MSIS 2631

Claire Zhang, Christopher Wong,
Dhanashree Mane, Faliha Zikra

Problem Statement

"We are a retail company "ABC Private Limited", want to understand the customer purchase behavior against various products of different categories and improve customer satisfaction along with the sales of the company."

Data Summary

	A	B	C	D	E	F	G	H	I	J	K	L
1	User_ID	Product_ID	Gender	Age	Occupation	City_Category	Stay_In_Current_City_Years	Marital_Status	Product_Category_1	Product_Category_2	Product_Category_3	Purchase
2	1000001	P00069042	F	0-17	10	A	2	0	3			8370
3	1000001	P00248942	F	0-17	10	A	2	0	1	6	14	15200
4	1000001	P00087842	F	0-17	10	A	2	0	12			1422
5	1000001	P00085442	F	0-17	10	A	2	0	12	14		1057
6	1000002	P00285442	M	55+	16	C	4+	0	8			7969
7	1000003	P00193542	M	26-35	15	A	3	0	1	2		15227
8	1000004	P00184942	M	46-50	7	B	2	1	1	8	17	19215
9	1000004	P00346142	M	46-50	7	B	2	1	1	15		15854
10	1000004	P0097242	M	46-50	7	B	2	1	1	16		15686
11	1000005	P00274942	M	26-35	20	A	1	1	8			7871
12	1000005	P00251242	M	26-35	20	A	1	1	5	11		5254
13	1000005	P00014542	M	26-35	20	A	1	1	8			3957
14	1000005	P00031342	M	26-35	20	A	1	1	8			6073
15	1000005	P00145042	M	26-35	20	A	1	1	1	2	5	15665
16	1000006	P00231342	F	51-55	9	A	1	0	5	8	14	5378
17	1000006	P00190242	F	51-55	9	A	1	0	4	5		2079
18	1000006	P0096642	F	51-55	9	A	1	0	2	3	4	13055
19	1000006	P00058442	F	51-55	9	A	1	0	5	14		8851
20	1000007	P00036842	M	36-45	1	B	1	1	1	14	16	11788
21	1000008	P00249542	M	26-35	12	C	4+	1	1	5	15	19614

Exploratory Data Analysis

Item prices ranged from 12 - 23691 dollars.

Each item had a different selling price - due to discounts

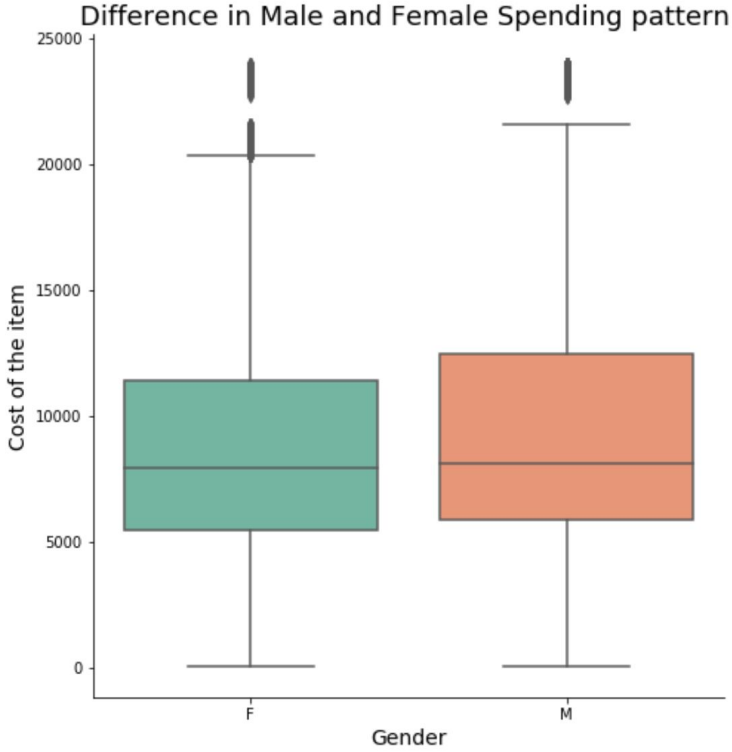
Customers could get 0-80 percent discount on items.

