BLACK FRIDAY PREDICTION

Submitted by:

YASH JAISWAL

ACKNOWLEDGMENT

I would like to thank you Flip Robo Technologies for providing me with the opportunity to work on this project from which I have learned a lot

INTRODUCTION

- Business Problem Framing

Shopping are one of the necessary need of each and every person around the globe and therefore Black Friday Focusing on changing trends in Black friday sales and purchases predictive Discount market mix Dataset Using Visualization in order to predict the actual values of the how many sales prospective and decide whether to amount to the Product.

• Conceptual Background of the Domain Problem

Boost your business with industry-premium products and services, at prices that won't break your budget. If it doesn't provide you with a better Day of shopping Black Friday experience, we simply don't offer it.

• Review of Literature

This is a comprehensive summary of the research done on the behalf You have to data set at least 783667 rows of data. You can data set more data as well, it's up to you, More the data better the Visualization In this section you have to Black Friday of the data Black Friday Prediction.

• Motivation for the Problem Undertaken

It traditionally marks the start of the Christmas shopping season in the united states many stires offer highly promoted sales at discounted prices often open early sometimes as early as midnight or even on Thankgiving.

Analytical Problem Framing

• Mathematical/ Analytical Modeling of the Problem

```
# lets check the type of dataset
data.dtypes
```

```
User_ID                       int64
Product_ID                   object
Gender                       object
Age                          object
Occupation                    int64
City_Category                object
Stay_In_Current_City_Years   object
Marital_Status                int64
Product_Category_1            int64
Product_Category_2          float64
Product_Category_3          float64
Purchase                    float64
dtype: object
```

• Data Sources and their formats

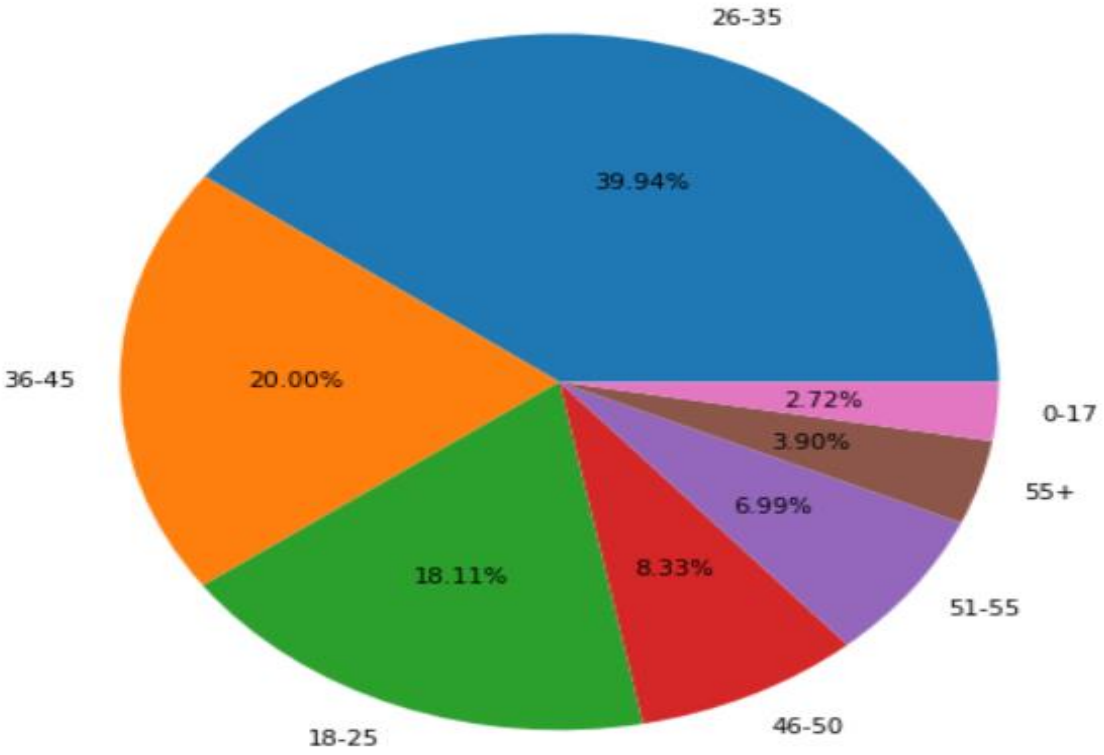
```
# Statistical summary
data.describe()
```

