

**Project: BLACK FRIDAY EDA**

Submitted by:

Prashant Pathak (INT33)

Problam Statement:

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 summary of various customers for selected high-volume products from last month. The data set also contains customer demographics (age, gender, marital status, city\_type, stay\_in\_current\_city), product details (product\_id and product category) and Total purchas\_amount from last month.

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

• Data

• Variable Definition

• User\_ID: User ID

• Product\_ID: Product ID

• Gender: Sex of User

• Age: Age in bins

• Occupation: Occupation (Masked)

• City\_Category: Category of the City (A, B,C)

• Stay\_In\_Current\_City\_Years: Number of years stay in current city

• Marital\_Status: Marital Status

• Product\_Category\_1: Product Category (Masked)

• Product\_Category\_2: Product may belongs to other category also (Masked)

• Product\_Category\_3: Product may belongs to other category also (Masked)

• Purchase: Purchase Amount (Target Variable)

**Acknowledgement**

First, I would like to thank my mentor Mr. Shwetank Mishra for giving me the opportunity to work on this project.

Black Friday EDA is a very interesting project and I enjoy it.

It’s a project to find purchase behavior of consumers.

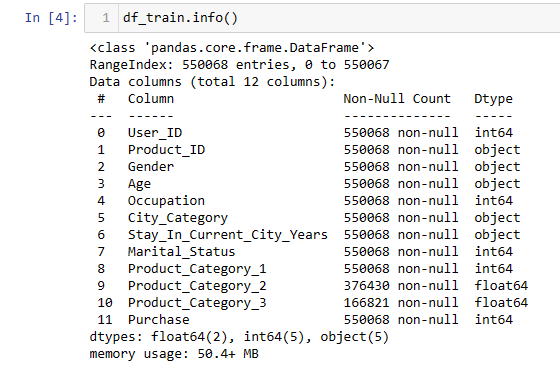
When I was working on this data, I found many interesting hidden patterns.

So now let's Go to the journey of Black Friday Sales EDA.

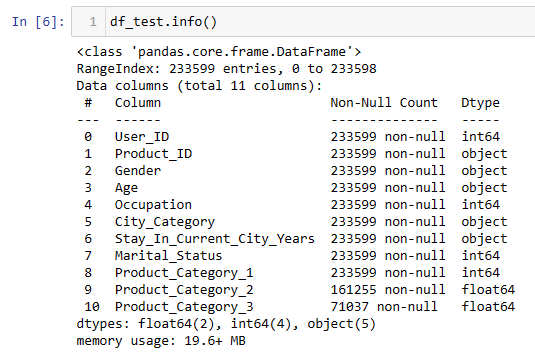
**About The Data**

We have 2 datasets one is for training and another one is for testing. Both data sets are in csv format.

Some information about training data:



Some information about testing data:



So, we have all the necessary information about both datasets.

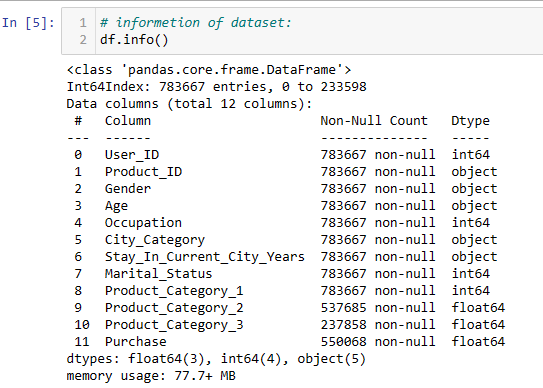
Train dataset has 550068 rows and 12 columns, and test dataset has 233599 rows and 11 columns.

Exploratory Data Analysis:

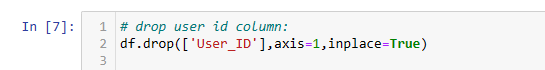
For Better EDA, we need to have a clear idea of our data.

So, to understand the data well, I took some steps like this: -

* Import all required modules and libraries.
* Lode the train dataset and test dataset.
* Marged both datasets using append method.
* Used df.info() to know about the dataset.

