Black Friday Sales Prediction

Presented to you by:

Raed Habib

Under Supervision of:

Dr. Doaa Mahmoud

Agenda

- Introduction
- Project goal
- Initial Questions to be Answered
- Exploratory Data Analysis
- Data Pre-processing and Problems
- Modeling and Results
- Conclusion

Introduction

A retail company “ABC Private Limited” wants to understand the customer purchase behavior (specifically, purchase amount) against various products of different categories. They have shared purchase summaries of various customers for selected high-volume products from last month. The data set also contains customer demographics (age, gender, marital status, city type, stayincurrentcity), product details (productid and product category), and Total purchase amount from last month.

Project goal

Now, they want to build a model to predict the purchase amount of customers against various products which will help them to create a personalized offer for customers against different products.

Questions to be Answered:

Initial Questions to be Answered:

1- How many products do we have?

2- How does gender affect the purchases?

3- How does age also affect the purchases?

4- What is the relationship between the occupation and the purchases?

5- Does the city category has any impact on the purchases?

6- Does the martial status have any impact?¶

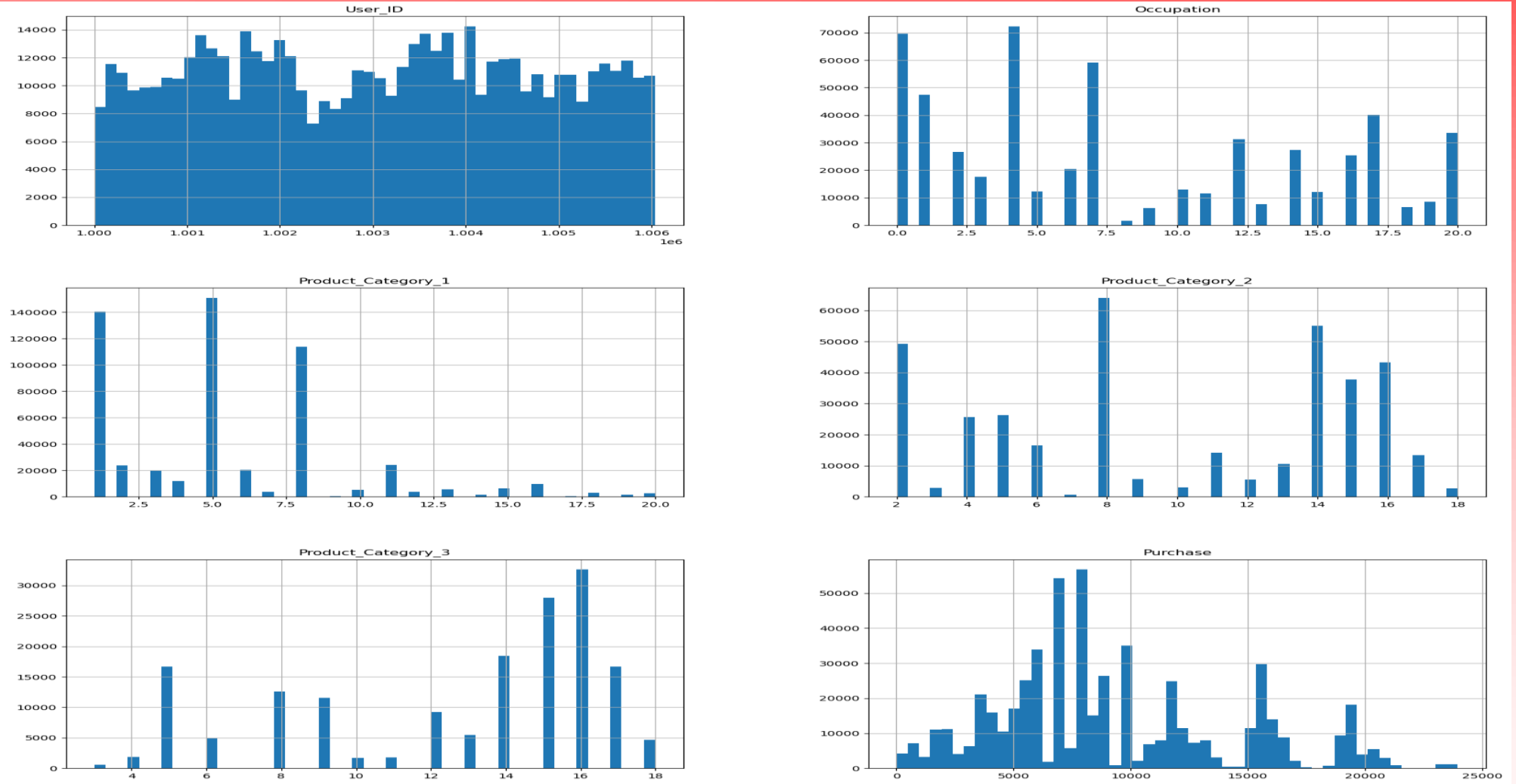
7- What are the most categories buyers have purchased from?

8- What are the top products?

9- Who are the top customers?