	Occupation	Stay_In_Current_City_Years	Marital_Status	Product_Category_1	Product_Category_2	Purchase
count	783667.000000	783667.000000	783667.000000	783667.000000	783667.000000	783667.000000
mean	8.079300	1.858247	0.409777	5.366196	9.579427	9263.968713
std	6.522206	1.288790	0.491793	3.878160	4.233575	4208.342958
min	0.000000	0.000000	0.000000	1.000000	2.000000	12.000000
25%	2.000000	1.000000	0.000000	1.000000	8.000000	6996.000000
50%	7.000000	2.000000	0.000000	5.000000	9.000000	9263.968713
75%	14.000000	3.000000	1.000000	8.000000	14.000000	9927.000000
max	20.000000	4.000000	1.000000	20.000000	18.000000	23961.000000

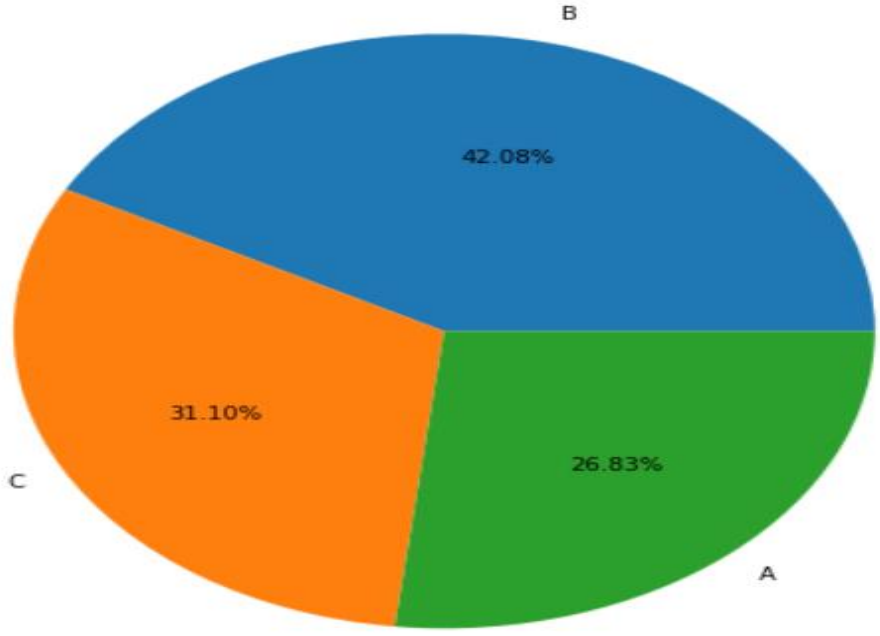
- Data Preprocessing Done

Loading the test and train data set as a dataframe , used pandas to set display I ensuring we do not see any truncated information , checked the number of rows and columns present in our test and train data set , checked for missing data and the number of rows with null values , verified the percentage of missing data in each columns are decide to dicard the once that value more than , dropped all the unwanted columns are duplicated data present in our data frame, separated categorical columns and numeric columns name in separate list variable for ease in visulazation , checked the unique values information in each column to get a gist for categorical data. Used pandas profiling during the visulaziation phase along with pie plot count plot scatter plot and the other, with the help of label encoding technique converted all object data type columns to numeric data types.