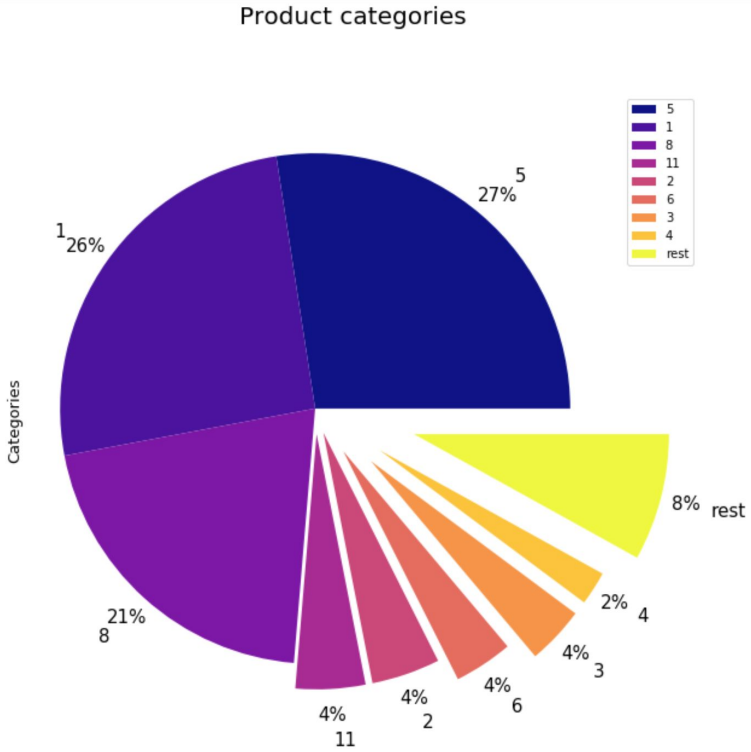
Most of the customers were Men.

Top 10 spenders : Bought 979 items - Spent 10 million dollars.