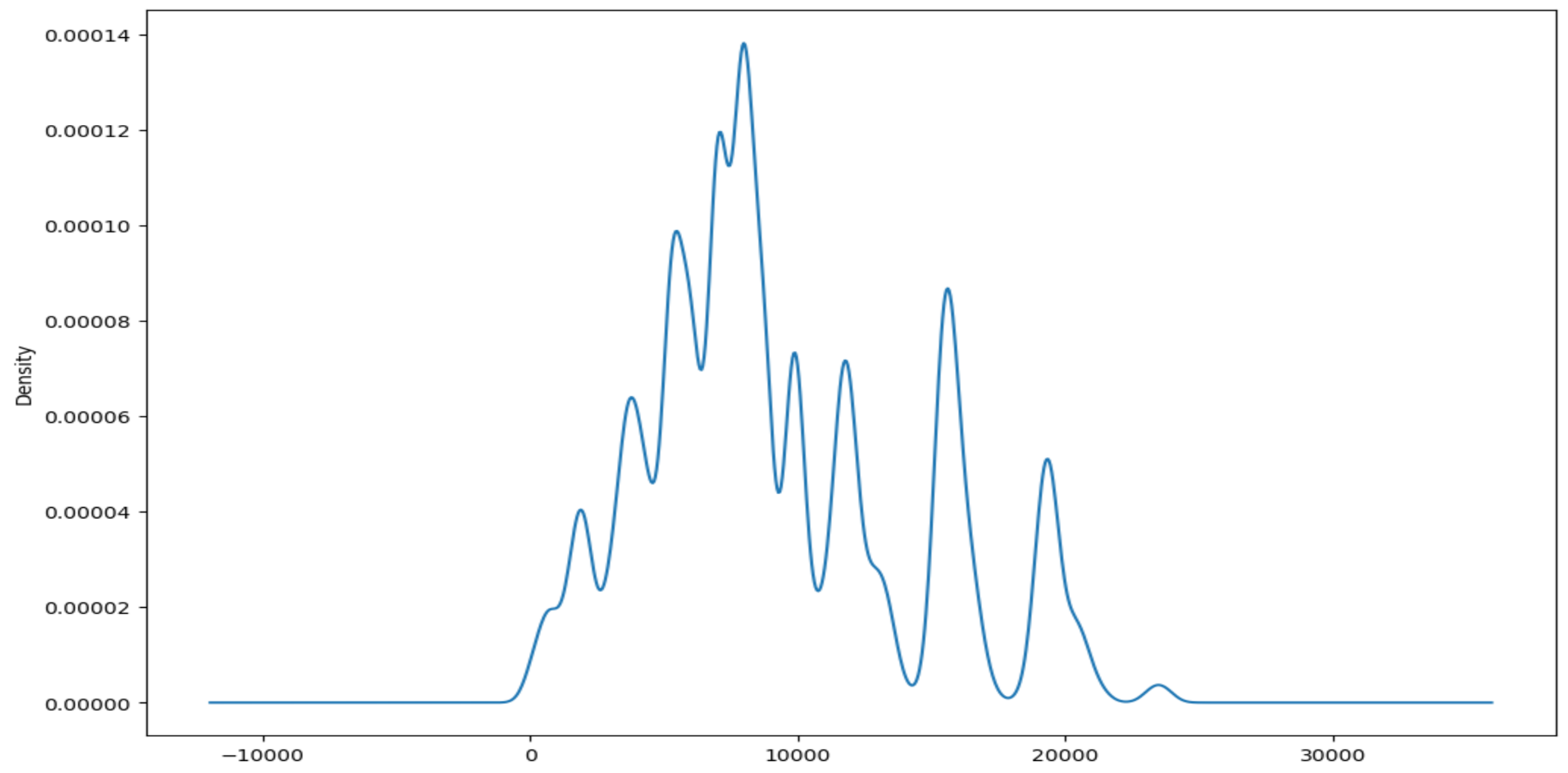
Exploratory Data Analysis

Data Distribution



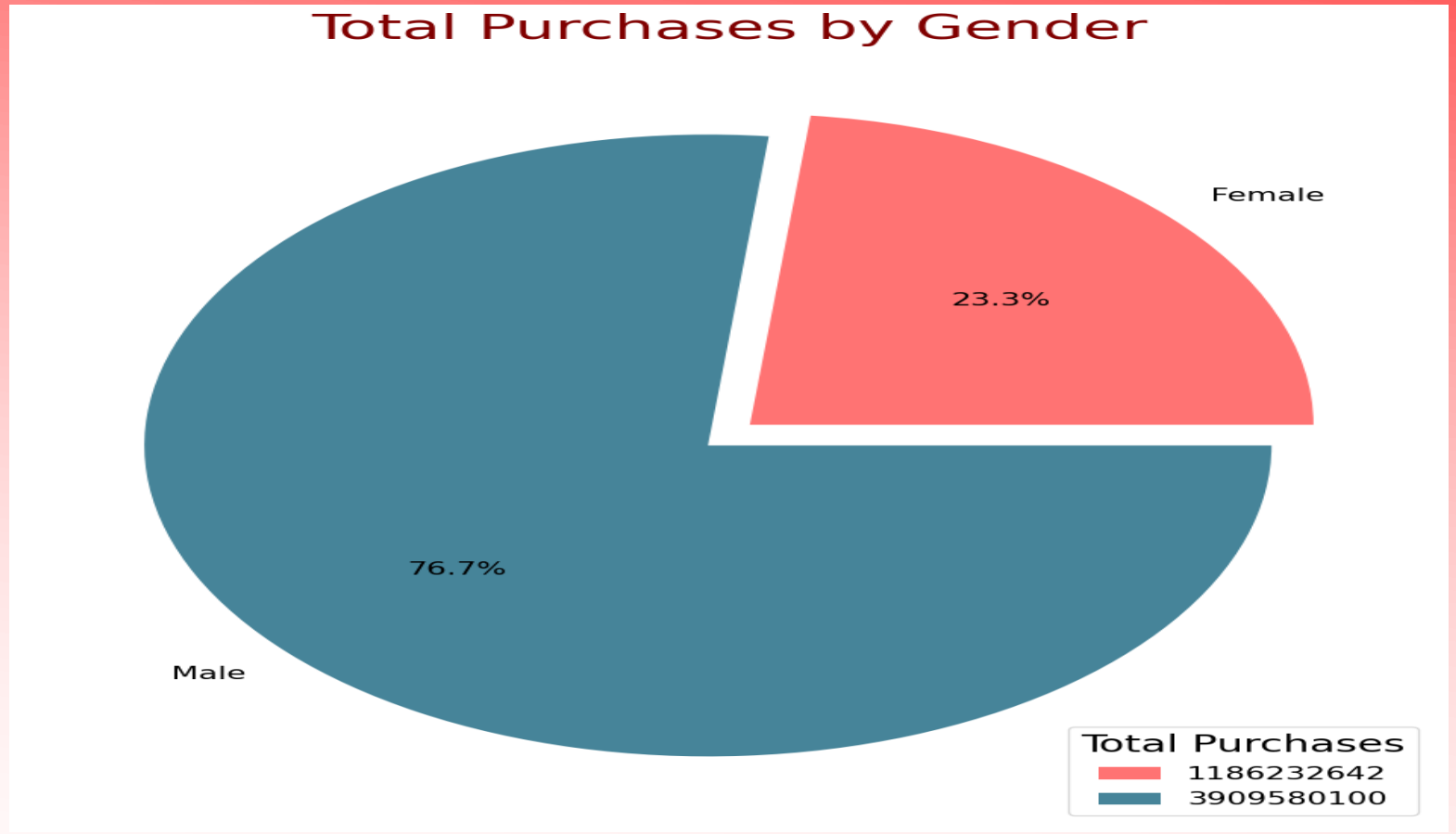
Exploratory Data Analysis

Target Distribution (Purchases)



