Black Friday Customer Purchase Prediction

by: Farzad Radmehr

Problem Statement

- Predicting customer purchase behavior with 3 product categories at store

- So the goal is to select the best product categories in order to increase the revenue

Data Description

- 550,000 observations in a retail store in black Friday
- Different types of variables either numerical or categorical
- Product Categories 2 and 3 have missing values
- No outliers detected

- ***Inputs:***

Gender

Age

Occupation

City Category

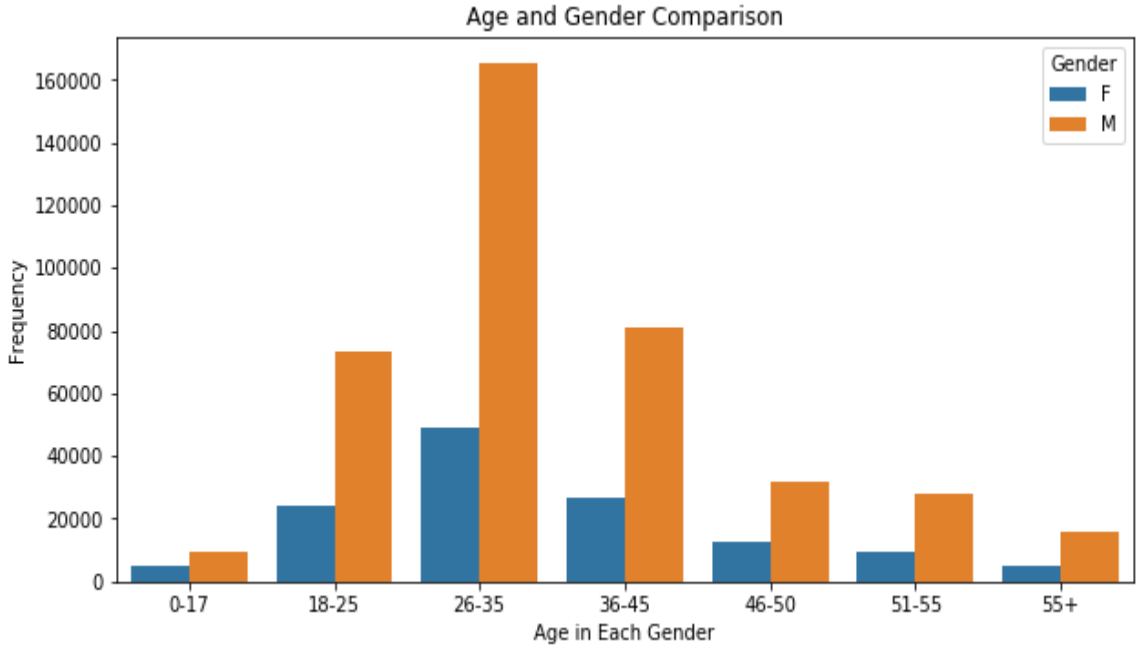
# of years of staying in current city

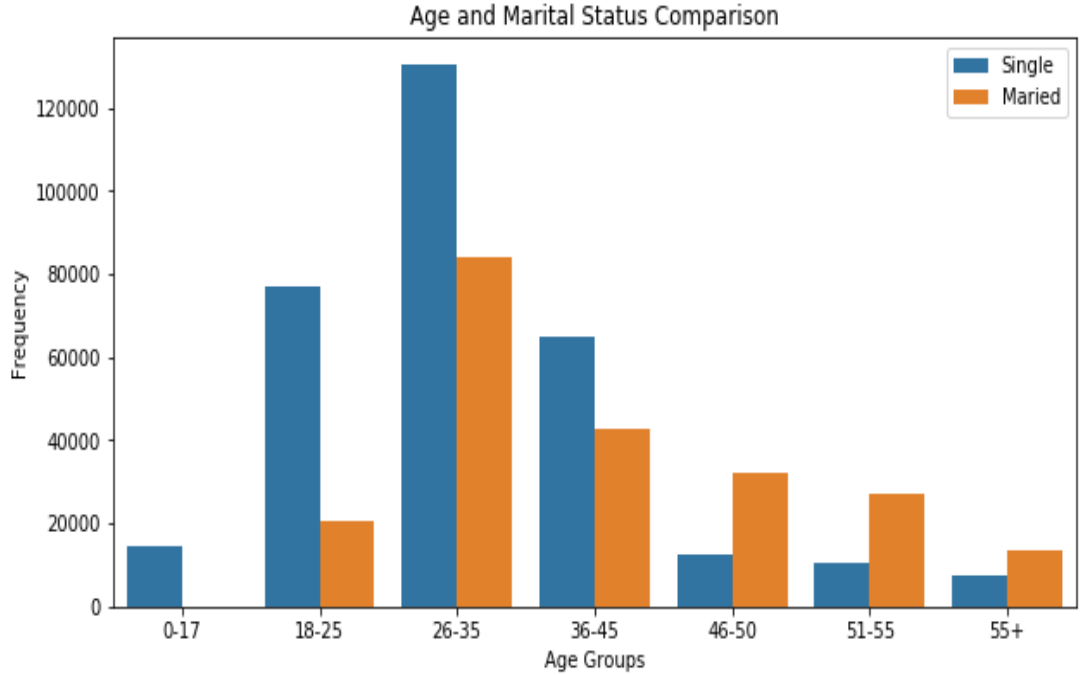
Product Category 1

