**BU51037**

**Data Visualization for Business**

**Step 1: Dataset Selection**

This dateset is from kaggle: <https://www.kaggle.com/datasets/mohammadtalib786/retail-sales-dataset/discussion?select=retail_sales_dataset.csv>

The dataset consist of 1000 rows and 9 columns

Each column in this dataset plays a pivotal role in unraveling the dynamics of retail operations and customer behavior.

This dataset represents a fictional retail environment, encompassing crucial factors influencing retail activities and customer engagements. It comprises important information like:

Transaction ID: A unique identifier for each transaction, allowing tracking and reference.

Date: The date when the transaction occurred, providing insights into sales trends over time.

Customer ID: A unique identifier for each customer, enabling customer-centric analysis.

Gender: The gender of the customer (Male/Female), offering insights into gender-based purchasing patterns.

Age: The age of the customer, facilitating segmentation and exploration of age-related influences.

Product Category: The category of the purchased product (e.g., Electronics, Clothing, Beauty), helping understand product preferences.

Quantity: The number of units of the product purchased, contributing to insights on purchase volumes.

Price per Unit: The price of one unit of the product, aiding in calculations related to total spending.

Total Amount: The total monetary value of the transaction, showcasing the financial impact of each purchase.

**Step 2: Data Exploration**

No missing values were identified in the dataset which is essential for conducting precise and dependable analyses

Among the various product categories being considered, the one related to clothing had the highest sales. This information is valuable for businesses as it highlights which product line is performing the best, The dataset indicated that the ages 64 and 43 are the most common or frequently observed. Understanding the age distribution is crucial for targeted marketing, product development, and tailoring services to the specific needs and preferences of different age groups.

There are more occurrences of female customers than male customers. Understanding the gender distribution is essential for businesses to tailor their products, services, and marketing strategies to the predominant demographic, ensuring they effectively reach and engage their target audience.

**Step 3: Visualization Creation**

A box plot illustrated the average total amount for each gender category, highlighting that female spent more on average.

A bar chart displayed the distribution of purchases based on 'Product Category,' 'Age Category,' and 'Gender,' providing insights into spending variations across different categories and age groups. Bar charts are suitable for this type of visualization.

I created a stacked bar chart comparing gender to age categories. The x-axis represents the age categories, and each bar is divided into segments for different genders. The legend indicates which color corresponds to each gender.

A heatmap was created to visualize the correlation between age, gender, and total amount spent, aiding in the identification of patterns or trends.

A pie chart was employed to depict age distribution after grouping into categories ('18-25', '26-35', '36-45', '46-55', '56-65', '66-75', '76-85', '86+').

A stacked bar chart showcased the distribution of purchases in each product category, segmented by age group and gender.

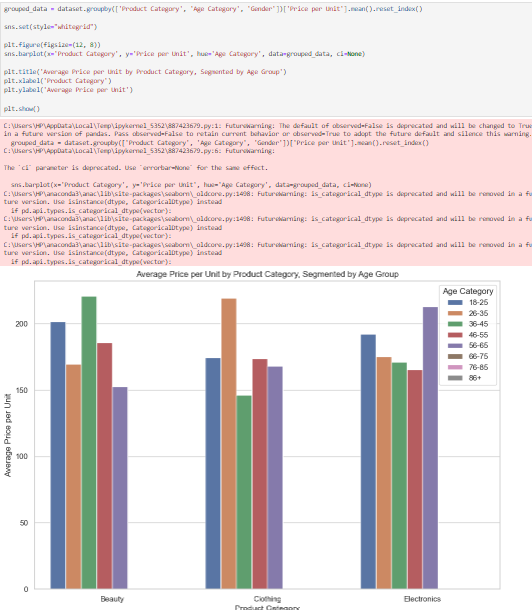
**Step 4: Interpretation and Analysis**

**Research Question:** How does customer age and gender influence their purchasing behavior?

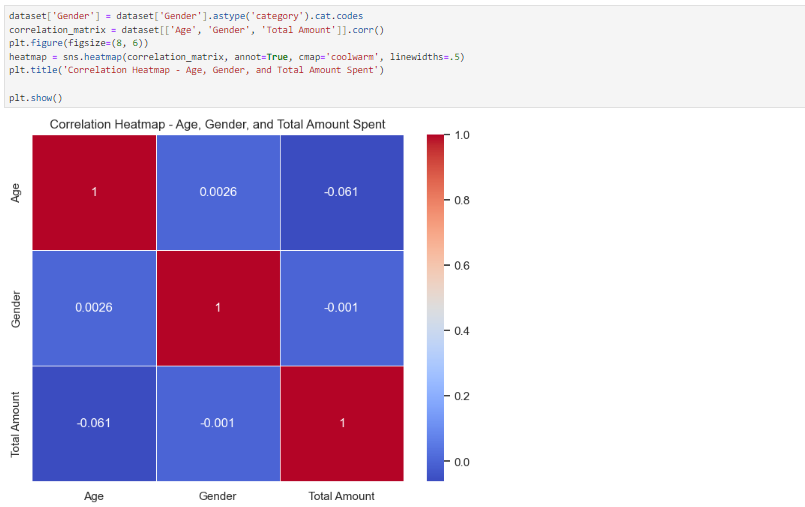
The analysis of purchases across product categories and age groups reveals a notable preference for clothing over beauty and electronic products. Specifically, the data indicates that females are the predominant consumers of clothing, aligning with expectations. Conversely, electronics exhibit lower purchase rates, with a higher representation of males, suggesting a greater interest in electronic products among males compared to females.



Now, if we check out the number of men and women in different age groups, we'll notice that there are more women overall. This means more ladies are around to buy stuff, especially clothes. So, the fact that we see more clothes sold than other things is because there are more women shopping. The connection between who's buying and how old they are really makes a big impact on what's getting sold, especially when it comes to clothes.



The correlation heatmap shows as expected that the diagonal is 1, which corresponds to the correlation of a number with itself, the matrix is symmetric and the features demonstrate low correlation, as evidenced by the proximity of their correlation values to zero.



In summary, this analysis shows how people buy different things based on their age and gender. It turns out that more ladies tend to buy clothes, which makes sense because there are simply more women overall. On the other hand, guys seem to be more interested in electronic stuff. The fancy chart we made also confirmed some basic things we expected, like the diagonal line being 1 and the patterns being kind of like a mirror. So, basically, what I noticed is that who's buying and how old they are really matters when it comes to what stuff gets sold, especially when it's about clothes and gadgets.



**APPENDIX**