Exploratory Data Analysis

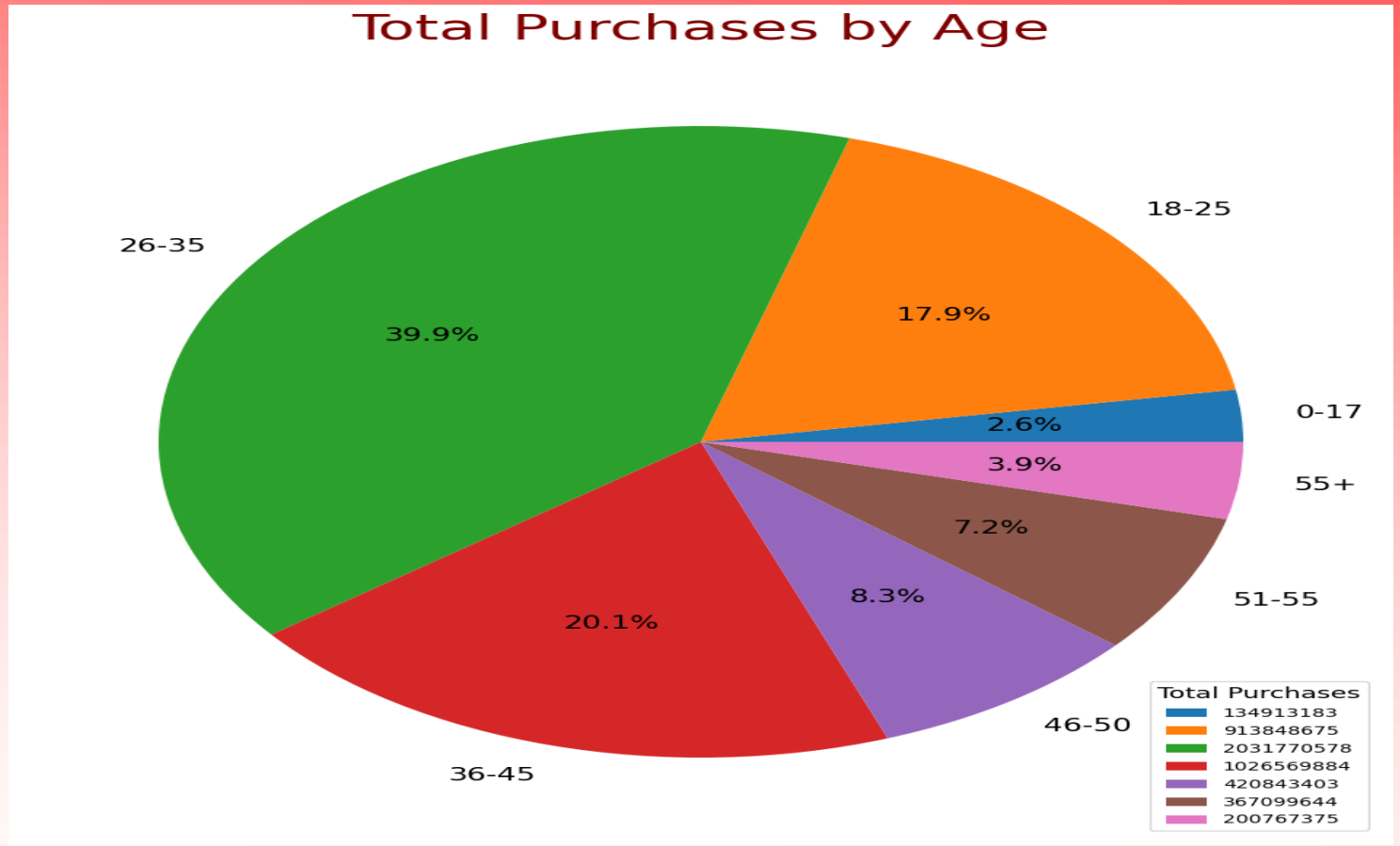
Effect of Gender on Purchases



We can see that males have the majority of purchases

Exploratory Data Analysis

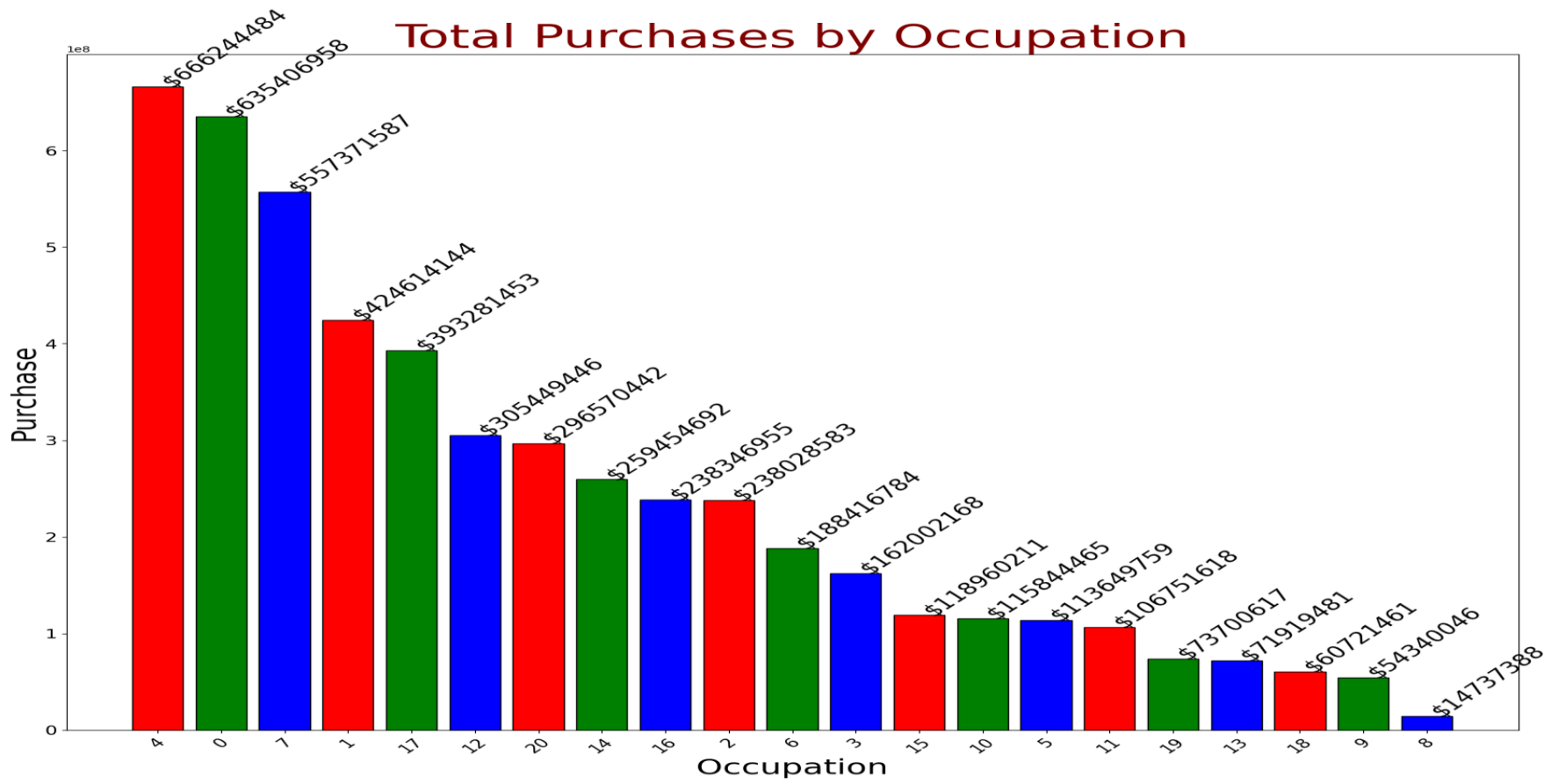
Effect of Age on Purchases



We can see that young adults have the majority of purchases

Exploratory Data Analysis

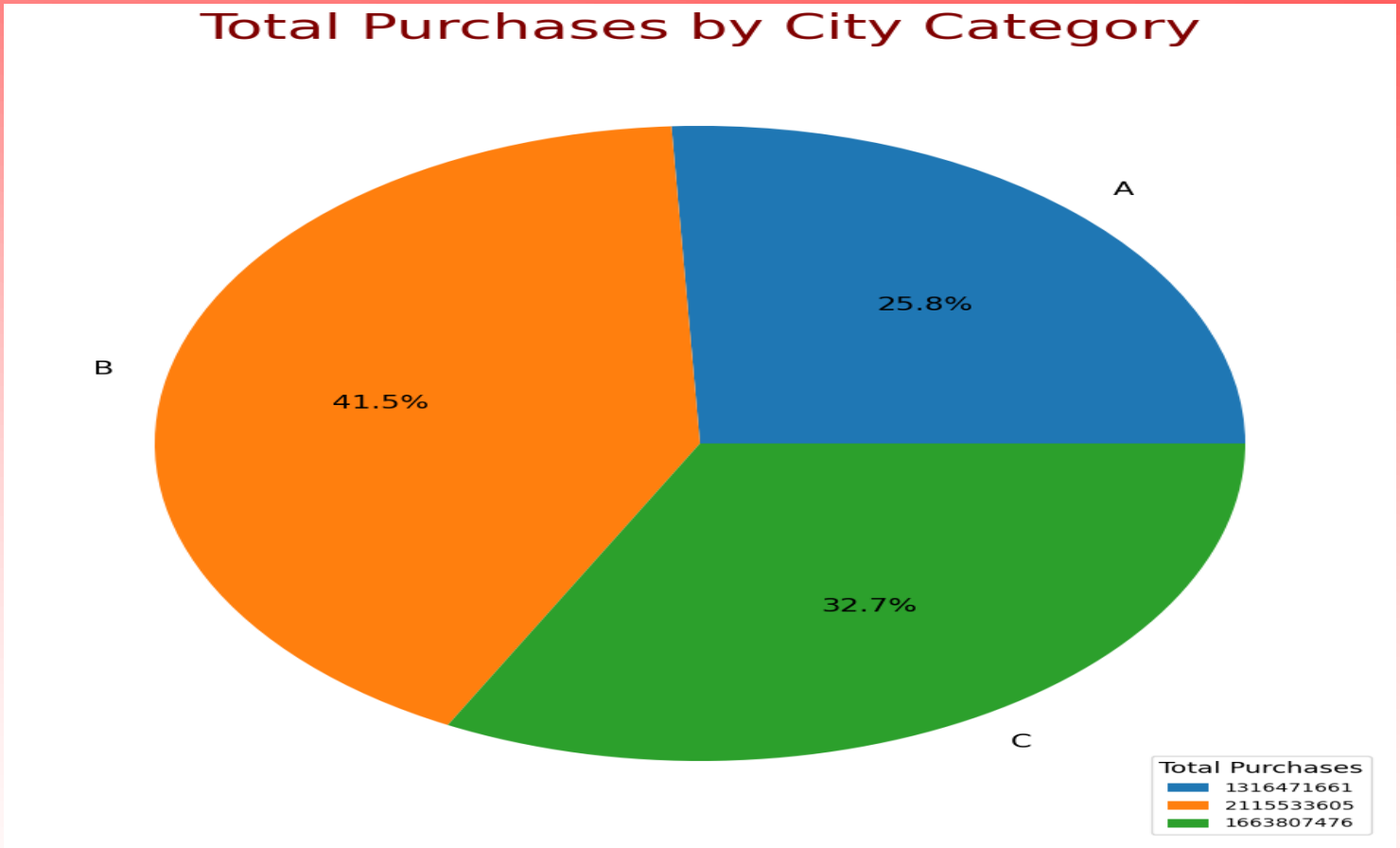
Relation between Occupation and Purchases