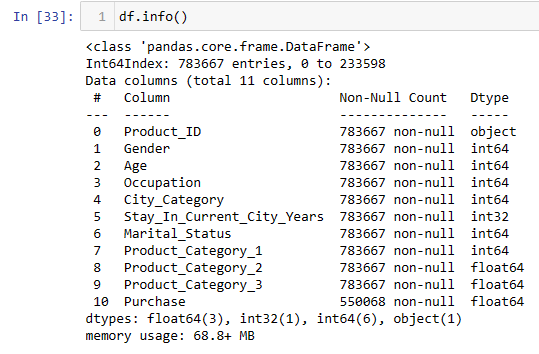


* Drop useless columns.



* Handle categorical features
* Handle missing values
* Handle datatypes of each column

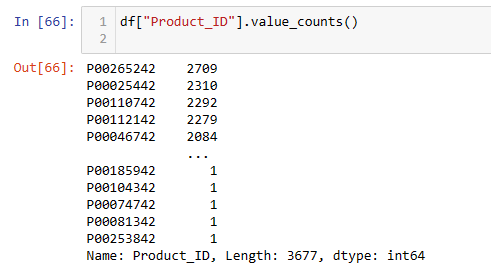
After the above steps, I have good data.



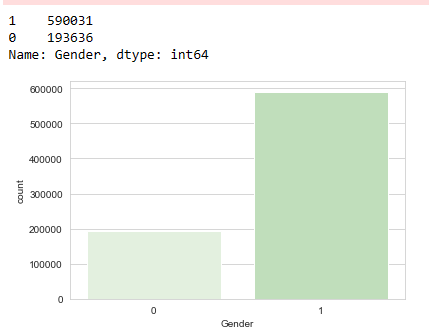
Visualizations

I basically use count plot and bar plot for visualization, and I found some interesting information about in data.

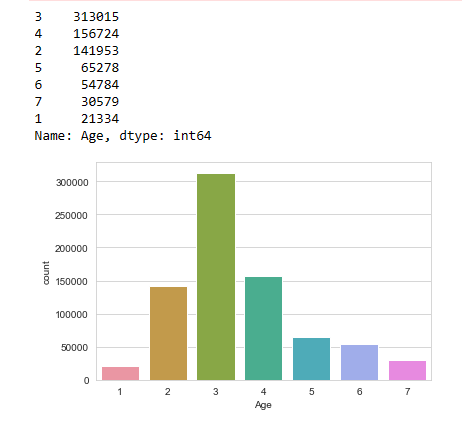
* Top 5 product\_id with counts. It means best-selling products.



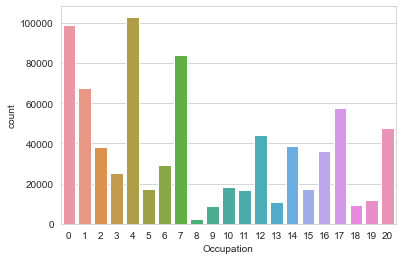
* Men purchase approx 3 times more than women. Is it possible? 😁



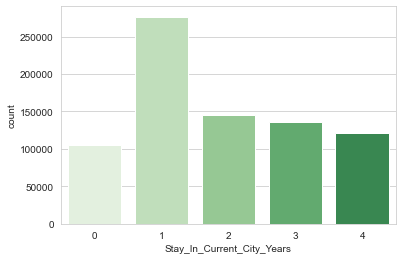
* Young people are the biggest purchaser which age group is 26y – 35y.



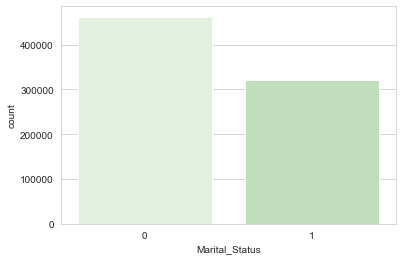
* Occupation 4 is the biggest purchaser in occupation categories.