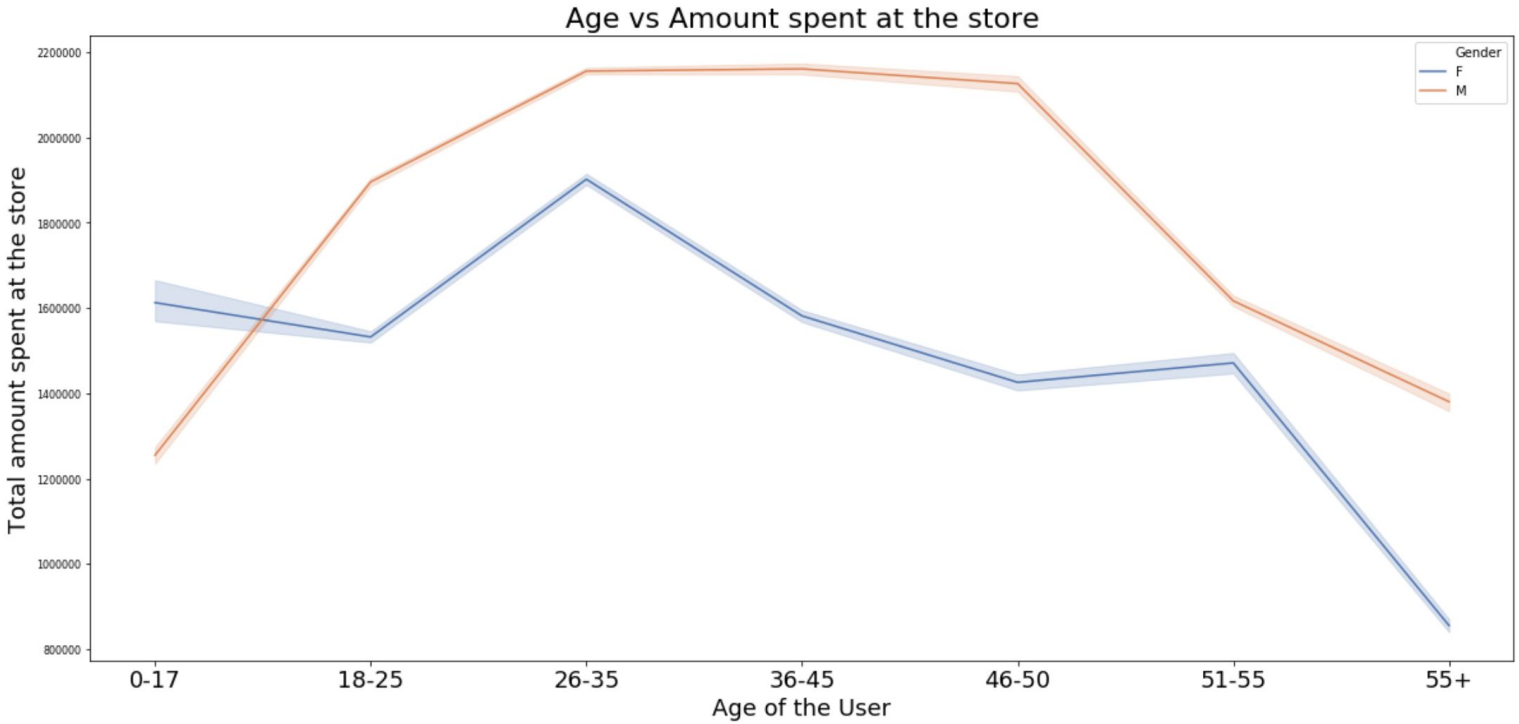
Men, Single, 36-45 , Occupation 16

Exploratory Data Analysis

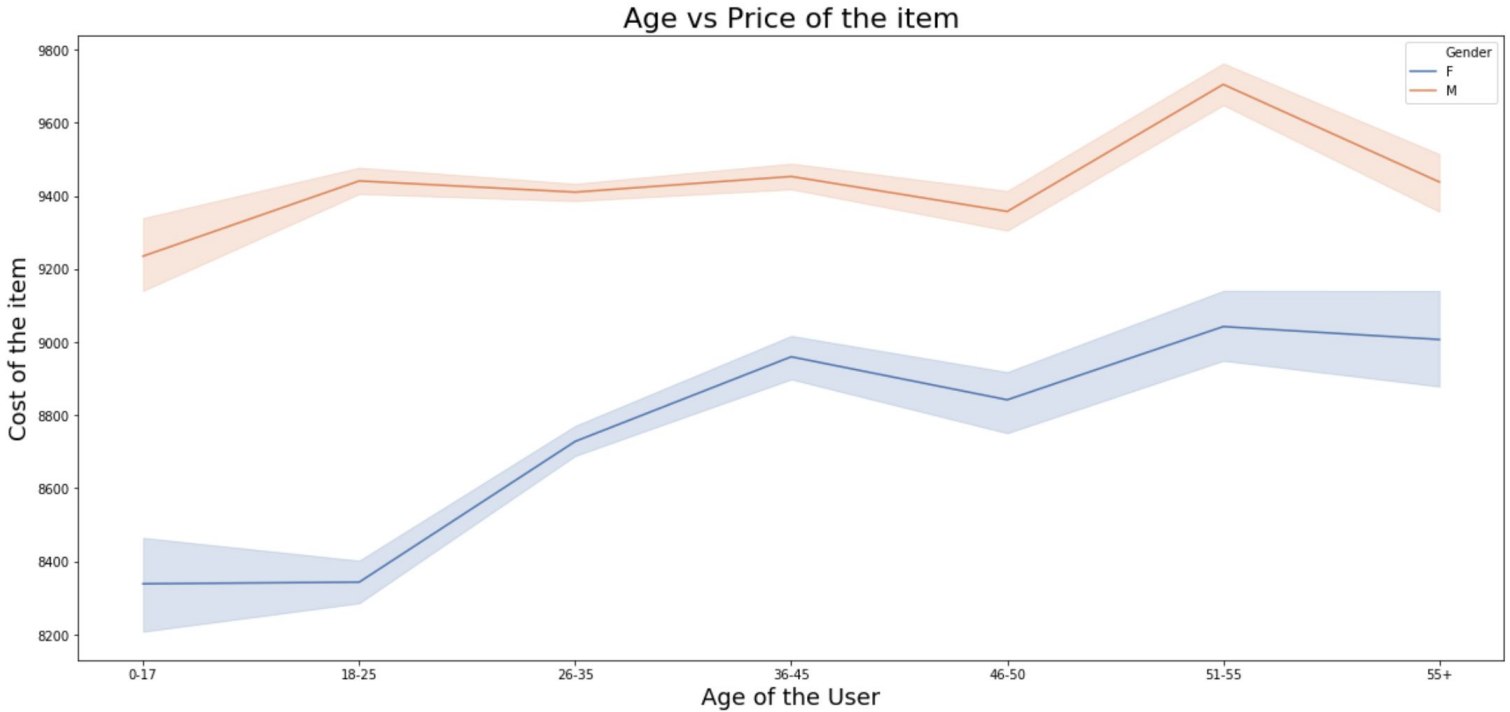