We can see that 4, 0 and 7 occupations have the majority of purchases

Exploratory Data Analysis

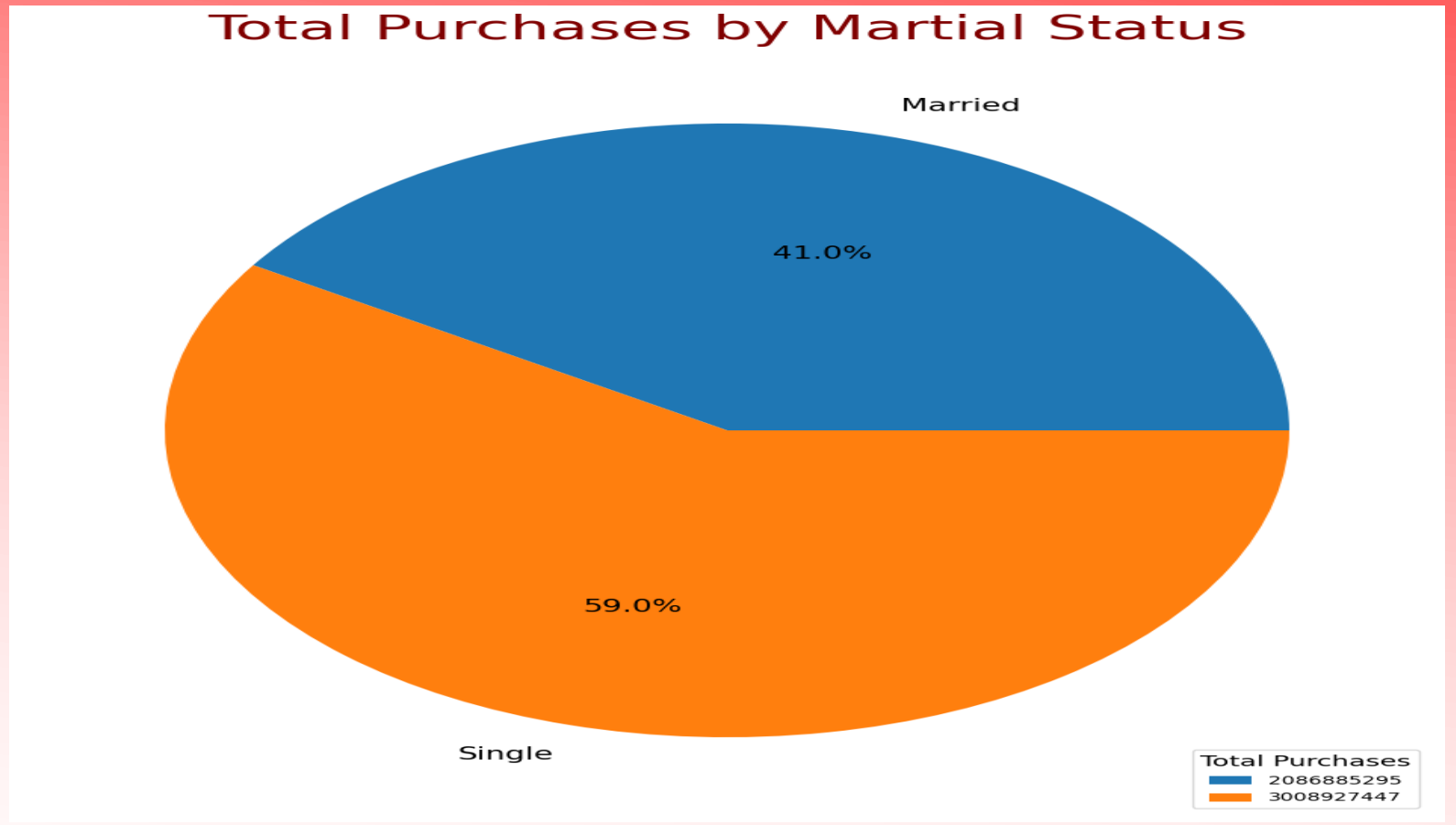
Effect of City location on Purchases



We can see that B has the majority of purchases

Exploratory Data Analysis

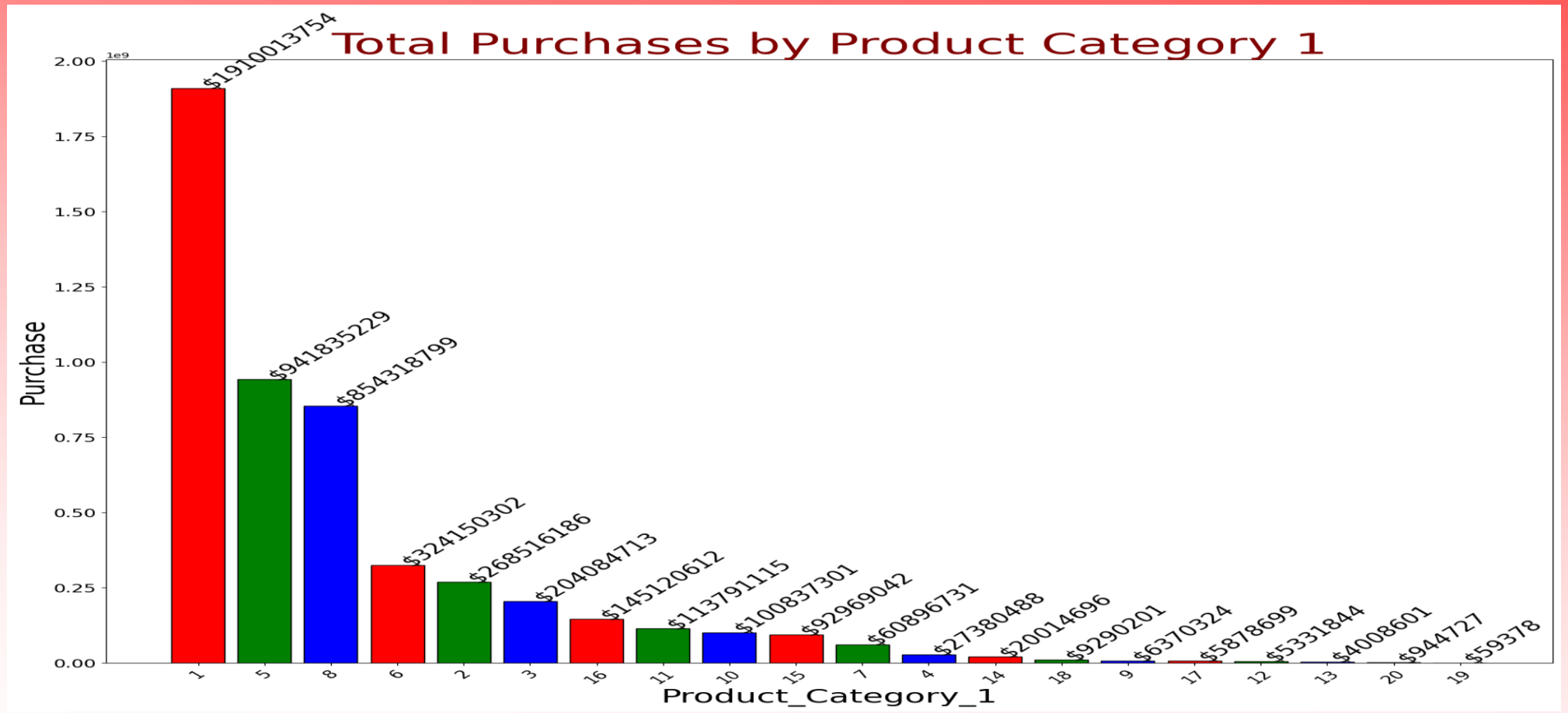
Effect of Martial Status on Purchases



We can see that single people have the majority of purchases

Exploratory Data Analysis

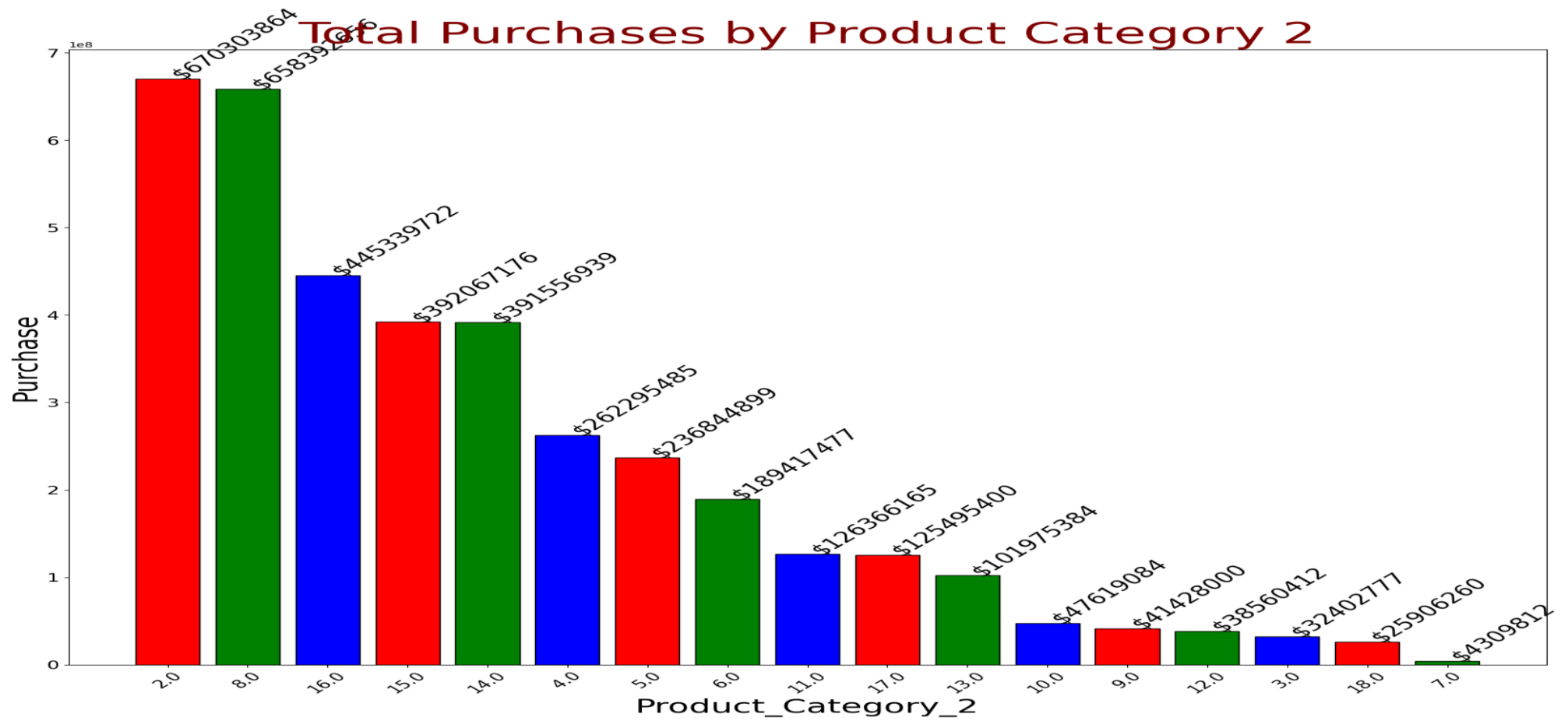