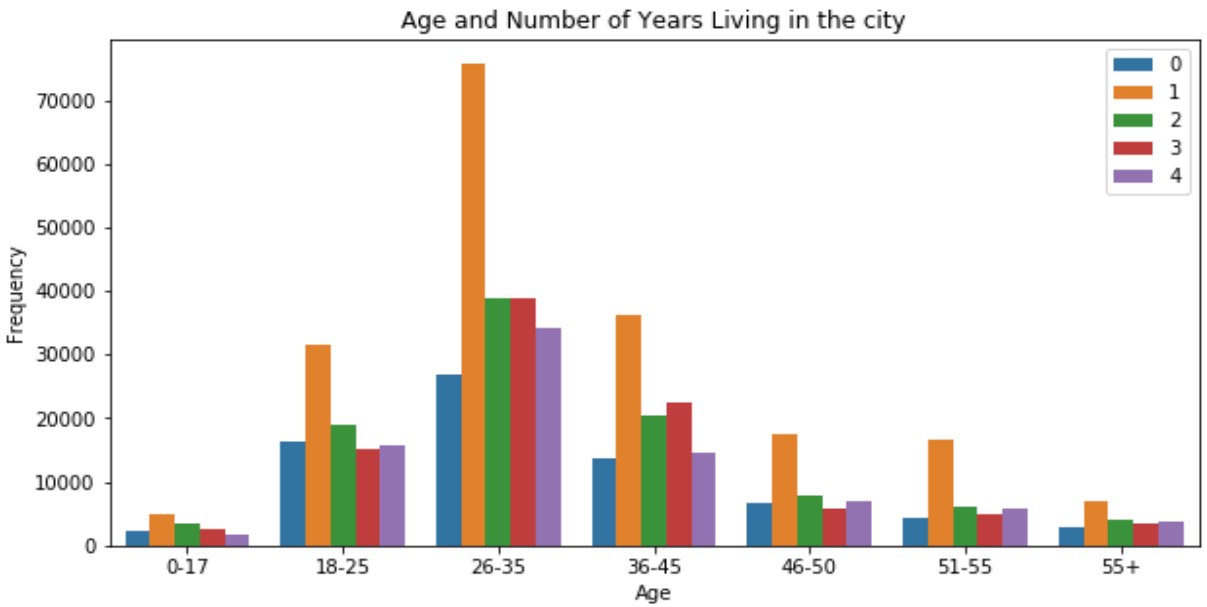
Product Category 2

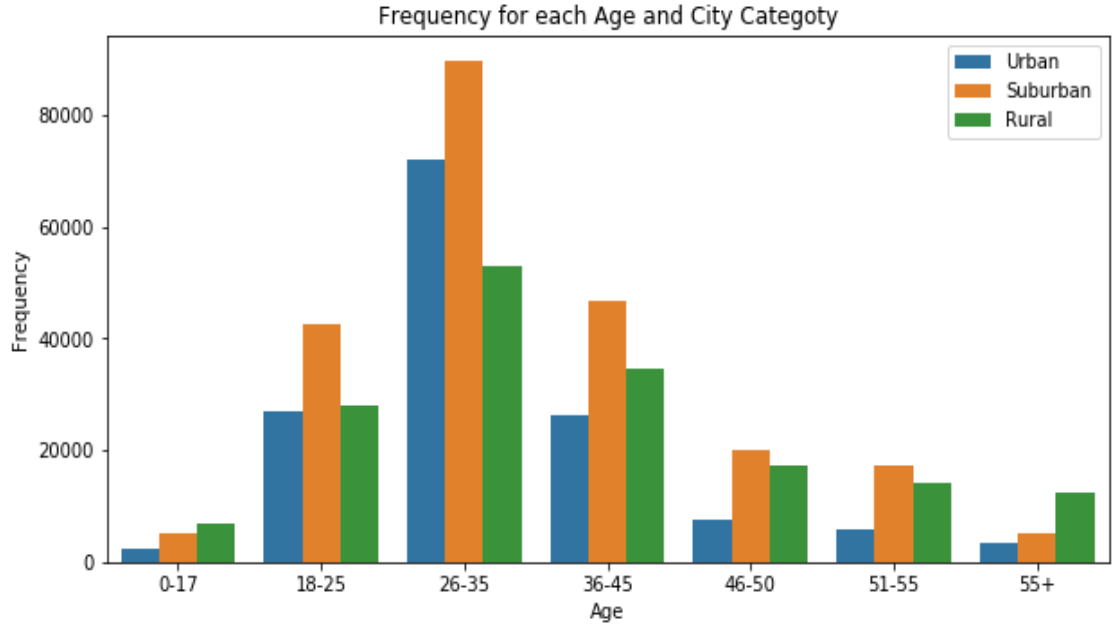
Product Category 3

- ***Output:***
  Purchase

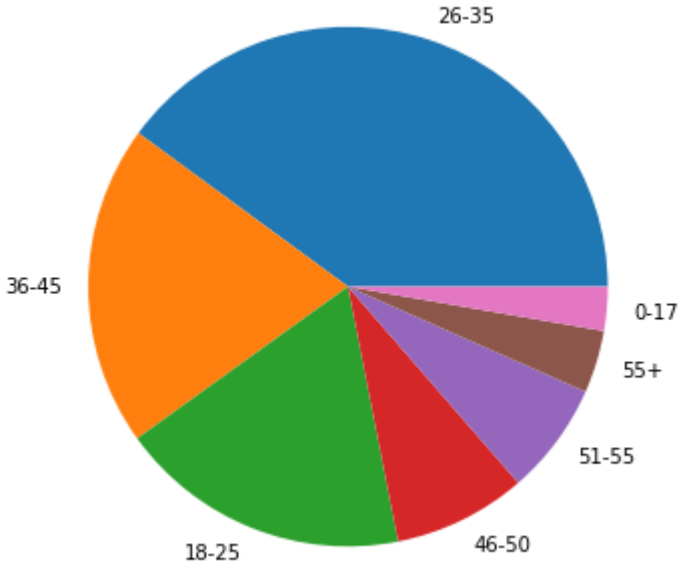

Age and Gender Comparison
Gender
F
M
Frequency
0
20000
40000
60000
80000
100000
120000
140000
160000
0-17
18-25
26-35
36-45
46-50
51-55
55+
Age in Each Gender


