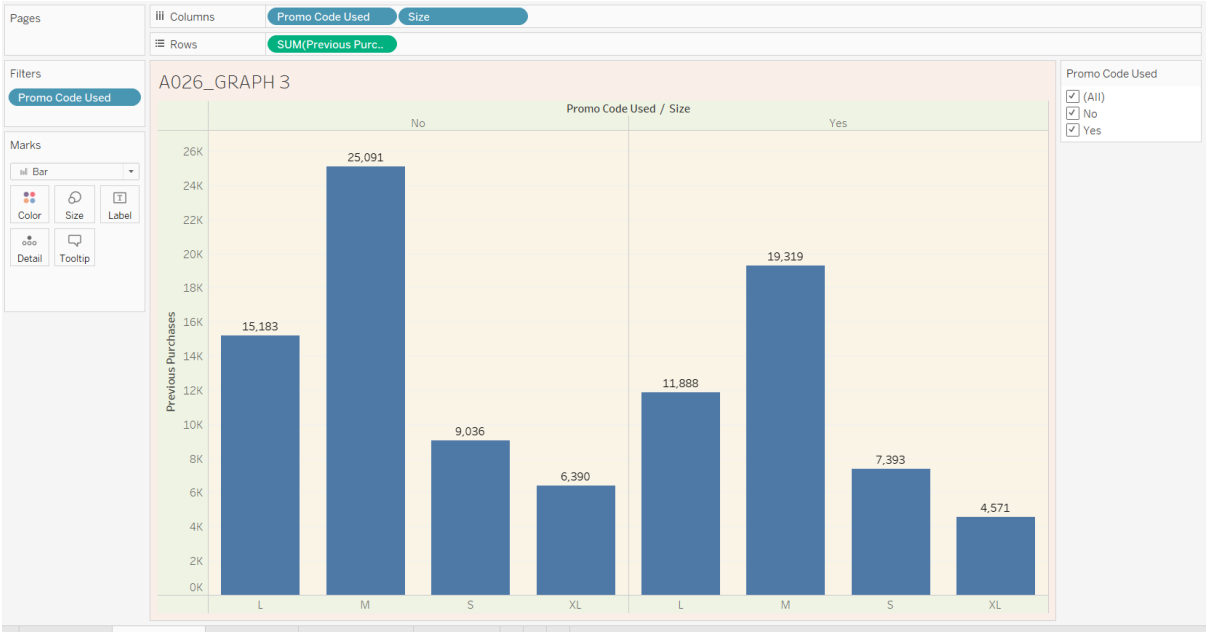
GRAPH 2



Created a line graph that visualized data with "LOCATION" on the x-axis and "REVIEW RATING" on the y-axis. The graph was designed to depict changes in "REVIEW RATING" over different "LOCATION" values. In this context, "LOCATION" served as the categorical variable, and "REVIEW RATING" was the quantitative

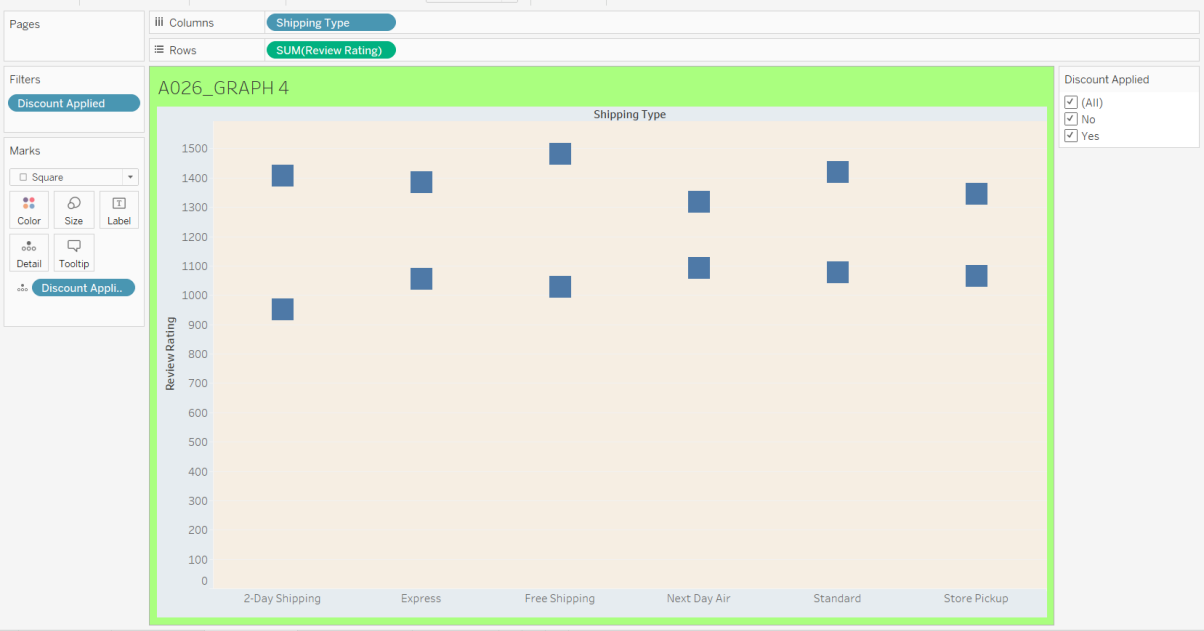
measure used to gauge the quality of a particular aspect or item. The line graph effectively showcased how "REVIEW RATING" varied across various locations, providing a clear and dynamic representation of the data.