Most Purchases by each Category



We can see that category “1” has the majority of purchases and by far

Exploratory Data Analysis

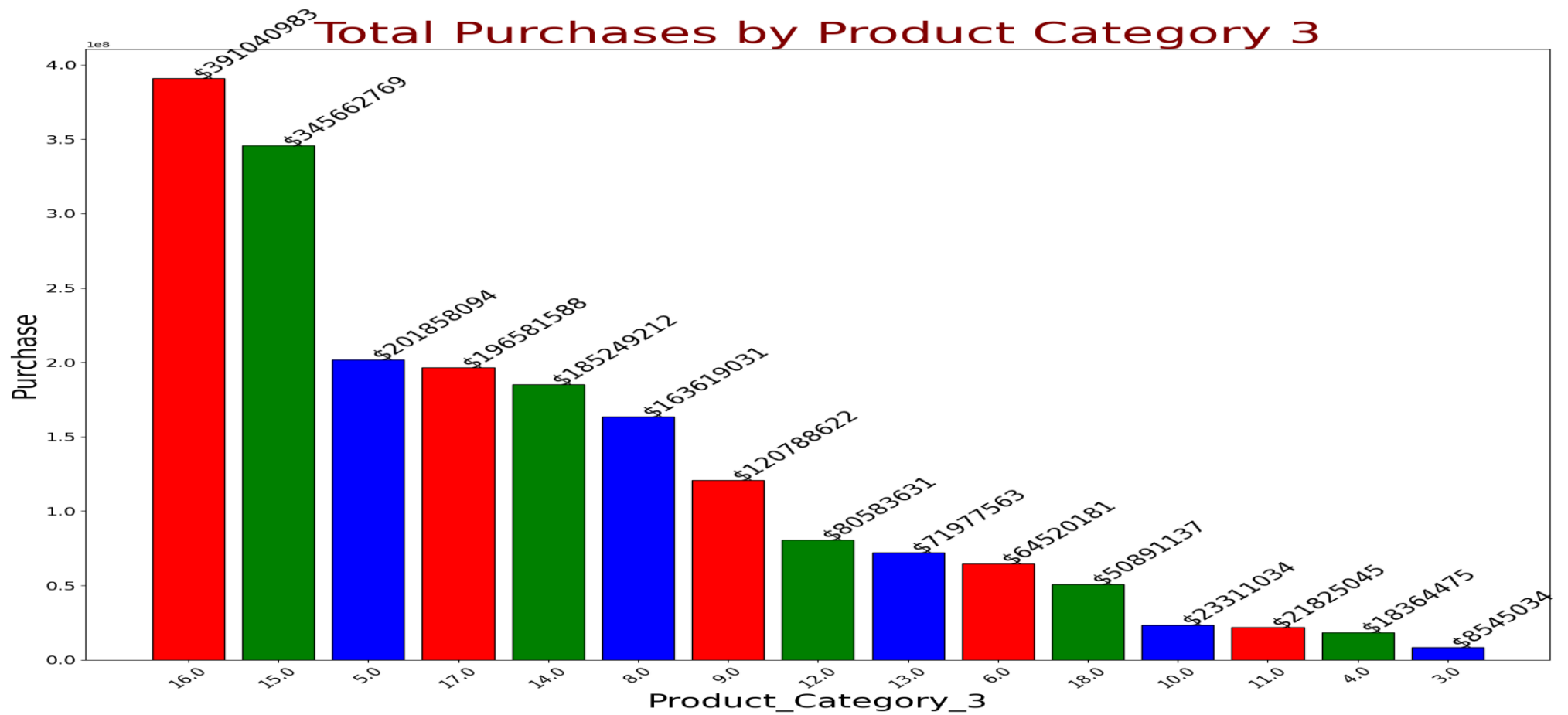
Most Purchases by each Category



We can see that both categories “2” and “8” have

Exploratory Data Analysis

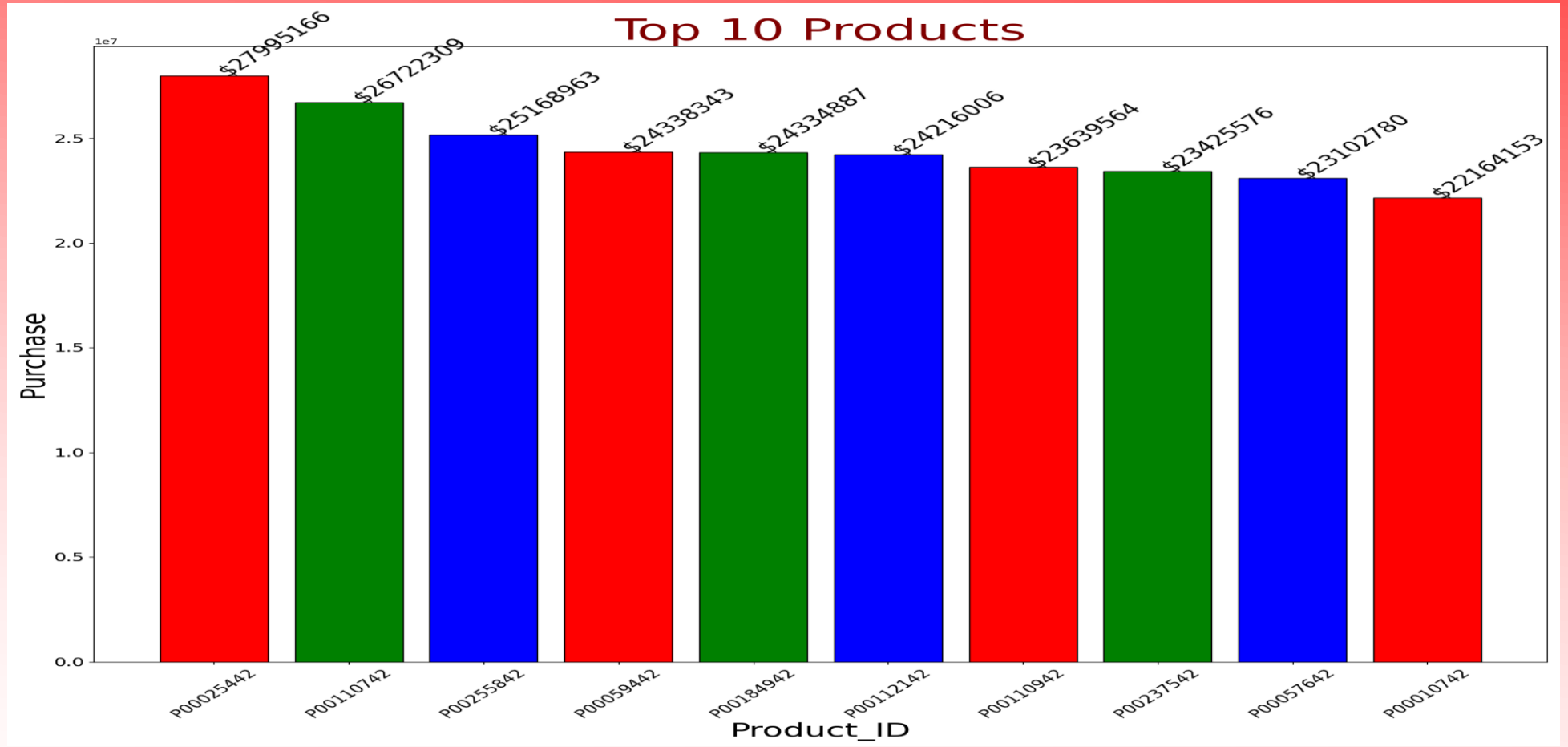
Most Purchases by each Category



We can see that both categories “16” and “15” have the majority of purchases and by far