Exploratory Data Analysis

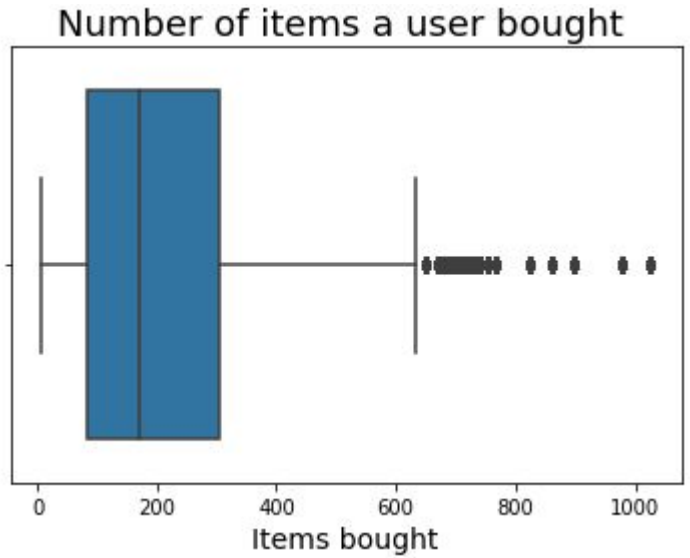


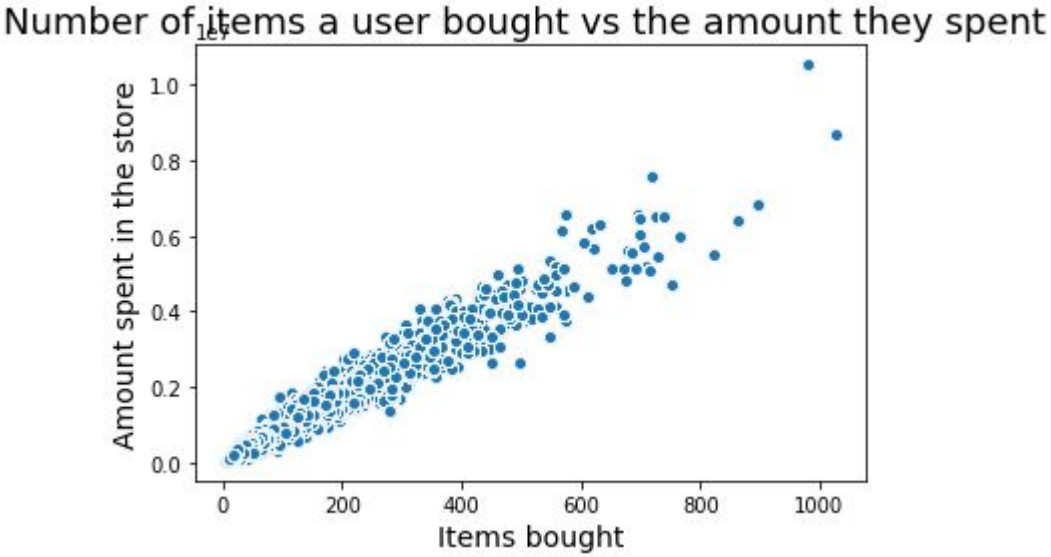
Exploratory Data Analysis