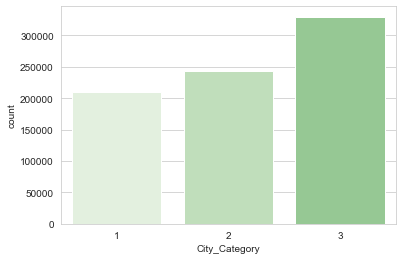
* People who have stayed in current city for 1 year buy more things in the Comparision of others.



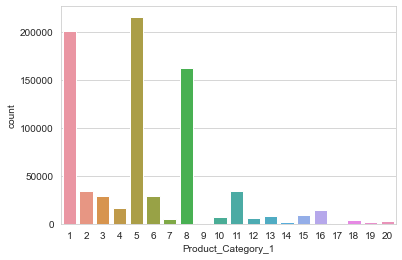
* Unmarried persons purchase more in Comparision of married persons. This fact is interesting. 😄



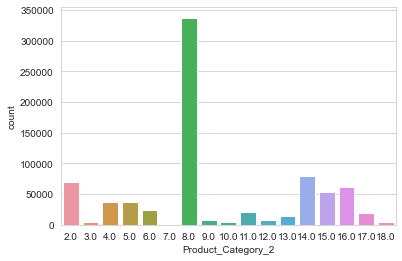
* City category “C” gets first rank of purchasing, and “B” gets second and “A” gets third. It means, citizens of city category “C” are good purchasers.



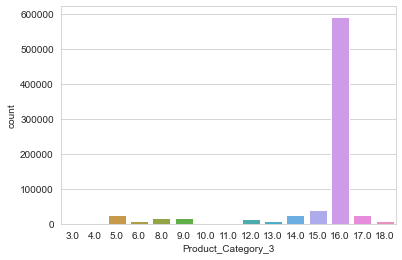
* Most purchased item in Product category\_1 is “16.0”.



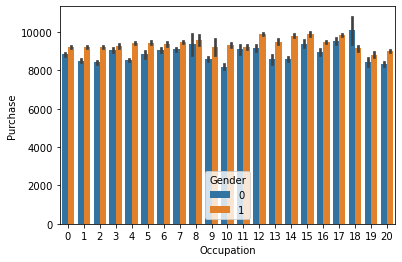
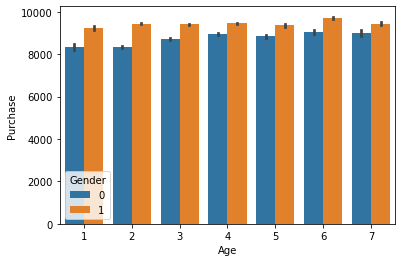
* Most purchased item in Product category\_2 is “8.0”.



* Most purchased item in Product category\_3 is “16.0”.

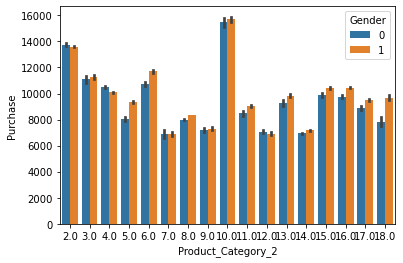
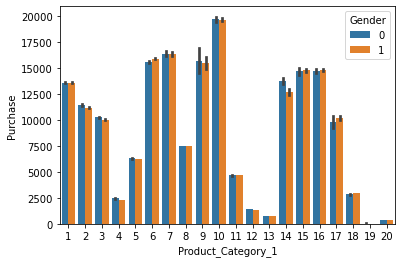


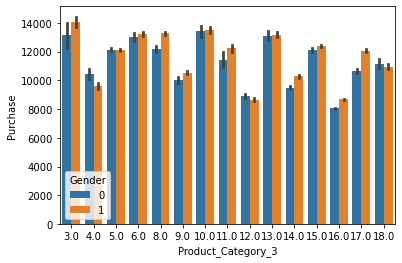
* Some more bar plots Showing relation with purchase.



* In above plot (Age vs Purchase) we see in all age category males are good purchaser in Comparision of females.
* In above plot (Occupation vs Purchase) we see in males are good purchaser in Comparision of females, except occupation “18”







* Through these graphs we can see the purchase price of each product of each product category.

**The end...**