Exploratory Data Analysis

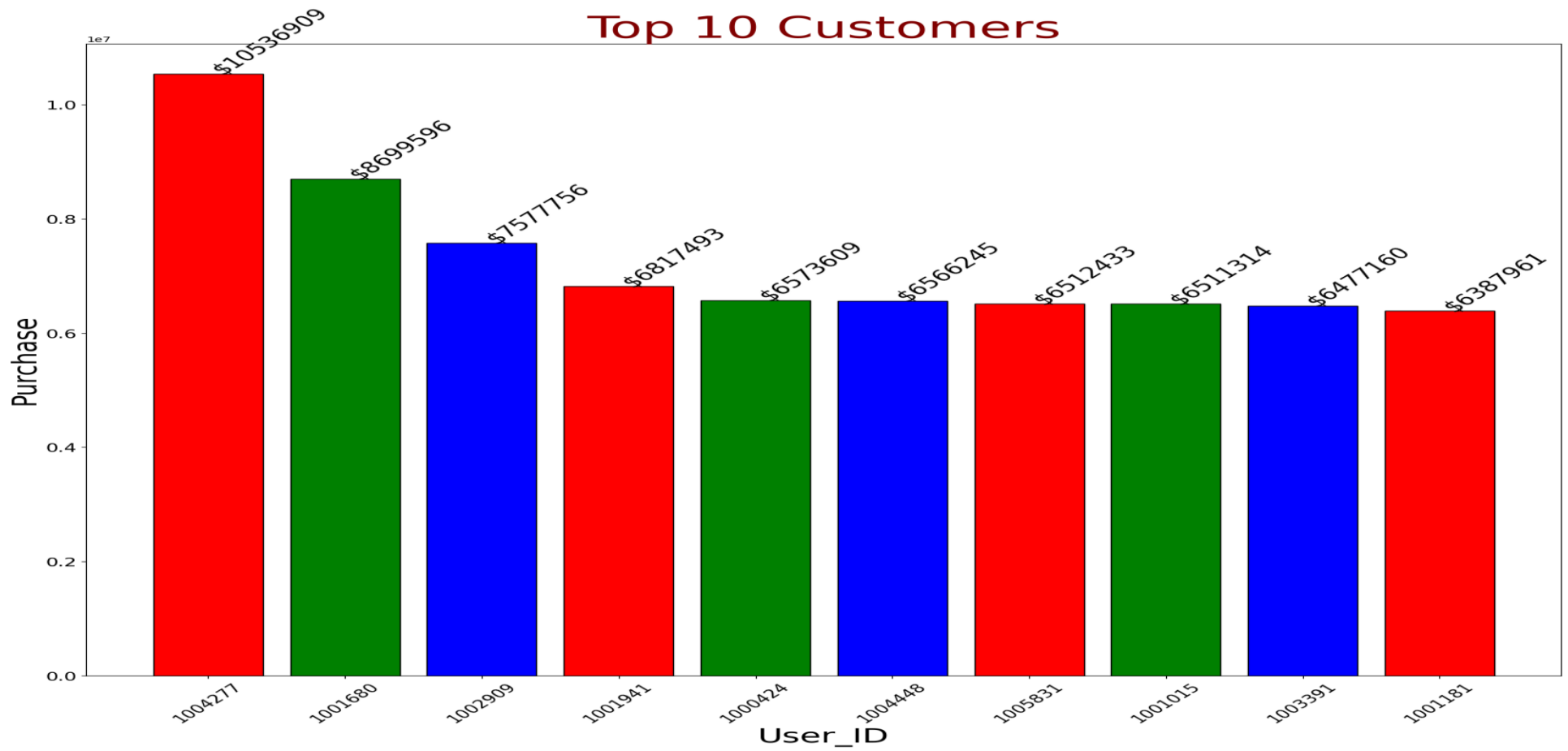
Top Products



The top 10 products

Exploratory Data Analysis

Top Customers

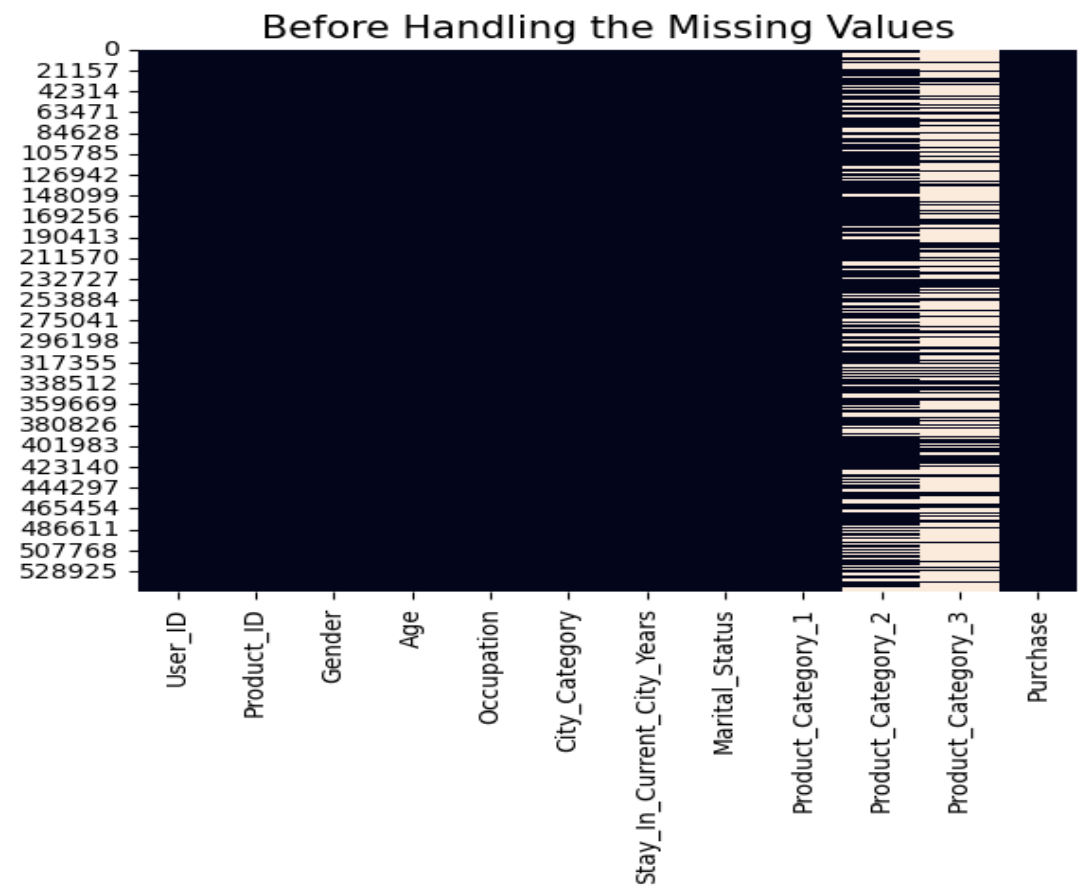


The top 10 Customers

Data Pre-processing and Problems

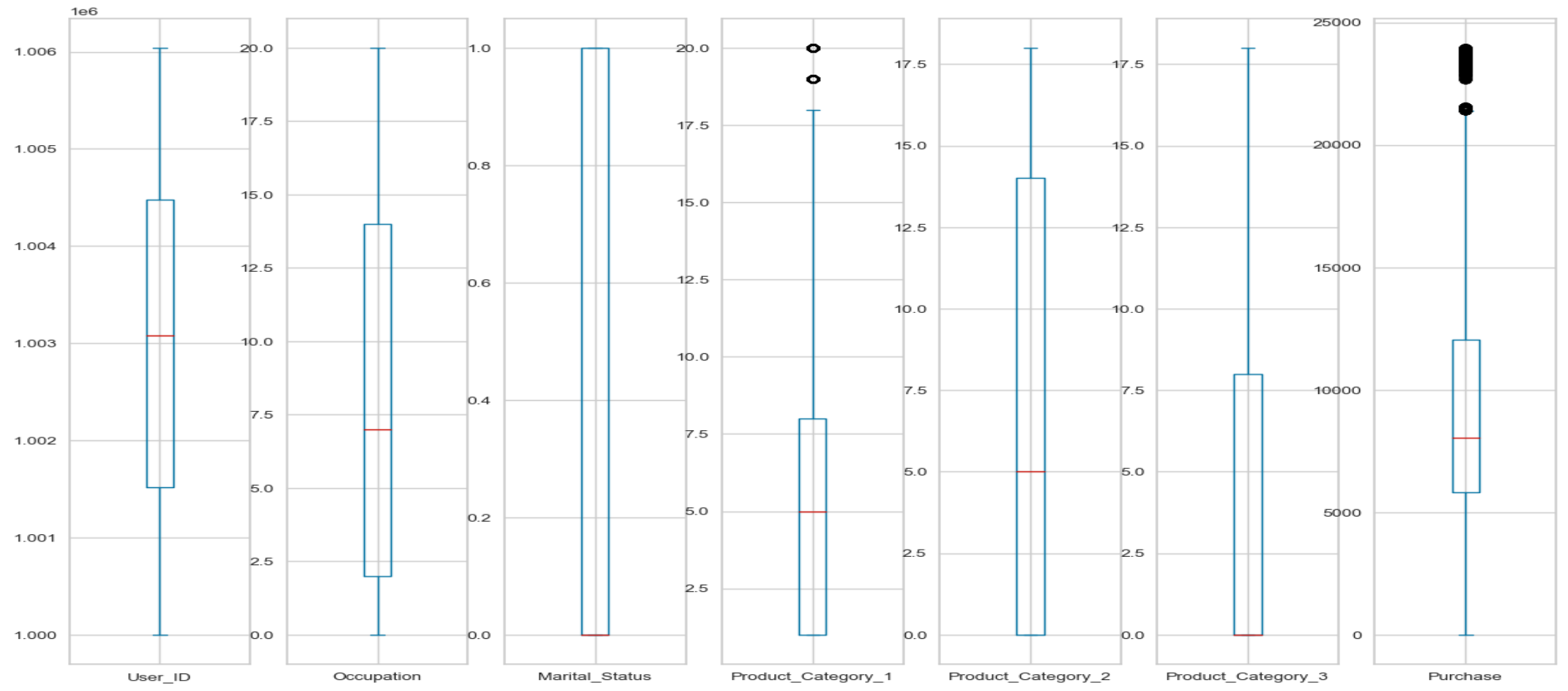
Data Pre-processing and Problems

Handling Missing Values



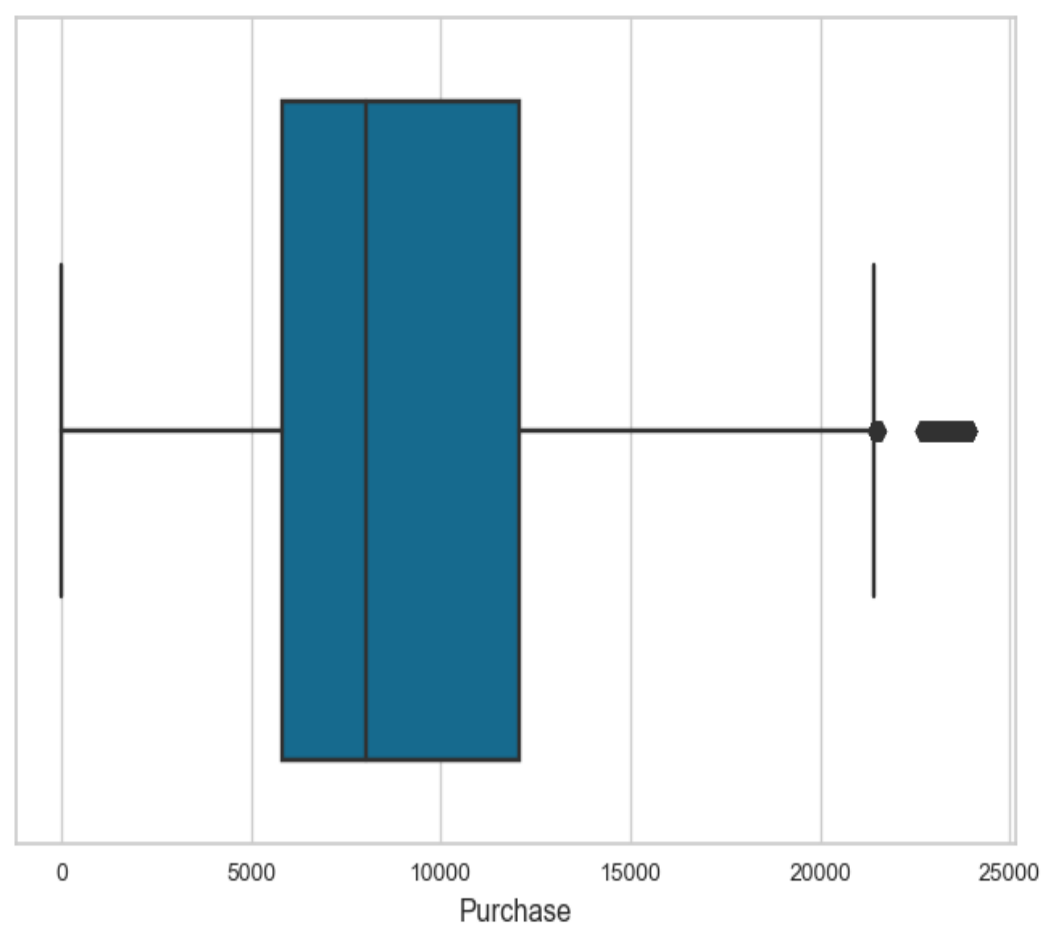
Data Pre-processing and Problems

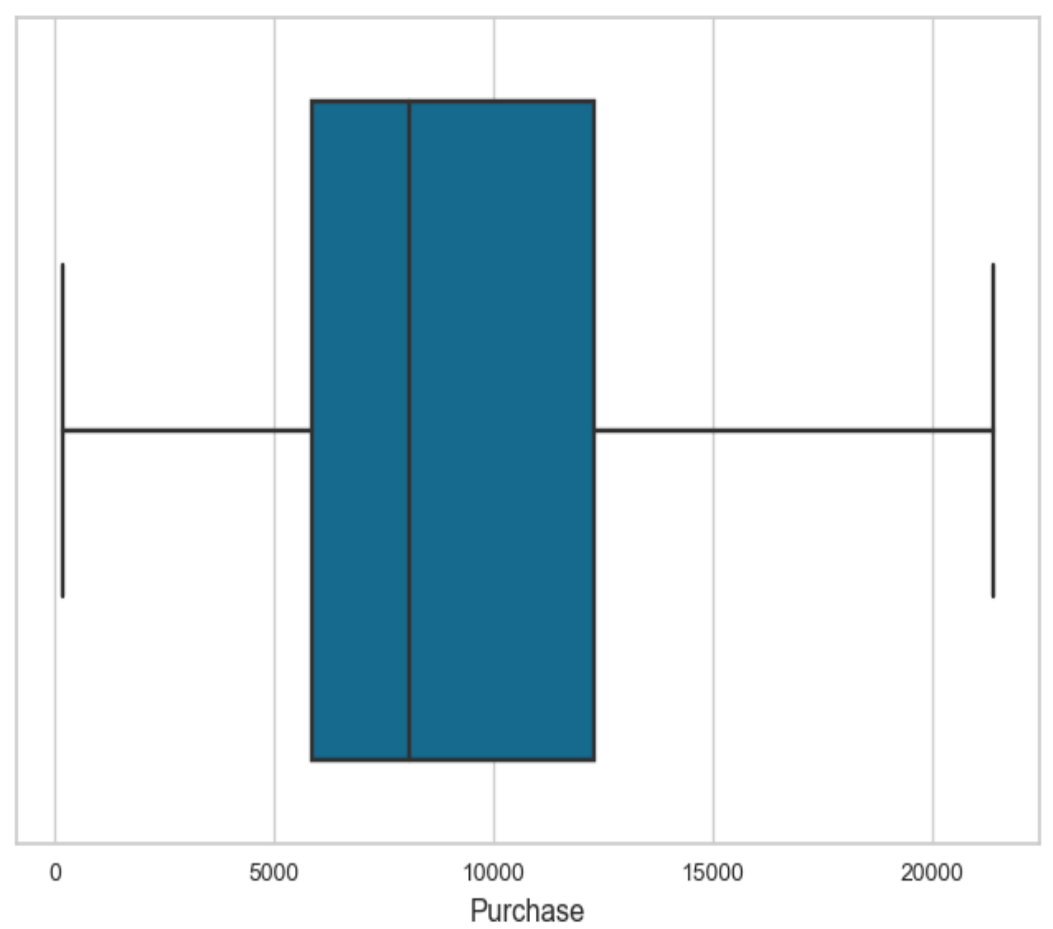
Handling Outliers



Data Pre-processing and Problems

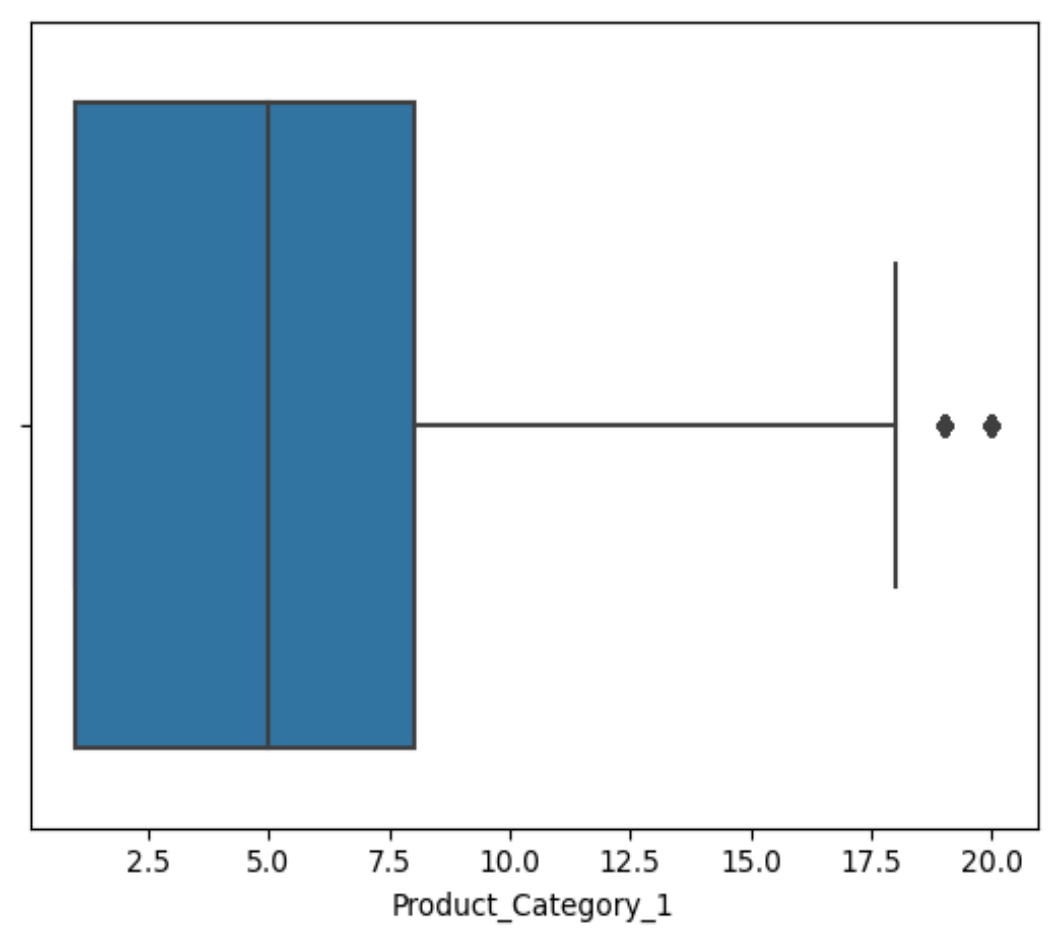
Handling Outliers

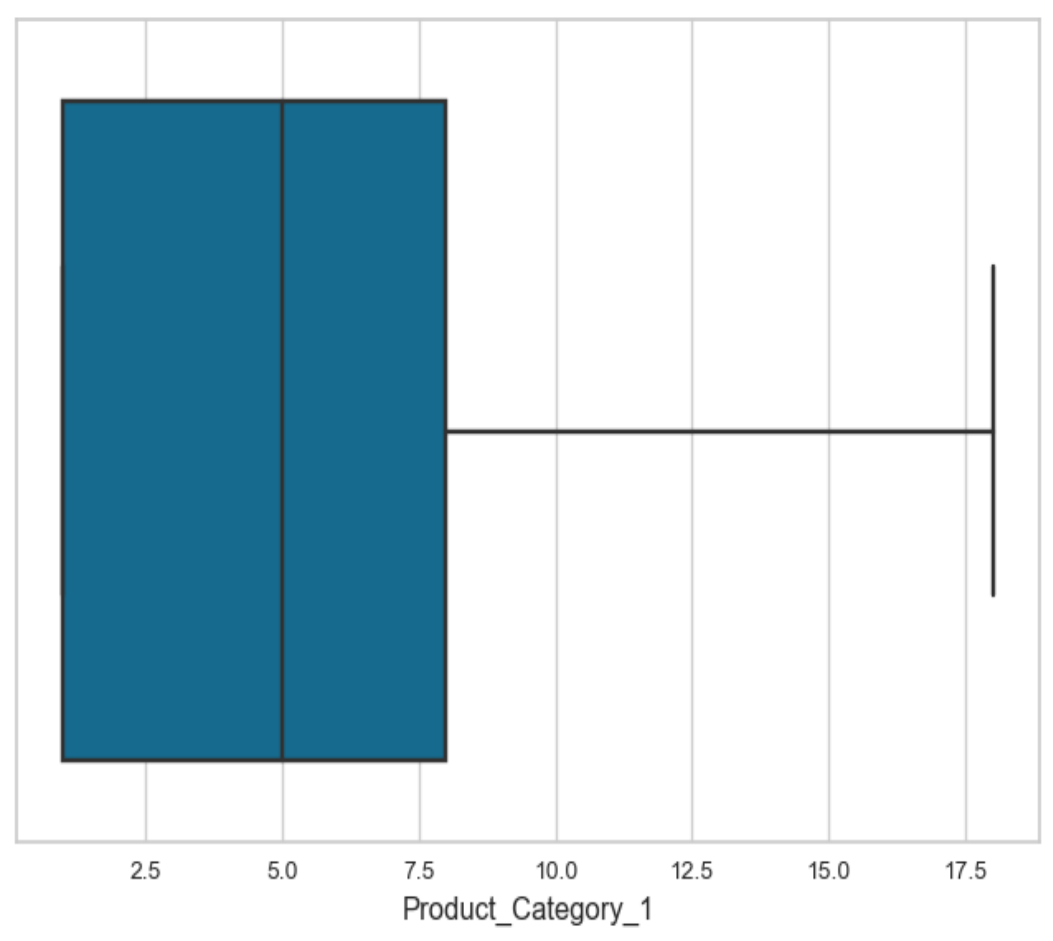




Data Pre-processing and Problems

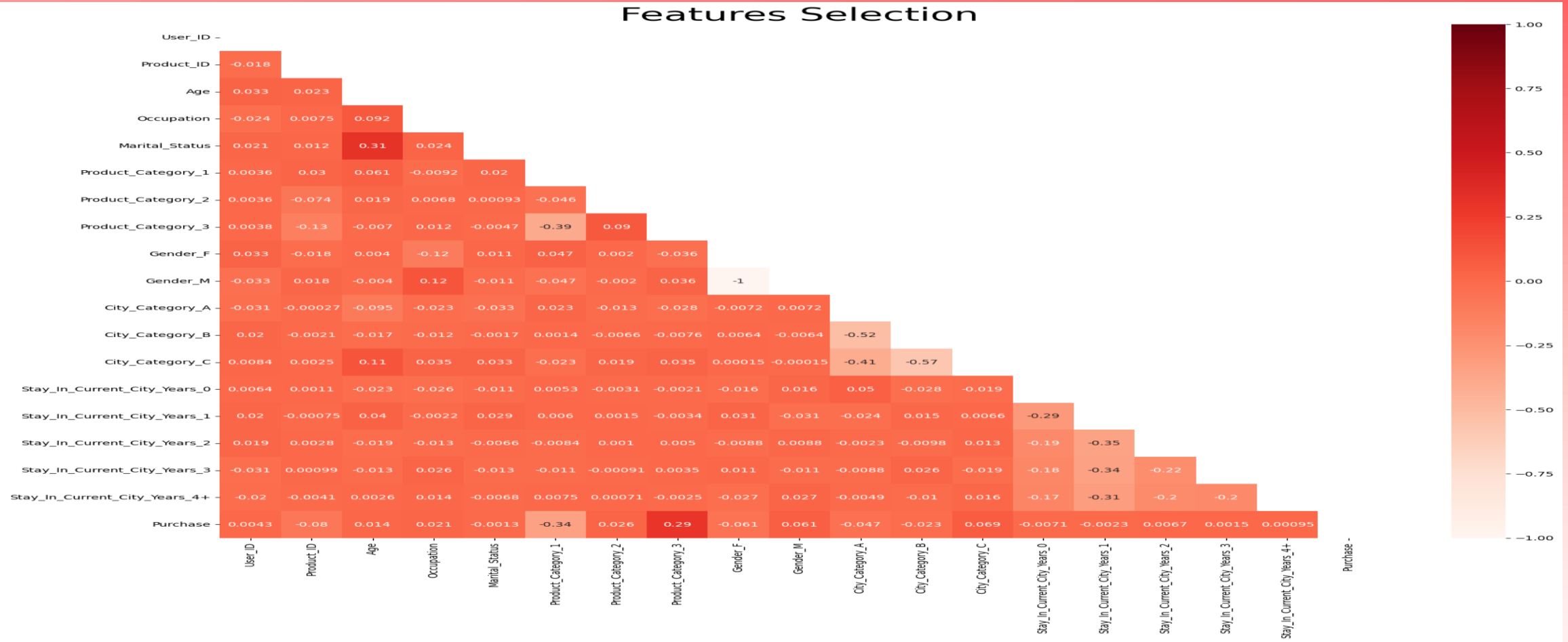
Handling Outliers