Feature Engineering

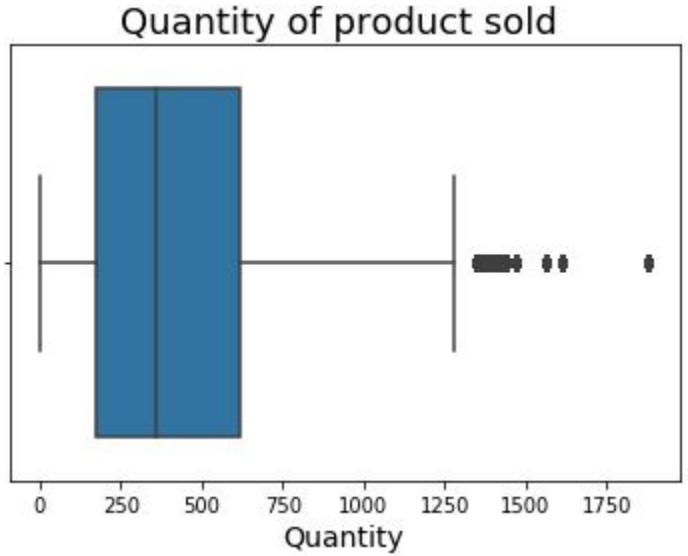
- Number of Products a Customer Buys
  - Count the number of products bought by customer ID





