

BLACK FRIDAY PROJECT

Submitted by:

RAVI PRAKASH BAJPAI

**ACKNOWLEDGMENT**

Data for this project has been provided by flip robo.reference for making this project was taken from many websites like kaggle ,wikepedia,medium.com and https://www.geeksforgeeks.org.

**INTRODUCTION**

* Business Problem Framing

In real world many company are dealing with sales of product so according to there features provided data can be analysed which will help in growth in sale and improvement of business.

* Conceptual Background of the Domain Problem

Domain is basically understanding how to increase the sale of any business by analysing data

* Motivation for the Problem Undertaken

Through this project i may able to help any business person dealing with sale of product either it is on big scale or it may be on small scale

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

In this project basically label that is purchase is focused as main target for this project was how to increase the purchase.

* Data Sources and their formats

Data for this project was already provided to us by our team from flip robo. Features in the data are of both the type that is object and numeric.

* Data Preprocessing Done

It can be observed from juyper notebook that there was two features with missing data or having null value, so for product\_category\_2 mode method has been applied to fill the null values and product\_category\_3 has been droped because there were nearly 70% values which were not present.

After filling the null value get dummies has been applied to convert object data type into numeric.

* Data Inputs- Logic- Output Relationships

From heat map it is observed that features in the dataset are not closely related to each other nor they tightly corelated to label of the dataset.

* Hardware and Software Requirements and Tools Used

Harware requirement is only laptop with minimum i5 processor and for software juypter notebook should be installed /anaconda.