Data Pre-processing and Problems

Features Engineering



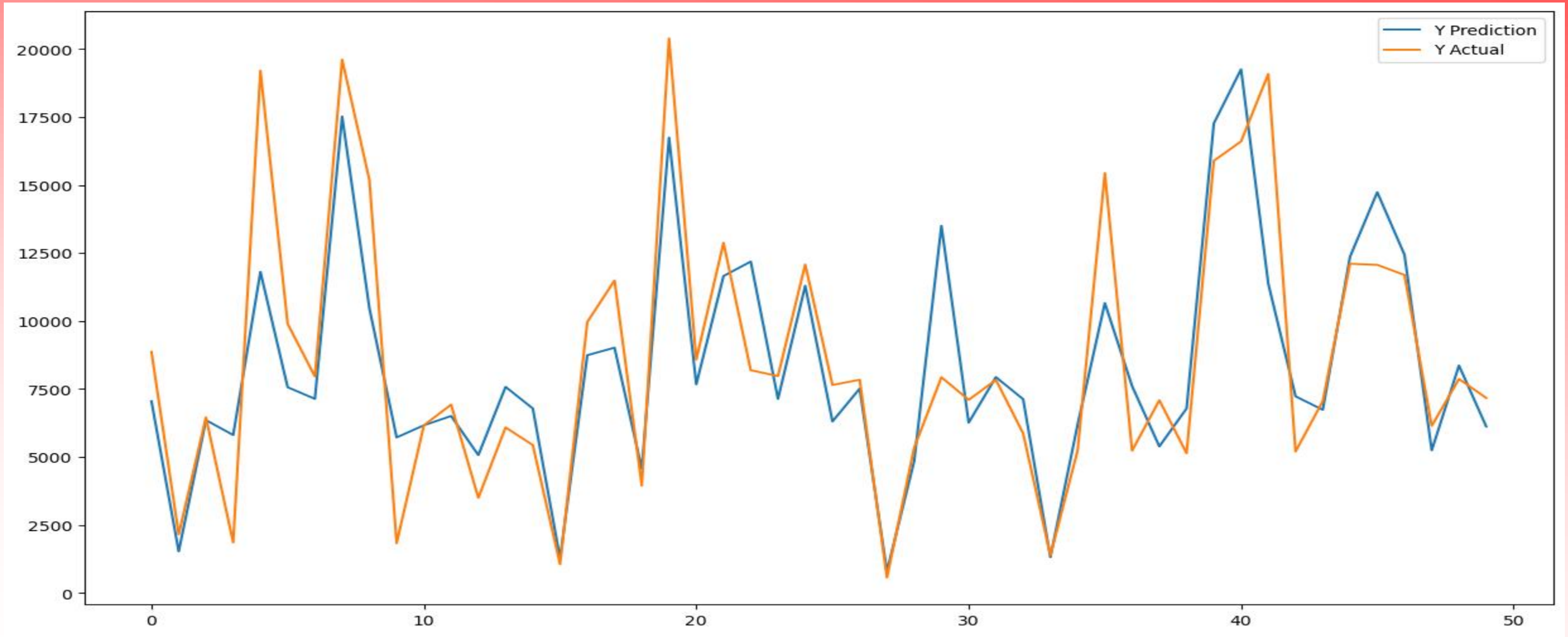
Modeling and Results

By using the Auto ML Library “Pycaret” on our data; we found that the following are the best models:

Algorithm Used (Model)	R2 Score	MAPE
Random Forrest Regressor	70%	0.301
Xtreme Gradient Boosting	71%	0.304

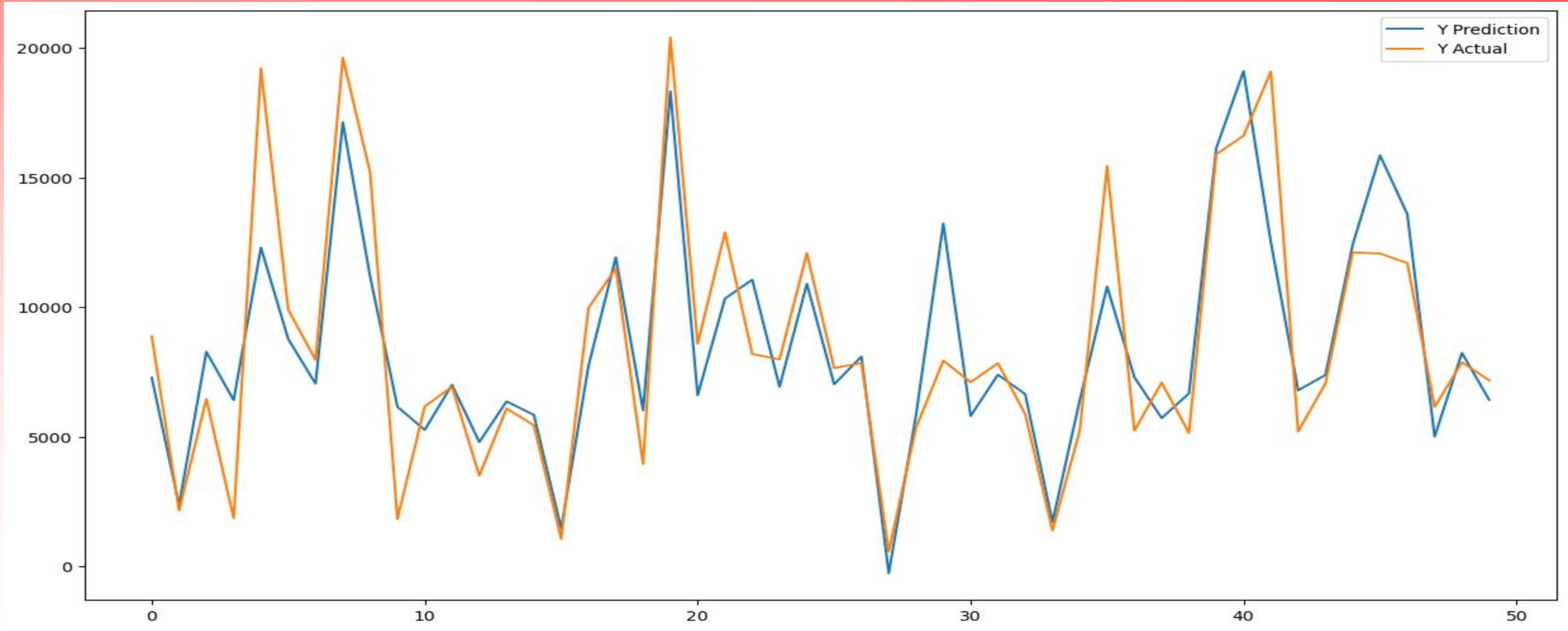
Modeling and Results

Actual vs Predicted Values Plot (Random Forrest Regressor)



Modeling and Results

Actual vs Predicted Values Plot (Xtreme Gradient Boosting)



Conclusion

- We have about 3631 different products
- We saw that the majority of the purchases have been done by the males.
- We saw that the majority of the purchases were done by young adults (26-35)
- We also noticed that some occupations had higher purchases.
- We also found single customers tend to but more than married ones.
- We found that there are certain categories that customers tend to buy more.
- We got the following insights from the top 3 customers:
  - They are all males.
  - 2 of them are from the young adults category (26-35), and the third is from the older adults category.
  - They all come from city category A.
  - 2 of them are single.
  - 2 of them didn't stay more than 1 year in the current city.
- We found that product categories had the most impact on the purchases.
- Eventually we managed to build the required model with about 70-72% R2 Score.

Thank You!