GRAPH 3



A bar graph was created to visualize the relationship between "PROMO CODE USED" and "SIZE," with "PREVIOUS PURCHASES" serving as the measure along the vertical axis. In this graph, each unique combination of "PROMO CODE USED" and "SIZE" represents a distinct bar, while the height of each bar corresponds to the value of "PREVIOUS PURCHASES." This graphical representation provides a straightforward and concise way to compare the effect of promotional codes and garment sizes on customers' previous purchase behavior. Such visualizations are instrumental in understanding trends and patterns within the dataset, allowing for informed decision-making and marketing strategies based on past customer behaviors and preferences.

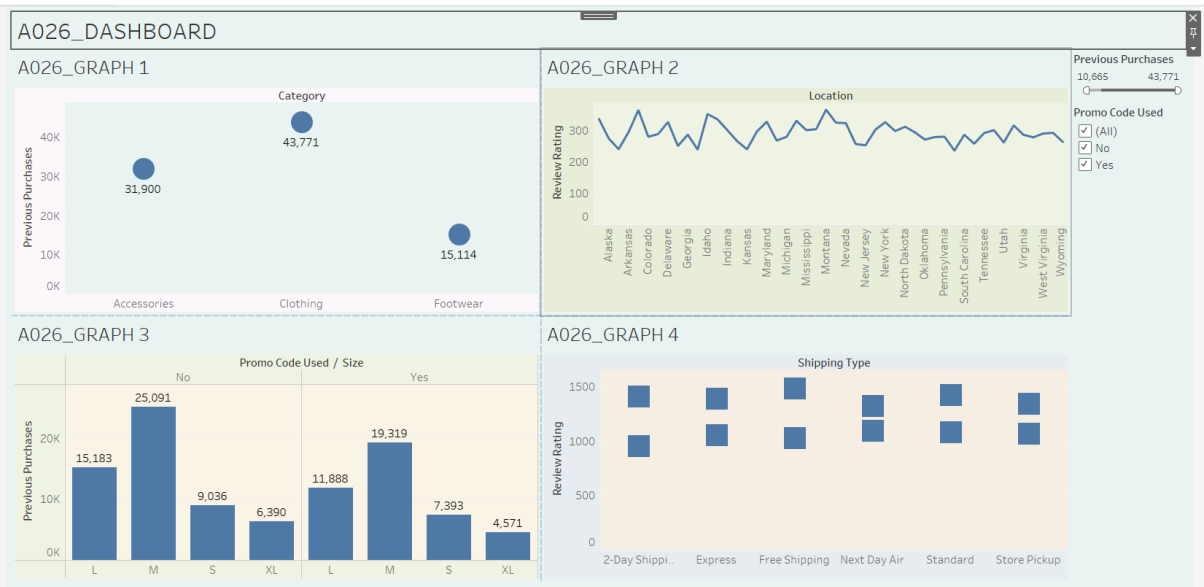
GRAPH 4



Created a square graph that visualizes the relationship between "Shopping Type" and "Review Rating." In this graph, "Shopping Type" is represented along the rows, while "Review Rating" is displayed along the columns. The graph was designed to help us understand the distribution of review ratings across various shopping types. The measure used for this graph is the "Review Rating," which serves as the data point that populates the cells of the square.

By constructing this square graph, past trends and patterns in review ratings associated with different shopping types can be effectively observed and analyzed. Such visualizations are particularly useful for gaining insights into how customers perceive and rate their shopping experiences, which can, in turn, inform businesses and marketers in tailoring their strategies to meet the specific needs and preferences of different shopping segments. This graph thus serves as a valuable tool for exploring and comprehending the relationship between shopping type and review rating based on historical data.

