FLIP ROBO

Black Friday Project

Submitted By :

Ruzaina Khan

ACKNOWLEDGMENT

I would like to thank Flip Robo Technologies for providing me with the opportunity to work on this project from which I have learned a lot. I am also grateful to Shwetank Mishra for his constant guidance and support.

INTRODUCTION :

A retail company “ABC Private Limited” wants to understand the customer purchase behaviour (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 purchase_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.

➢ Conceptual Background of the Domain Problem

So What is Black Friday . In in 1950 In Philadelphia (USA) there is football match Between American Army and American Navy . after one day of Thanks Giving So in that day there is a lots of crowd for watching the match . So for that reason for all the police is busy on that day to secure the city and in that day all police staff holiday is cancel .So for that reason we celebrate the day as called Black Friday.

So Basically Thanks giving is a festival in Eastern and Western Countries . and in every year it Celebrate in 4th Thrusday of Novemender.

So, In November 4th thrusday in 1950 every one celebrate Thanks Giving and after one day there is a football is help. And in that there is a lots of crowd . and all crowd are not wathing maximum crowd are shopping. And for that Black Day is the most shopping day as we can see in the EDA portion.

➢ Review of Literature

In Black Friday . Most of the peoples are shopping. Because in western countries Black Friday comes after one day of Thanks Giving. And for Black Day Peoples are waiting. Because Peoples get most of the offers for all the product. And no Matter what is his age . all age peoples are shoping in Black Friday.

And this festival Customers and Merchents both are get profit . Because in that day most of the population out for shopping in market and some are using online shopping.

➢ Motivation for the Problem Undertaken

So ,In my view. The main problem is to understand the Masked dataset. and in this dataset . we easily understand behaviour of the dataset..

But there is a one more big problem in set we have Nan values . In Product_category_2 and Product_category_3.

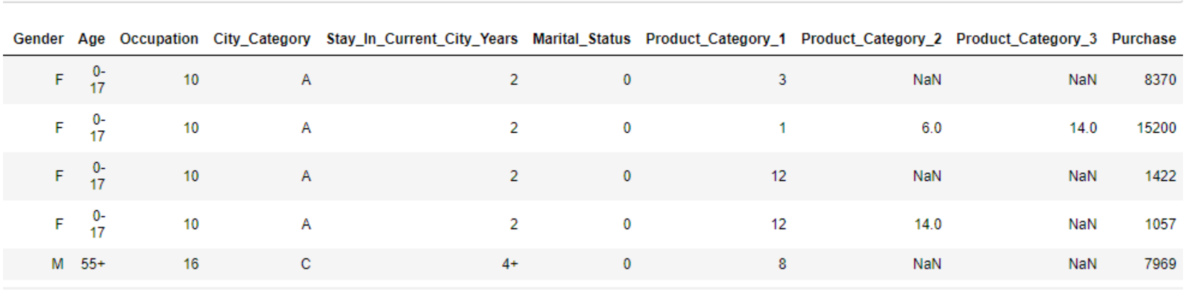
So, we keep or Remove this Nan values. so we decided in preprocessing step. And this dataset is very clear. SHOPPING SHOPPING SHOPPING in Black Friday.

Analytical Problem Framing

- Mathematical/ Analytical Modeling of the Problem

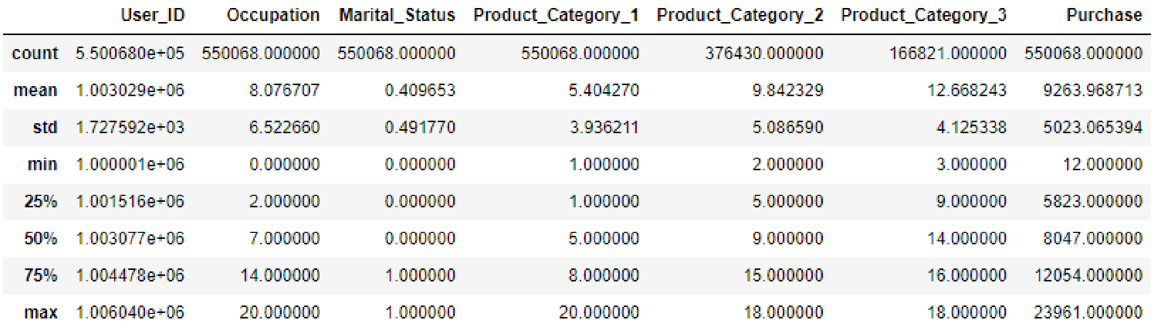
  This Project is a Regression Problem. And this dataset we have to predict Purchase amout of the product .