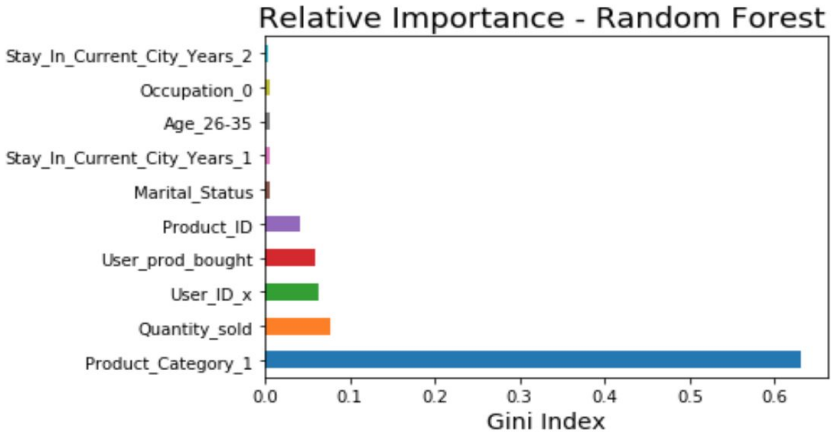
Feature Engineering

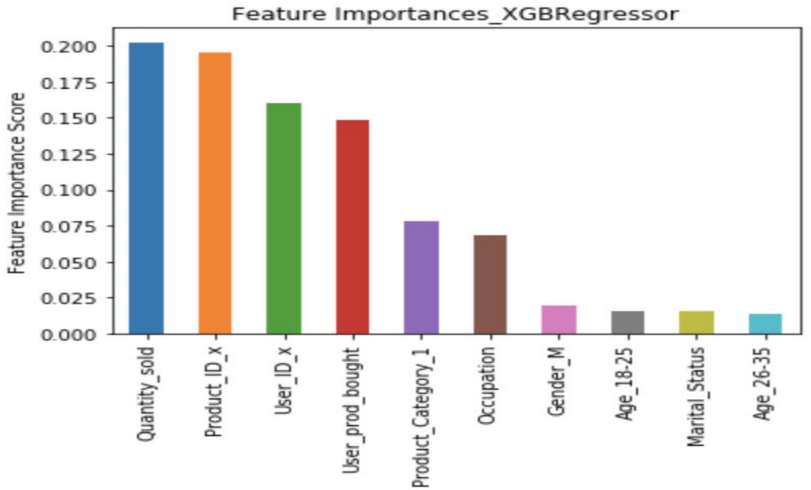
- Quantity of Item Sold
  - Count how much of each item is sold
  - We consider items that move more than 620 units (75th percentile) to be “High Sellers”



Modeling

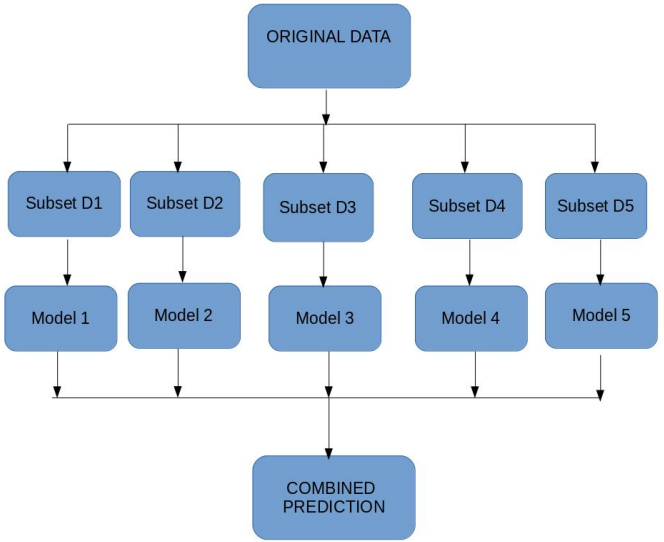
Model	RMSE_cv_ Before Feature Engineering	RMSE_cv_ After Feature Engineering	Change (%) between RMSE
Linear Regression	4382	4304	-1.78%
Random Forest	2865	2788	-2.67%
XGBRegressor	2659	2614	-1.69%





Future Work

- Original selling price to be the target value
- Request more features from retailer
- Ensemble techniques for better results



Recommendations

40% of the customers belong to age 26-35 and they are the high spenders and buy more number of items. Target this age group- advertising and marketing.