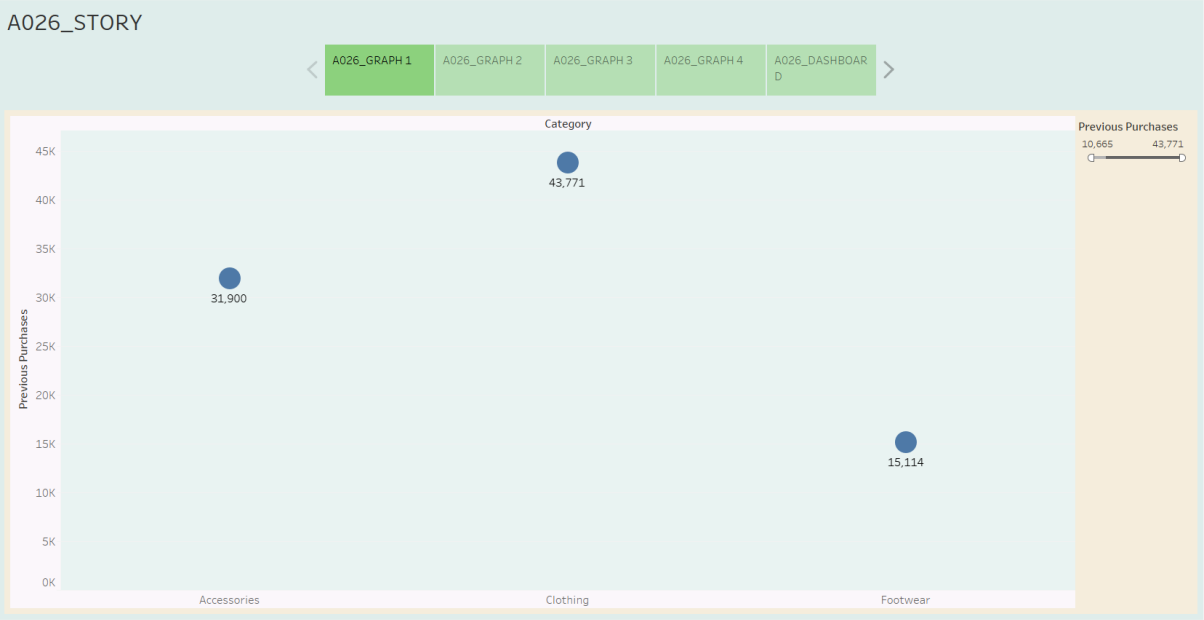
DASHBOARD



I made a dashboard with four different graphs to analyse shopping patterns. All of these graphs provided a thorough understanding of many facets of consumer preferences and behaviour related to purchasing. One graph illustrated the demographic makeup of the clientele by showing how different age groups were distributed among the shoppers. An additional chart centred on product categories, exhibiting the most popular item types among consumers. Four graphs were used to explore the frequency of purchases and provide insight into the frequency of transactions made by customers.

Decision-makers found the dashboard to be an effective tool that helped them quickly understand and analyse key buying trends.

STORY



In order to facilitate the analysis of a visual representation of shopping trends, a story comprising all four graphs and a dashboard was created. We can view each graph separately by clicking on the various graphs and dashboards, giving us a clear understanding of how shopping trends are visualised.

CONCLUSION

The analysis of shopping trends conducted through Tableau, comprising a collection of graphs, a meticulously designed dashboard, and a compelling story, has yielded invaluable insights into consumer behavior and preferences. The combination of these elements has provided a comprehensive view of the data, facilitating a deeper understanding of the dynamics at play.

The graphs within Tableau have visually represented trends in a clear and intuitive manner. These visualizations, encompassing various aspects of shopping, have allowed for the identification of patterns, anomalies, and correlations that might have otherwise remained hidden within the data. From the distribution of shopping types to the impact of promotional campaigns on sales, these visualizations have served as a powerful tool for data exploration and interpretation.

The dashboard, with its interactivity and real-time updates, has been a central hub for monitoring key performance indicators. Users can seamlessly explore data, filter information, and track changes in real time, which is vital for making informed decisions and spotting emerging trends as they unfold.

Additionally, the story built within Tableau has woven these insights into a compelling narrative. It has seamlessly connected the visualizations and dashboards, allowing stakeholders to follow a logical flow of information and grasp the key takeaways from the data. The story serves as a bridge between the raw data and actionable insights, providing a framework for decision-making and strategy development.

REFRENCES :

https://www.tableau.com/blog/top-metrics-every-sales-executive-cant-live-without

https://www.tableau.com/10-dashboards-every-retailer-should-use

https://www.datacamp.com/cheat-sheet/tableau-cheat-sheet

https://www.analyticsvidhya.com/blog/2017/07/data-visualisation-made-easy/

https://www.qualtrics.com/support/integrations/tableau-integration/