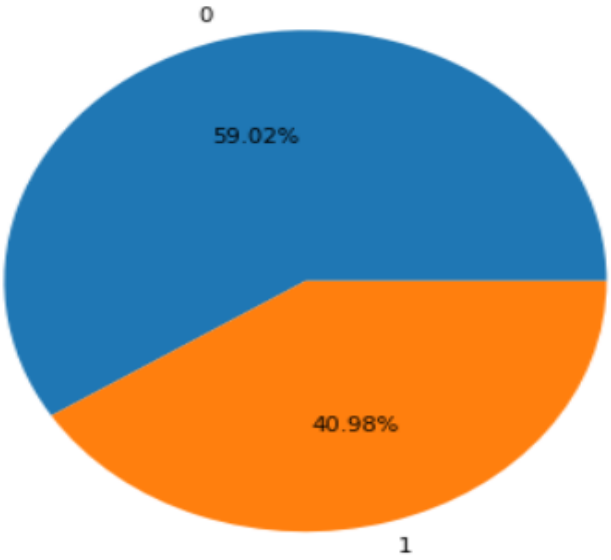
Age distribution



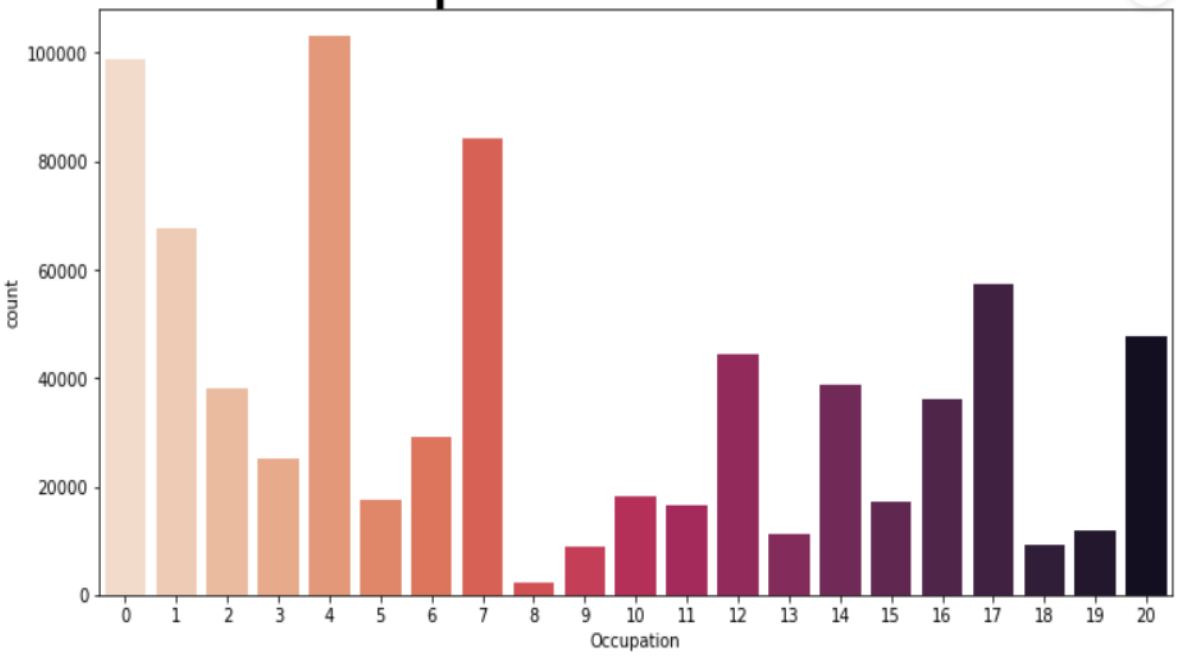
City Category distribution