  Lets see the Dataset How its Look:



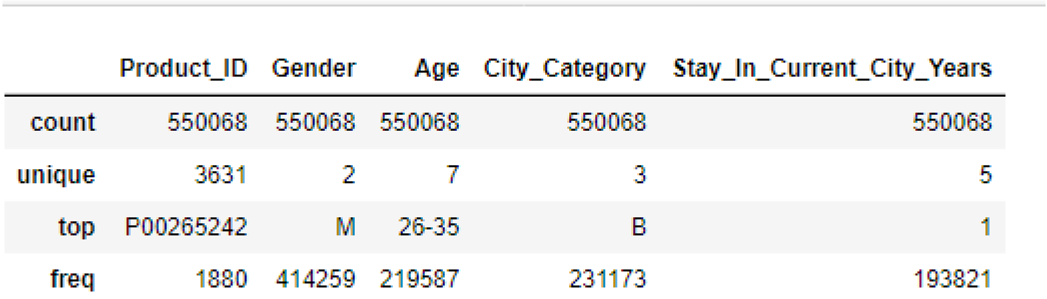
Gender	Age	Occupation	City_Category	Stay_In_Current_City_Years	Marital_Status	Product_Category_1	Product_Category_2	Product_Category_3	Purchase
F	0-17	10	A	2	0	3	NaN	NaN	8370
F	0-17	10	A	2	0	1	6.0	14.0	15200
F	0-17	10	A	2	0	12	NaN	NaN	1422
F	0-17	10	A	2	0	12	14.0	NaN	1057
M	55+	16	C	4+	0	8	NaN	NaN	7969

Lets Statistical Summary of Int and Float Type Data:



	User_ID	Occupation	Marital_Status	Product_Category_1	Product_Category_2	Product_Category_3	Purchase
count	5.500680e+05	550068.000000	550068.000000	550068.000000	376430.000000	166821.000000	550068.000000
mean	1.003029e+06	8.076707	0.409653	5.404270	9.842329	12.668243	9263.968713
std	1.727592e+03	6.522660	0.491770	3.936211	5.086590	4.125338	5023.065394
min	1.000001e+06	0.000000	0.000000	1.000000	2.000000	3.000000	12.000000
25%	1.001516e+06	2.000000	0.000000	1.000000	5.000000	9.000000	5823.000000
50%	1.003077e+06	7.000000	0.000000	5.000000	9.000000	14.000000	8047.000000
75%	1.004478e+06	14.000000	1.000000	8.000000	15.000000	16.000000	12054.000000
max	1.006040e+06	20.000000	1.000000	20.000000	18.000000	18.000000	23961.000000

Lets See Statistical Summary of Object Data Type:



	Product_ID	Gender	Age	City_Category	Stay_In_Current_City_Years
count	550068	550068	550068	550068	550068
unique	3631	2	7	3	5
top	P00265242	M	26-35	B	1
freq	1880	414259	219587	231173	193821

- About dataset

- Here , We have 550068 Rows and 121Columns
- We have null values in 2 Columns
- We have Duplicated Value and we remove that.
- We have 5 Object Type data type and 5 integer type data type and 2 float type (float type because it contains null values)
- This Data Usage 54.6+ MB storage .

- **EDA- Exploaratory Data Analysis**

- As we know we have 2 Columns with Nan Values . So we don’t think to delete this right . First I do EDA Then I will take my action against Nan values. Lets See Column Who have Nan Values.
- Total Rows -> 550068
- 1 -> Product_Category_2   -> 173638 Null Values

- 2-> Product_Category_3 -> 383247 Null Values
- So, I decided to not delete Nan values and move foreword to EDA . To se exact behaviour of customers.

- Hardware and Software Requirements and Tools Used

- Anaconda Navigator -> Jupyter Notebook
- Hardware -> AMD Ryzen 3 Processor with Vega Graphics 2200U.
- RAM -> 8GB
- SSB -> 120 GB

- **Visualization Done in PPT.(Please Follow PPT)**

❖ Conclusion :

Please Follow PPT :

- If you see in Slide -10 then you can see the counts of Male – 414259 and Female-135809 . And after that you can see in Slide 30, 31 and 32 . that the graph of all Products with his Price and hue by Gender. And you can find there . For all product female and male both are equal in counts. So its means is all the peoples are shopping in Black Friday..
- If we include the City also. the count of A, B and C city. So that is not properly balanced but if see the hue with Product and Price Then we can find all the cities order same counts for all the products.
- Black Day is a Shopping Day for Western and Eastern Countries. And Now we can see which which product needed more in this Black Day.
- So , Most of People Come From Occupation 4 then 0 then 7 then 1 then 17 then 20 then 20 and 12 .(Slide 10)
- Maximum Ordered Product for Product_Category_1 is 1 , 5 , 8 ,11 , 2,6,3 and 4
  Least Order Product -> 9,17, 14 , 19 ,20, 18, 7, ,12 ,10, 13, 15 and 16 .(Slide ->16)
- Maximum Ordered Product for Product_Category_2 is 8.0,14.0,2.0,16.0
  Least Order Product -> is 7.0 and 18,3.0,10.0,12.0 and 9.0. (Slide->17)
  Product_Category_2 Contain Null Values of Counts -> 173638
- Maximum Ordered Product for Product_Category_3 is 16.0,15.0,14.0,17.0,5.0,8.0, and 9.0

Least is 3.0,10.0,11.0,4.0,18.0 and 6.0 (Slide->18)
Product_category_3 Cotain Null values of counts 383247 (Big Number) Safe to delete this column

- In Slide 27 You can see that Maximum money spend from 51-55 age group then 55+ then 36-45 and 18-25 are same then 0 -17. If you see in the Graph then you can see that all age group Shopping in equal in counts. So you have to focus for age category products.
- We don’t check for outliers and Distribution plot for all columns because we don’t need to check because column are Float and Int Type but it’s a Categorical data .
- And I decide to fill null or remove null after model building in both types. Without null or fill null. So then I decide null is important or not.

- **Happy Black Friday to You and Your Customer.**

Thankyou 😉 😉