Libraries that are used are mainly

Matplotlib

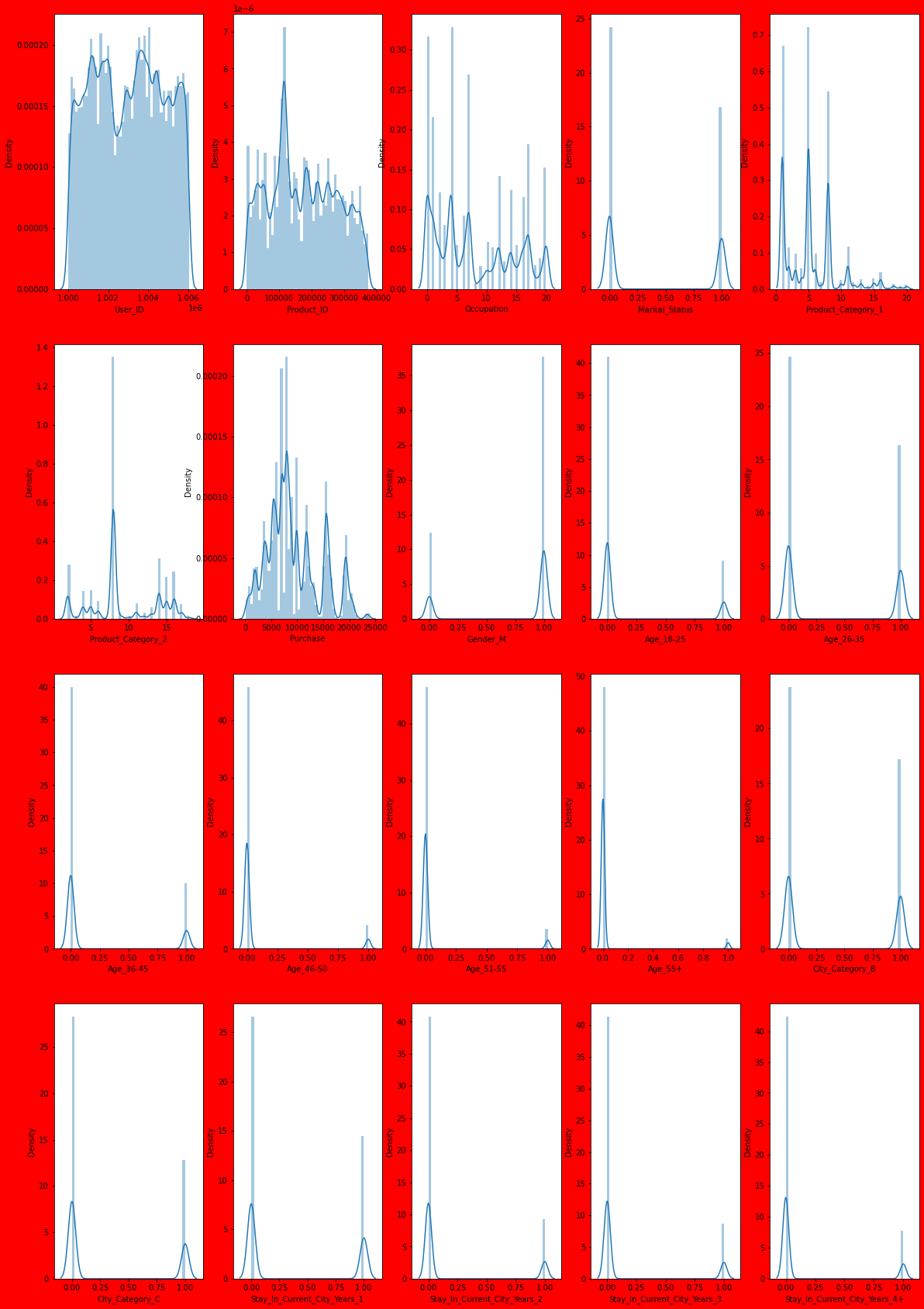
Pandas

Numpy

Seaborn

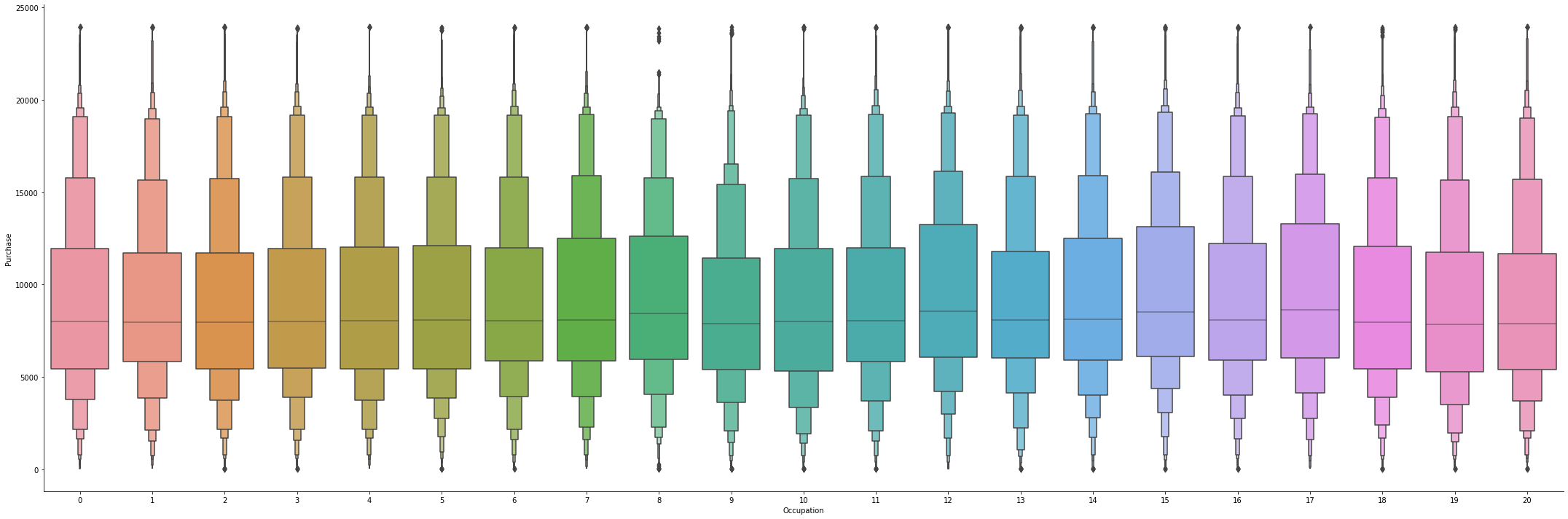
* Visualizations

Displot

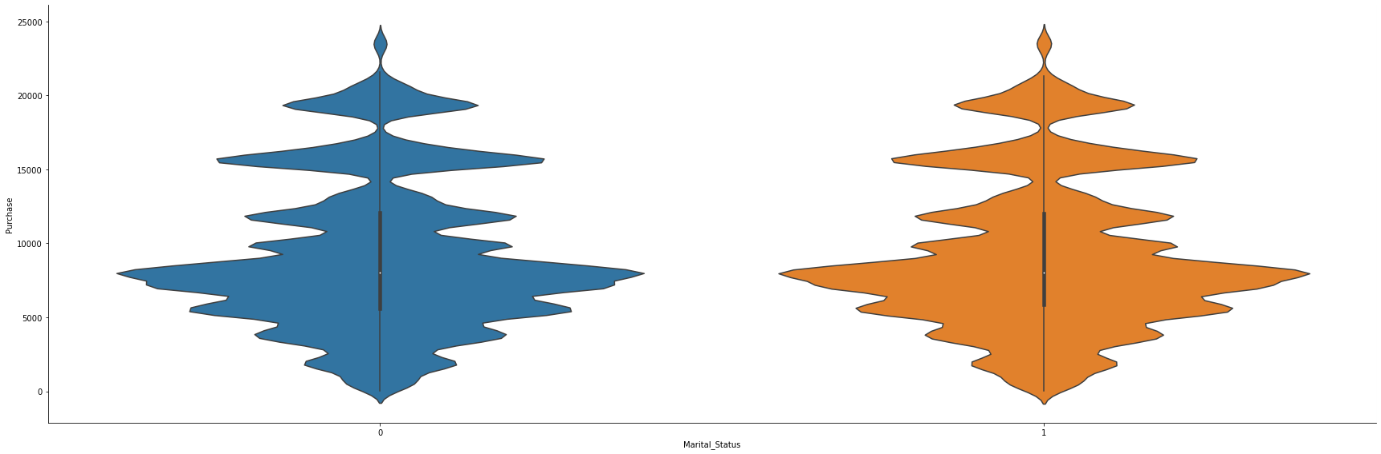


From distplot not much is observed only we can see that most of features are categorical.

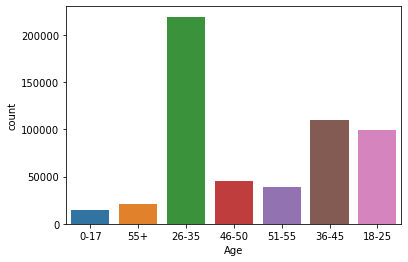
Catplot



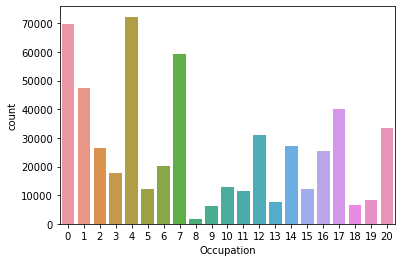
Outlier can be clearly observed from the above graph it is showing maximum in 8 category.