Age and Marital Status Comparison
Single
Maried
Frequency
0
20000
40000
60000
80000
100000
120000
0-17
18-25
26-35
36-45
46-50
51-55
55+
Age Groups

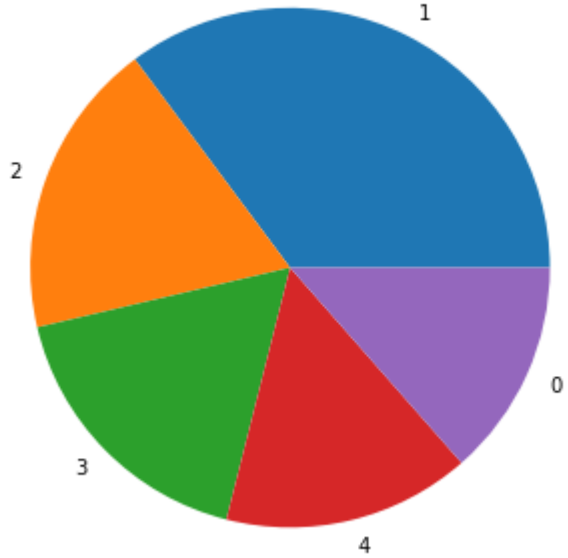

Age and Number of Years Living in the city
0
1
2
3
4
Frequency
0
10000
20000
30000
40000
50000
60000
70000
0-17
18-25
26-35
36-45
46-50
51-55
55+
Age


Frequency for each Age and City Categoty
Urban
Suburban
Rural
Frequency
0
20000
40000
60000
80000
0-17
18-25
26-35
36-45
46-50
51-55
55+
Age