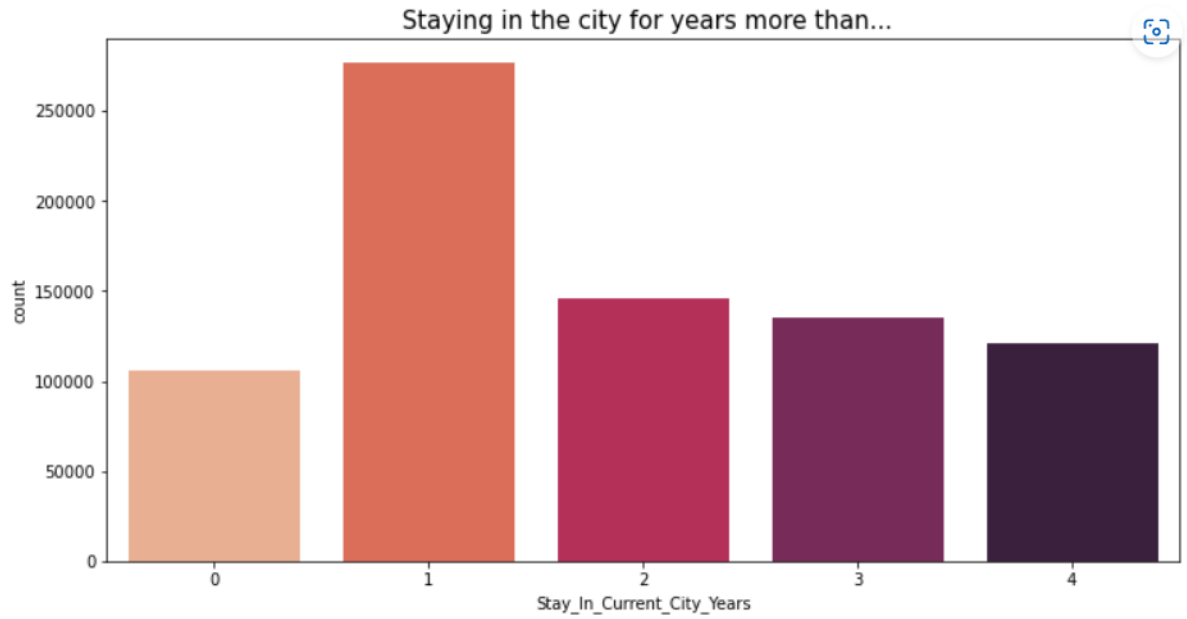


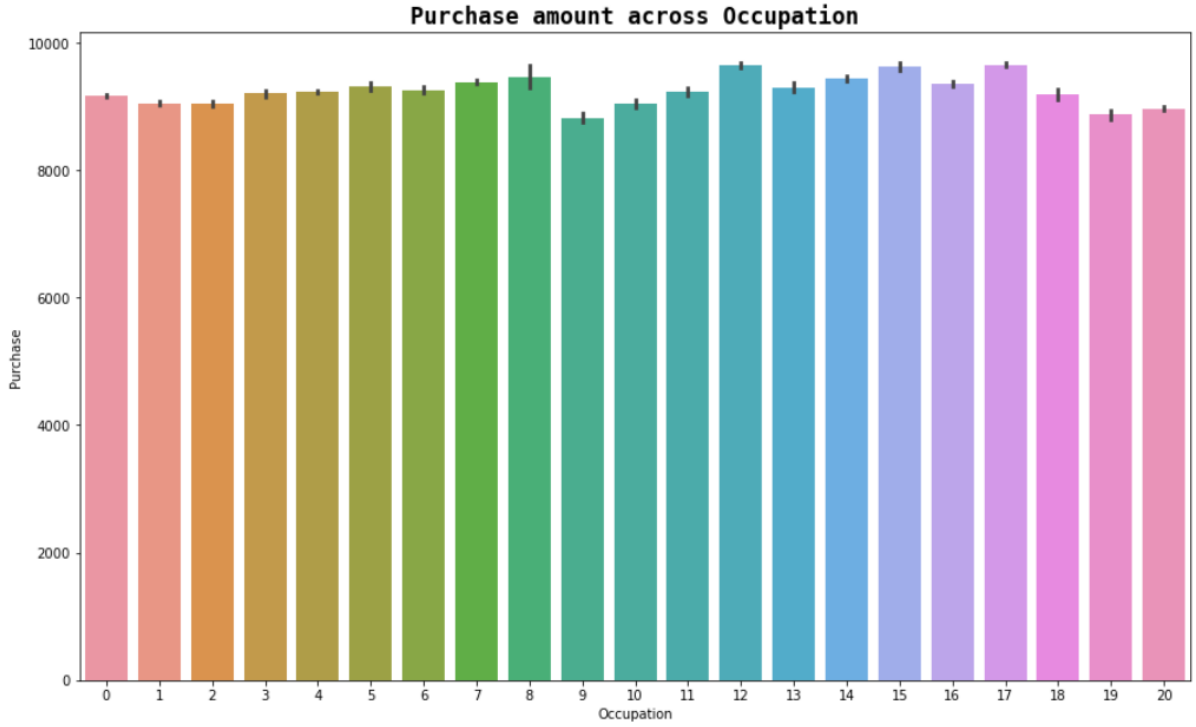
Marital Status distribution

