Data Analysis Report: Black Friday Dataset

Dataset Name: Black Friday Dataset

Executive Summary

In this project, we aimed to build a predictive model for customer purchase amounts in order to help a retail company, "ABC Private Limited," create personalized offers for their products. The dataset contains information about customer demographics, product details, and total purchase amounts for selected high-volume products from the previous month. Our primary focus was on model building, but we conducted some initial exploratory data analysis (EDA) to gain insights into the dataset.

Data Collection

Data Source:

The Black Friday dataset used in this analysis was obtained from kaggle platform. The dataset contains detailed information about customer purchases, including demographics, product categories, and purchase amounts.

Data Collection Method:

The data used in this analysis was sourced from kaggle platform and was made available for the purpose of this project. The data collection process involved the extraction of the following key information:

- **Gender**
- **Age**
- **Occupation etc.**
- **Stay_In_Current_City_Years**
- **Marital_Status**
- **Product_Category_1**
- **Product_Category_2**
- **Product_Category_3**

Data Preprocessing

So many preprocessing are done all are explained in jupyter notebook file with proper documentation command.

Observation and Analysis

For this project, the primary focus is on model building. However, we have conducted some preliminary exploratory data analysis to gain insights into the dataset.

1. Which Product Category Has the Highest Purchase?

- Product_Category_1 has the highest purchase. This is indicated by the red color in the graph, suggesting that products in category 1 are the most popular and generate the highest revenue.

2. How Does Category A Differ from B and C?

- Detailed analysis comparing categories A, B, and C is given with the representation of graph in jupyter notebook.

3. Comparison Between Some Important Features

- **Gender**: Men make more purchases than women. The purchasing pattern indicates a gender-based disparity in buying behavior.
- **Age**: Across all age categories, men tend to make higher purchases. There is a significant difference between the purchasing patterns of different age groups.
- **Occupation**: The 18th occupation category shows that males make purchases exceeding 10,000. Occupation plays a role in determining purchase behavior.
- **Product Categories**:
    - Product Category 1: The 10th product in this category is purchased the most, with sales surpassing 16,000.
    - Product Category 2: The 10th product is also popular in this category, with purchases exceeding 14,000, followed by the 2nd product with over 12,000 in sales.
    - Product Category 3: Most products in this category are purchased within the range of 10,000 to 12,000, suggesting relatively uniform sales across the category.
- **Marital Status**: In the "0-17" age category, unmarried individuals do not appear to have made any purchases during the Black Friday sales. However, other age groups show purchases ranging from 8,000 to 10,000.

Conclusion

The preliminary analysis provides some insights into the dataset, including the highest-selling product categories, differences in purchase behavior based on gender and age, and the influence of occupation and marital status. To build an accurate model for predicting purchase amounts, further data preprocessing, feature engineering, and modeling steps will be necessary. This EDA serves as the foundation for the model-building phase of the project.

NOTE:

All the exploration and visualization are present in the jupyter notebook, If you need more interactive dashboards and visualization use visualization tools like tableau and powerbi.

Please note that my data analysis reports based on the Black Friday dataset feature unique observations and analyses. These insights are the result of my own knowledge and analytical skills, and they differ significantly from online reports. My observations and the code used in my analysis are entirely distinct from what is available online.

ONLY FOR EDUCATIONAL PURPOSE

Prepared By,
Ahamed Basith