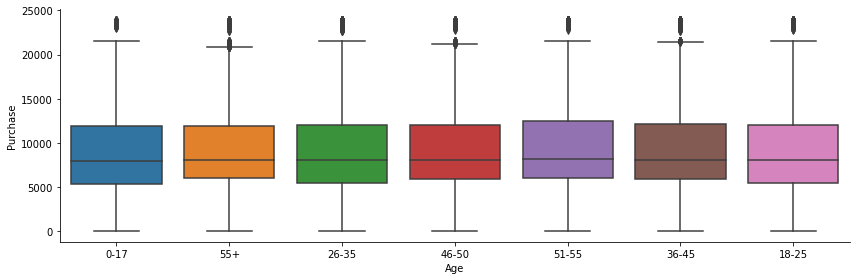
From martial\_status of any person effect on purchase do not vary that either 0 or 1 both are showing nearly same.



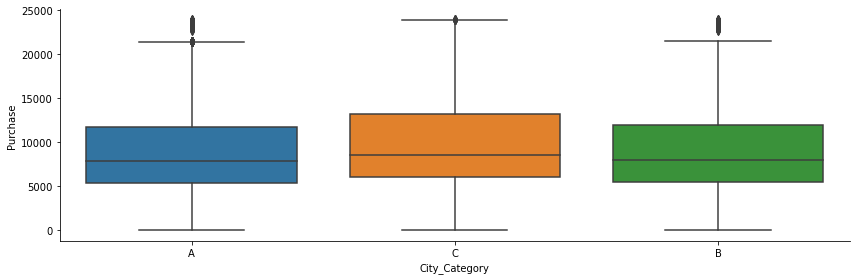
From count of age it can be infered that age group of 26-35 are more in count.



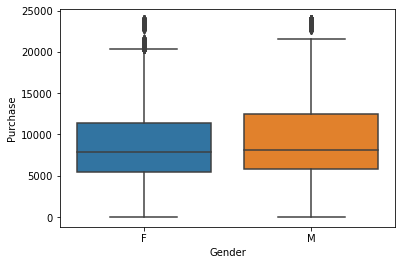
it can be observed that occupation type 4 has maximum number abd type 8 are least in numbers.



