Customer Shopping Trends Dataset

Journey into Consumer Insights and Retail Evolution with Synthetic Data

About Dataset

Context

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.

Content

This dataset encompasses various features related to customer shopping preferences, gathering essential information for businesses seeking to enhance their understanding of their customer base. The features include customer age, gender, purchase amount, preferred payment methods, frequency of purchases, and feedback ratings. Additionally, data on the type of items purchased, shopping frequency, preferred shopping seasons, and interactions with promotional offers is included. With a collection of 3900 records, this dataset serves as a foundation for businesses looking to apply data-driven insights for better decision-making and customer-centric strategies.

Dataset Glossary (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** - Number of previous purchases made by the customer
- **Payment Method** - Customer's most preferred payment method
- **Frequency of Purchases** - Frequency at which the customer makes purchases (e.g., Weekly, Fortnightly, Monthly)

Question:
- Task is to complete the below sub questions and clearly give comments / justify your answer with visualizations, plots wherever possible.
- Assessment will be focused not just on the running code snippet in jupyter notebook file, but also it has to be supported with proper reasons why you chose a particular approach / plotting technique, and why not others could be used in that place instead.
- Please note: Since it's a group task involving 6 people at least, hence cleaner code i.e. modularized way of calling function to be used which rewards more marks rather than hard coding the plots.

Hints:
You would fetch more marks if following things are covered:
- to suppress display of warnings
- Pandas is used for data manipulation and analysis
- Numpy is used for large, multi-dimensional arrays and matrices, along with mathematical operators on these arrays
- Using Matplotlib as a data visualization library
- Using Seaborn, which aids in drawing attractive and informative statistical graphics
- display all dataframe columns (avoid cutting of display of columns due to presence of many columns in a dataframe when showing head(5))
- to set the floating numbers precision limited to 3 decimals
- display all dataframe rows
- Check dimensions of the dataframe in terms of rows and columns
- Check data types. Ensure your data types are correct. Refer data definitions to validate
- If data types are not as per business definition, change the data types as per requirement
- Study summary statistics
- Check for missing values
- Study correlation
- Detect outliers

- Use various plotting techniques where possible to clearly convey the analysis insights, give justification of the insights observed from each plot in words.
- Conclude with inferences or business decisions that are take away points if possible.

Source:
https://www.kaggle.com/datasets/iamsouravbanerjee/customer-shopping-trends-dataset/data