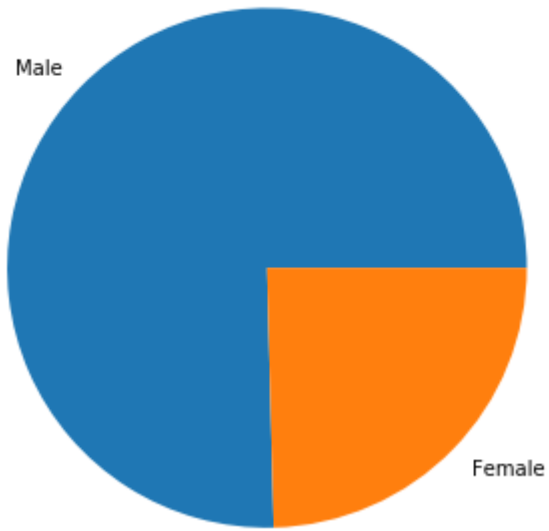
Age Percentage



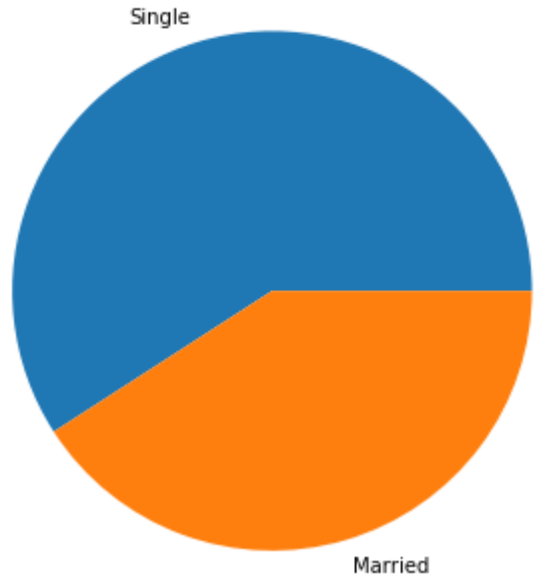
Length of Stay in City Percentage

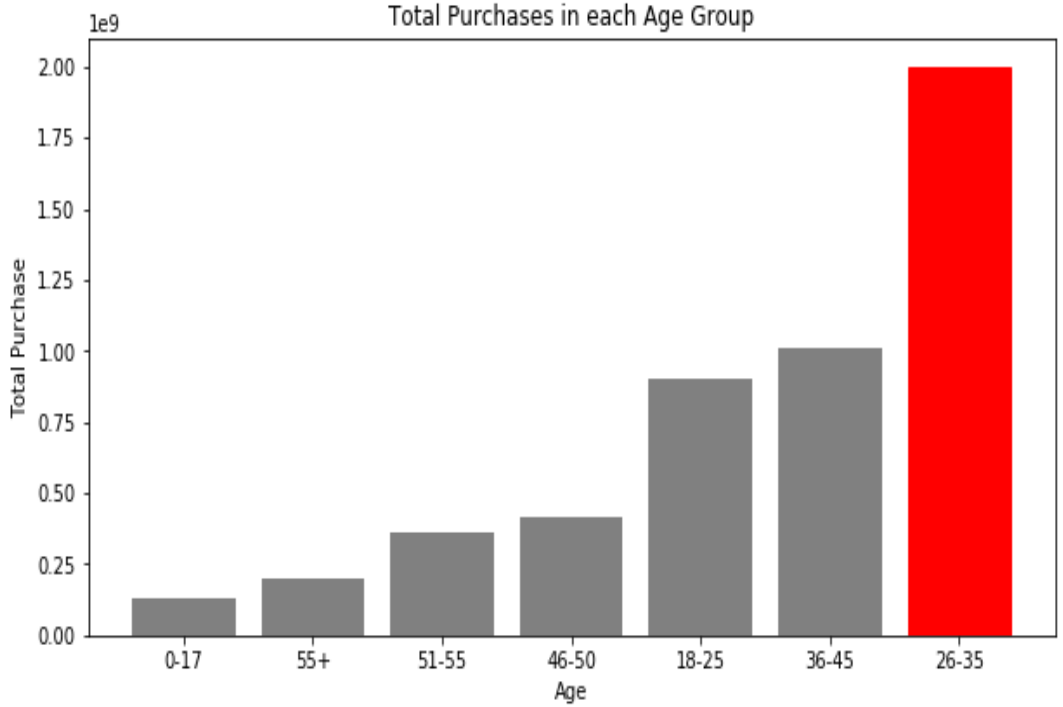


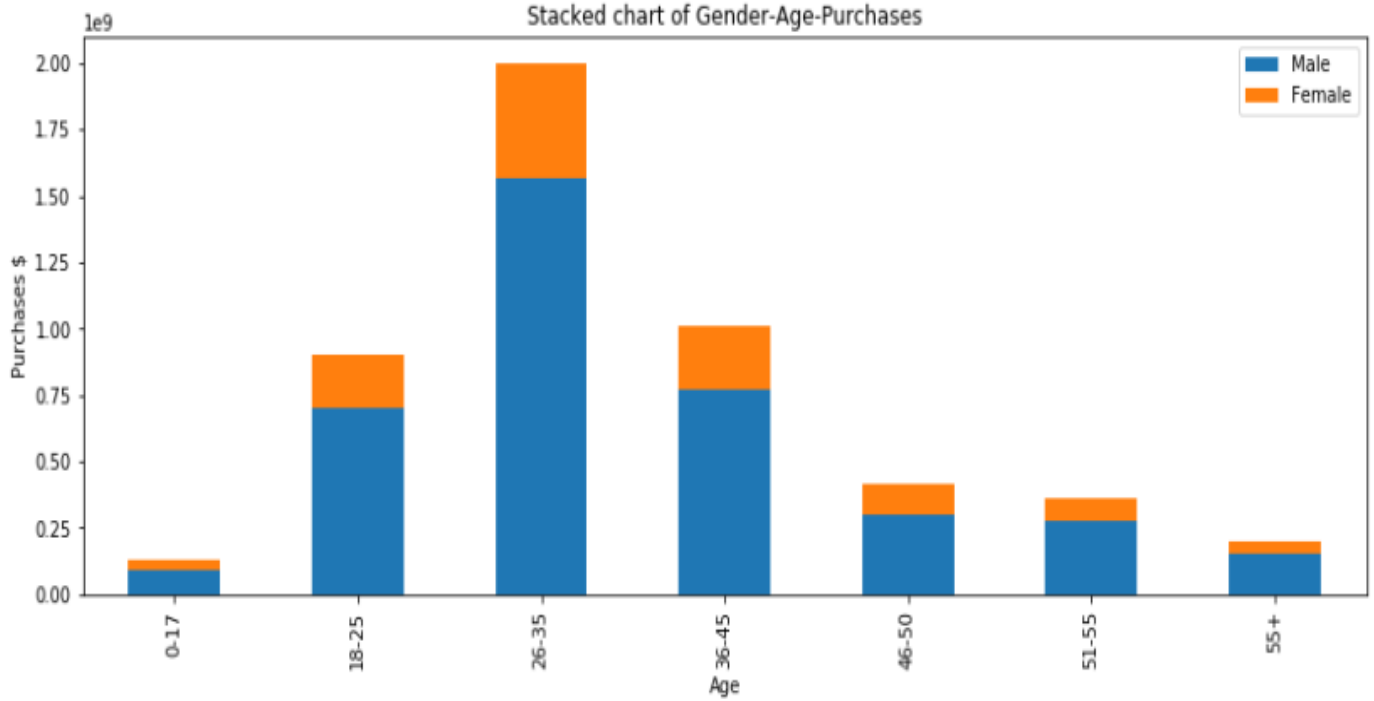
Gender Percentage



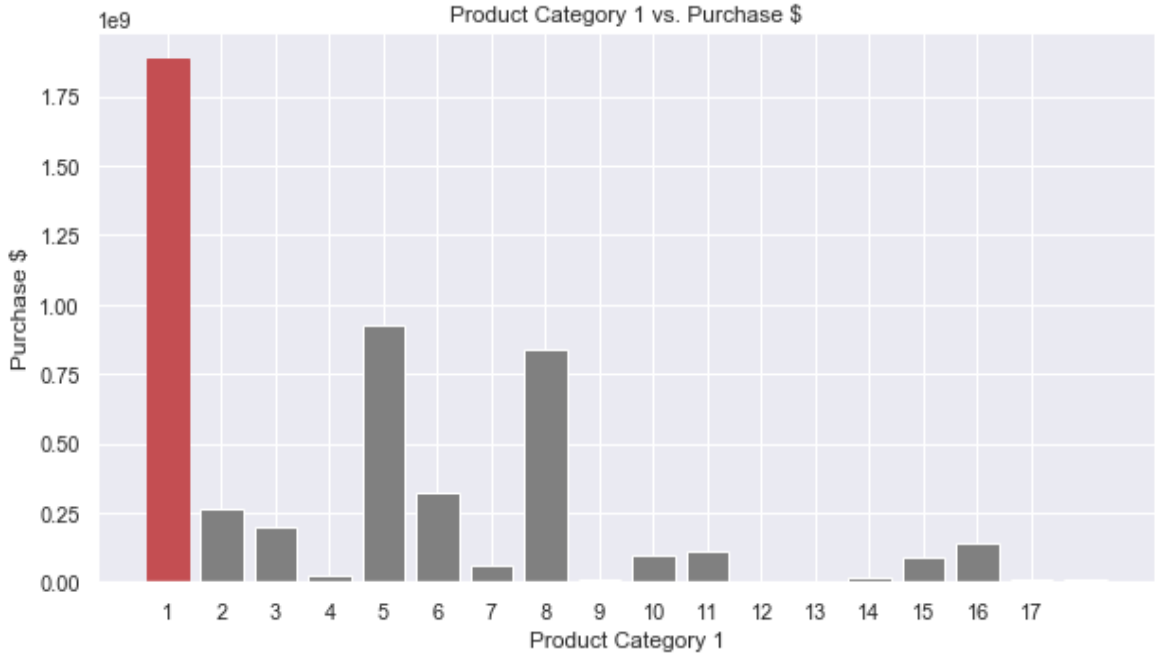
Marital_Status Percentage

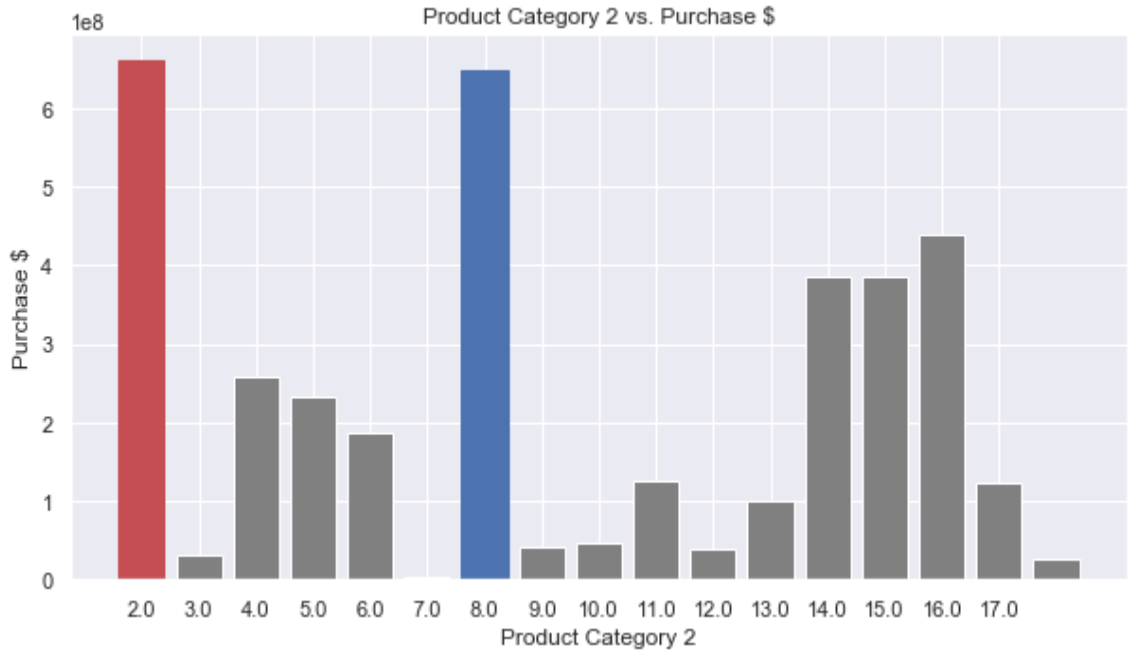


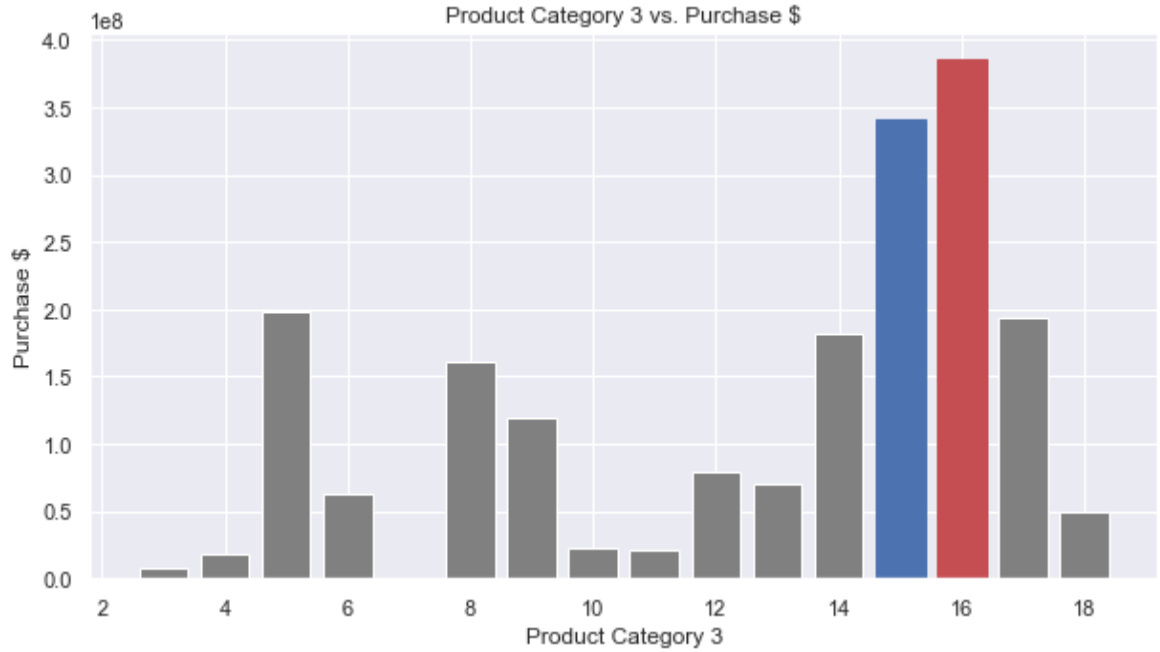

Total Purchases in each Age Group
1e9
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Total Purchase
0-17
55+
51-55
46-50
18-25
36-45
26-35
Age


Stacked chart of Gender-Age-Purchases
1e9
Male
Female
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Purchases $
0-17
18-25
26-35
36-45
46-50
51-55
55+
Age


Purchases among Occupations
1e8
6
5