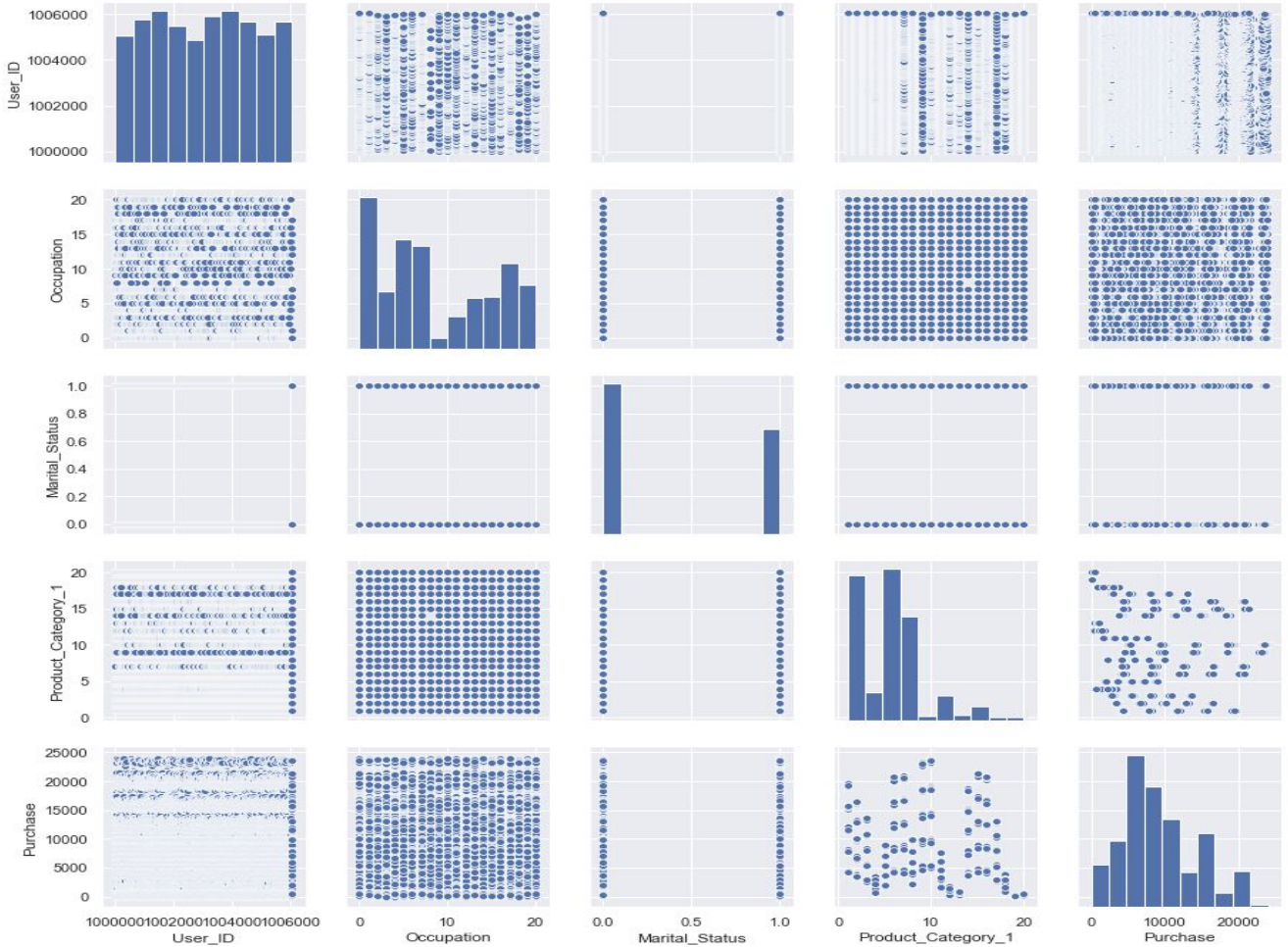
144 items were sold only in 1 quantity with no discount. We can increase the demand for this item by discounting the price or product placement in the store.

Since we already have a female customer base, we think of expanding the our store that sells female products as well.