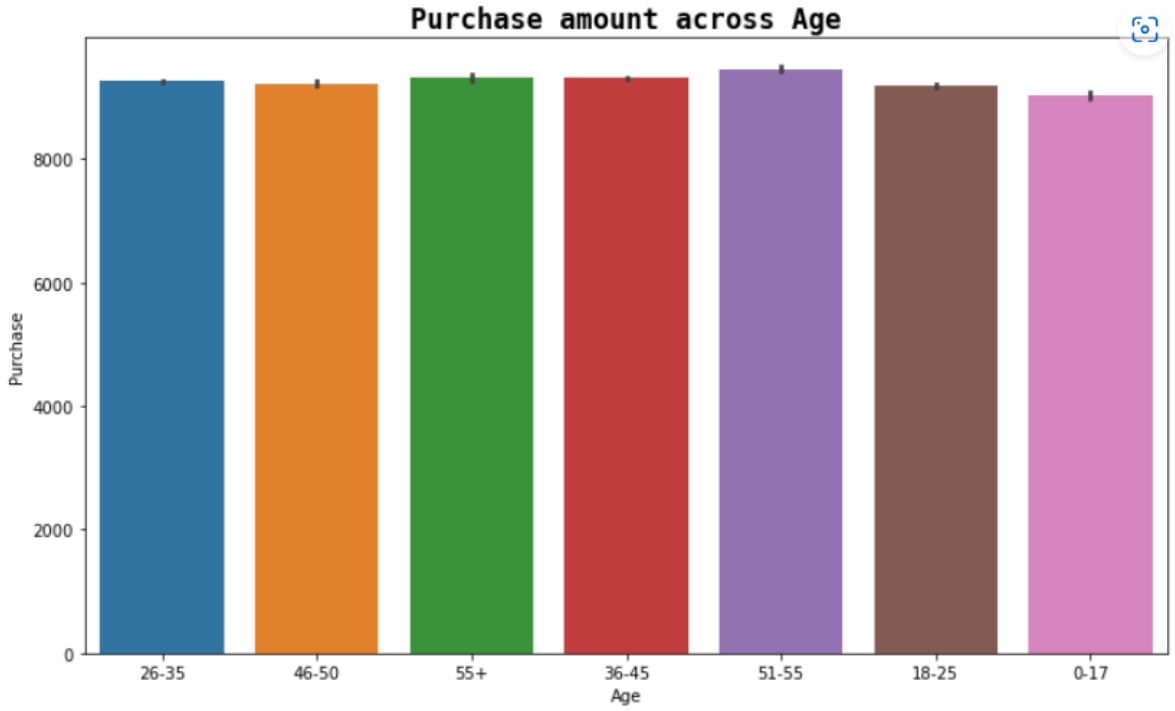


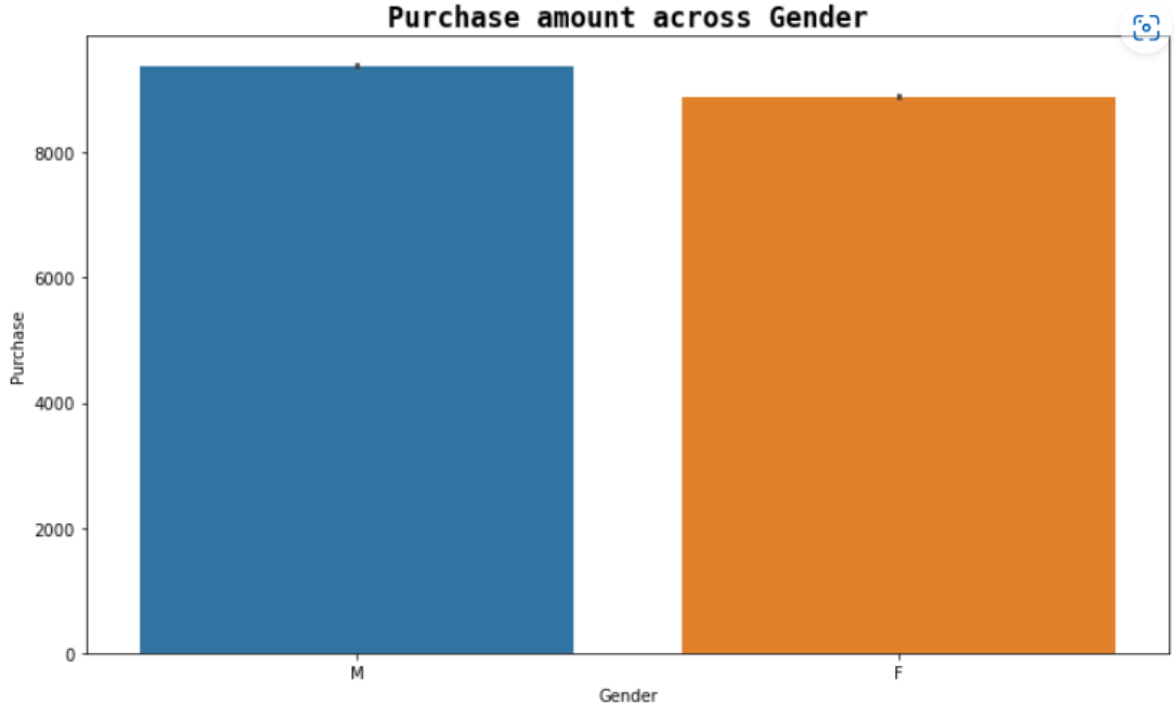
Occupation distribution



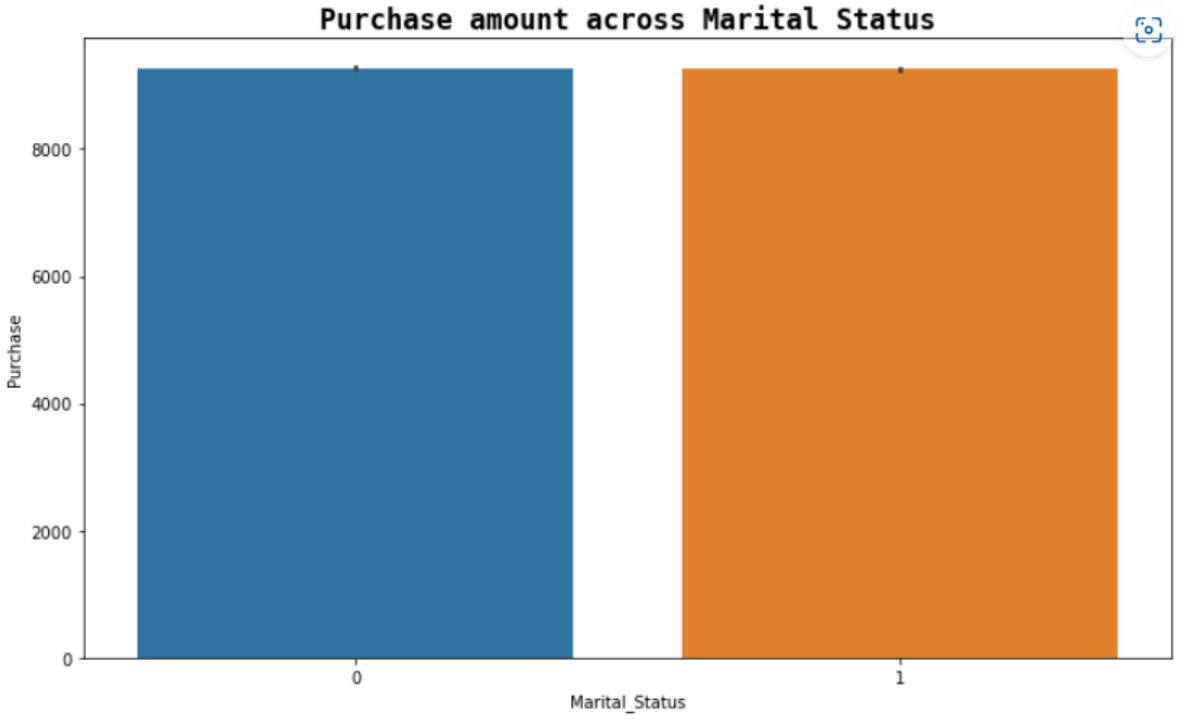

Staying in the city for years more than...
count
250000
200000
150000
100000
50000
0
0
1
2
3
4
Stay_In_Current_City_Years