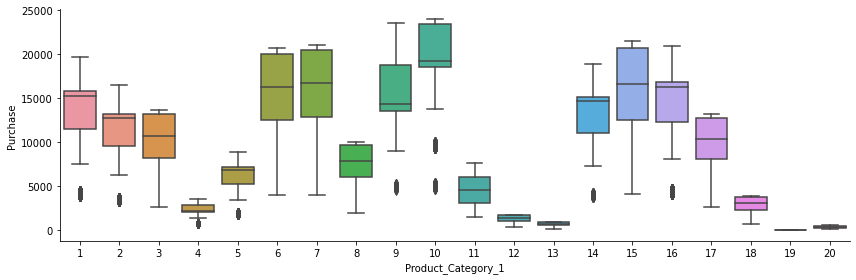
From above plot it can be analysed that there are many outliers present in age features.



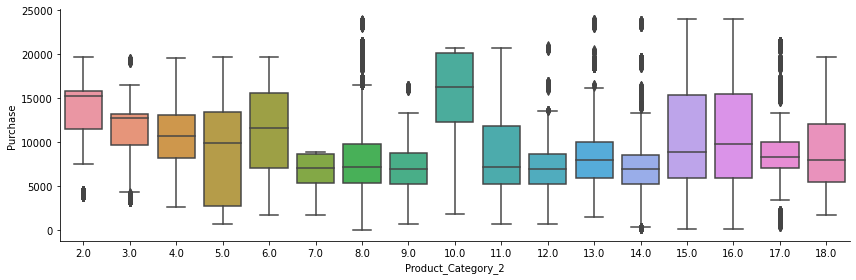
From above cat plot it can be seen that city category B purchase more than other two category and category A purchase is least of all three.



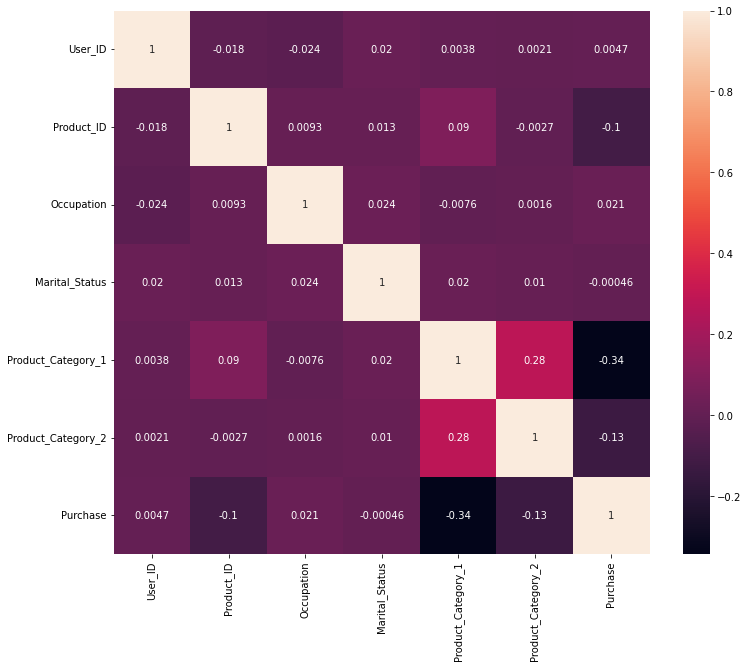
From above plot it can be seen that in female category of gender there are more outliers and purchase of male is more than female.



it can be observed from above graph that category 10 of product\_category\_1 is most purchased and 19 and 20 is least purchased.



it can be observed from above graph that 10,15 and 16 is purchase maximum, and maximum outliers are present in 17 and 8.



From heat map it can be observed that no two features are tightly corelated with each other nor they are corelated with label.

* Interpretation of the Results

From above visualization it can be interpreted that different features are influencing label differenly.it can also be seen that particalur categorical data is much effective than whole feature, for example in occupation count plot it can be oberved that 0,4,7,12,17,20 are covering most of the count so seller can focus on these these of people have above occupation type.

**CONCLUSION**

Key Findings and Conclusions of the Study

In this project it can be concluded that how particular area in ploted graph or we can say particular categorical data of features can be more effective than the whole features. This project gives idea that in which area of any given feature seller should concentrate more to increase the purchase and attract the purchaser by giving offer to such customer.