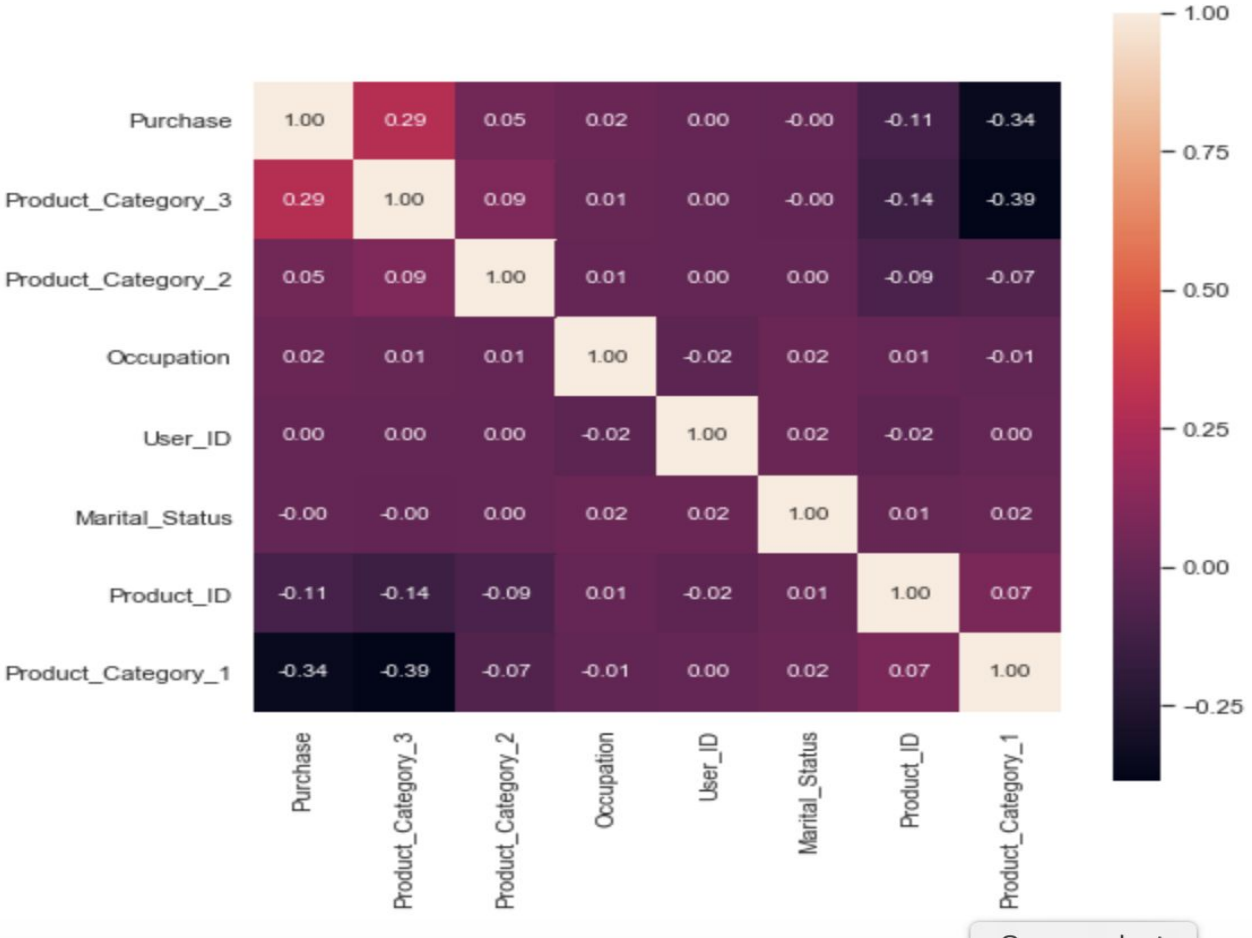
Thank You!

Extra

Pairplot


User_ID
Occupation
Marital_Status
Product_Category_1
Purchase
1006000
1004000
1002000
1000000
20
15
10
5
0
1.0
0.8
0.6
0.4
0.2
0.0
25000
20000
15000
10000
5000
0
0.5
10000
20000

Correlation Map



Null Value Pct

	Total	Percent
Product_Category_3	383247	0.696727
Product_Category_2	173638	0.315666
Purchase	0	0.000000
Product_Category_1	0	0.000000
Marital_Status	0	0.000000
Stay_In_Current_City_Years	0	0.000000
City_Category	0	0.000000
Occupation	0	0.000000
Age	0	0.000000
Gender	0	0.000000
Product_ID	0	0.000000
User_ID	0	0.000000