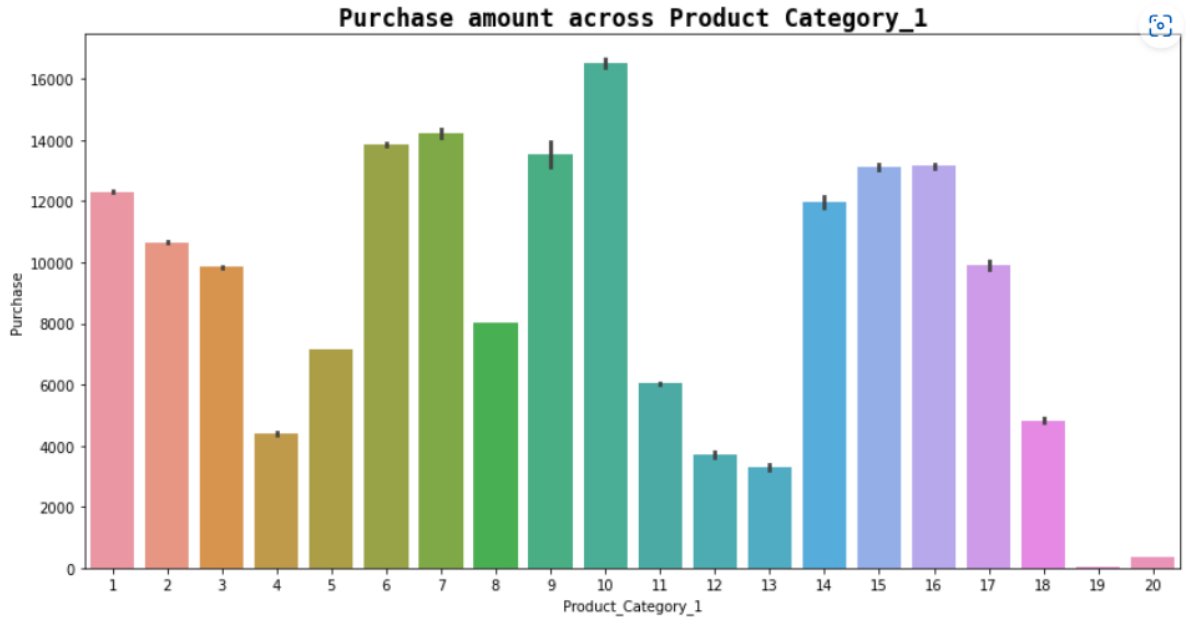

Purchase amount across Occupation
Purchase
10000
8000
6000
4000
2000
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Occupation

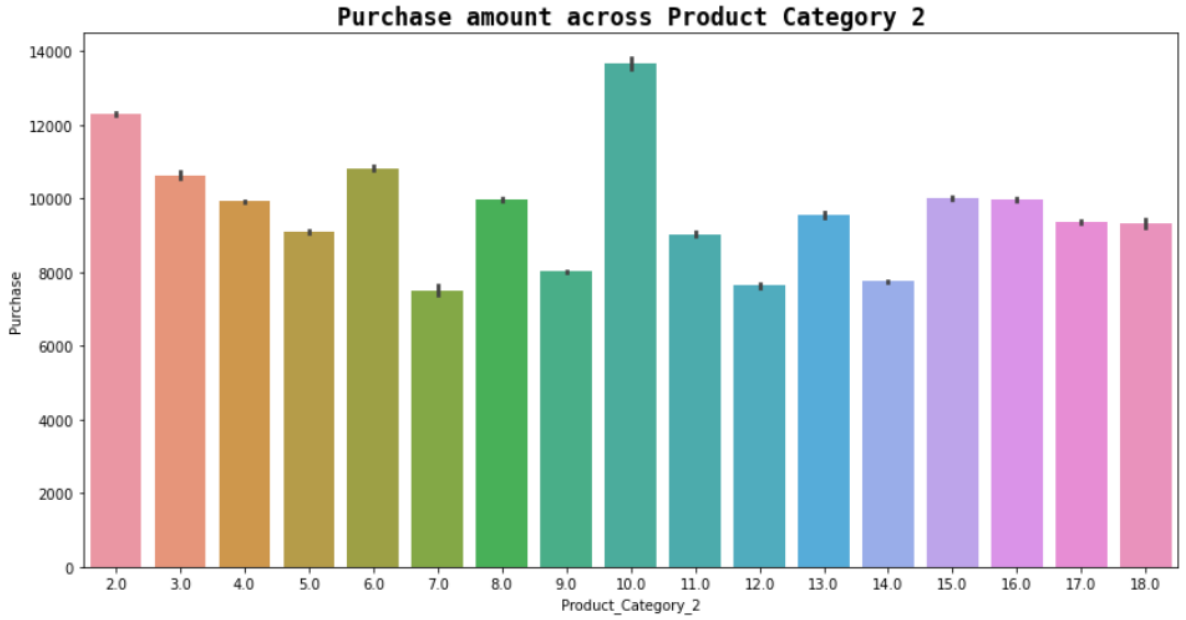

Purchase amount across Age
Purchase
0
2000
4000
6000
8000
26-35
46-50
55+
36-45
51-55
18-25
0-17
Age


Purchase amount across Gender
Purchase
0
2000
4000
6000
8000
M
F
Gender


Purchase amount across City Category
10000
8000
6000
4000
2000
0
Purchase
A
C
B
City_Category


Purchase amount across Marital Status
8000
6000
4000
2000
0
Purchase
0
1
Marital_Status





- not to create machine learning models because Flip Robo Technologies says only findings and conclusions in detailed data analysis .

CONCLUSION

After visualizing the data. I found Black Friday is the best online shopping day because Male and female both of use to shopping more. customer satisfaction and customer trust appeared as the outcomes of overall Retail Company 24 November is Black Friday shop are discount 40% to 50%.