4
3
2
1
0
Purchase $
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Occupation


Product Category 1 vs. Purchase $
1e9
Purchase $
Product Category 1


Product Category 2 vs. Purchase $
1e8
Purchase $
Product Category 2


Product Category 3 vs. Purchase $
1e8
Purchase $
Product Category 3

People from rural areas paid significantly higher.

People in age group 26-35 paid significantly highest.

No difference between single and married.

product category 3 has the highest impact on total purchases.

Categories	Results 1	Results 2
Product Category 1	Women are signicantly more interested	Singles are significantly more interested
Product Category 2	Both men and women have same interests	Both singles and married ones have same interest
Product Category 3	Men are signifantly more interested	singles are signicantly more interested

Machine Learning Techniques

Method	Features	Parameters	Training RMSE	Test RMSE
Linear Regression	Gender, Age, Occupation, City Category, Current City # of years, Marital Status, Product Category 1,2,3	Test % = 30, Random State = 42	4630.9	4633.8
Decision Tree Regressor	Gender, Age, Occupation, City Category, Current City # of years, Marital Status, Product Category 1,2,3	Test % = 30, Random State = 42	2960.2	2983.8
Random Forest Regressor	Gender, Age, Occupation, City Category, Current City # of years, Marital Status, Product Category 1,2,3	Test % = 30, Random State = 42	2945.1	2968.4

Conclusion

- women like category 1 AND men like category 3.
- Singles like the store more.
- People in the first year of living are interested to this store, but loose interest in second year.
- People of rural areas buy more in this store.
- Regression techniques can be used to predict the purchase.
- Random Forest Regressor has the best performance among other techniques.

Next Steps

2 steps:

Step 1: Add new categories of products to make the following groups interested to the store:

- Married people
- People who live more than one year in this area
- People of urban and suburban areas
- Other age groups

Step 2: Re-do this research to review the changes.

Thank you

https://github.com/fradmehr/Springboard-Capstone-Project-1

https://www.linkedin.com/in/farzad-radmehr-97132143/