**Customer Behavior Analysis Black**

**Friday**

**A PROJECT REPORT**

*Submitted by*

**PONNURI ANIRUDDHA [Reg No: 2112704010015]**

*Under the Guidance of*

**Dr. Kalpana A V**

(Assistant Professor, Department of Data Science and Business Systems)

*In partial fulfillment of the Requirements for the Degree of*

**BACHELOR OF TECHNOLOGY COMPUTER SCIENCE AND BUSINESS SYSTEMS**



**DEPARTMENT OF DATA SCIENCE AND BUSINESS SYSTEMS**

**FACULTY OF ENGINEERING AND TECHNOLOGY**

**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY**

**NOVEMBER 2022**

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY KATTANKULATHUR-603203

BONAFIDE CERTIFICATE

Certified that this project report titled “**CUSTOMER BEHAVIOR ANALYSIS ON BLACK FRIDAY”** is the bonafide work of **“PONNURI ANIRUDDHA [Reg No:**

**2112704010015]”** who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion for this or any other candidate.

|  |  |
| --- | --- |
| Dr. Kalpana A V Dr.G. Vadivu | Dr. M.Lakshmi |
| **Guide Ass.Professor** | **HEAD OF THE DEPARTMENT** |
| Assistant Professor Dept. of DSBS  Dept. of DSBS | Dept. of DSBS |
| Signature of Internal Examiner | Signature of External Examiner |

## ABSTRACT

On the day of the Black Friday sale, customers can be found crowding the aisles of every retail store. Customers rush in to acquire the products as most of them have been reduced in price and discounts have been applied. Even with a well thought out strategy, it is difficult for buyers to purchase the products. However, the owners of the store have an even harder time controlling the crowd with the little number of employees that they have and focusing on potential consumers.

The resolution of this issue has been approached from a few different angles, but none of them have proven to be particularly effective. A method that has shown some promise in terms of finding a solution to the issue is called a prediction model.

This study focuses on the topic of prediction models with the goal of developing an accurate and efficient algorithm to evaluate the spending patterns of customers in the past and output the spending patterns that same customers will have in the future using the same characteristics. For the purpose of developing a prediction model, this research makes use of a variety of machine learning approaches, including regression and neural networks, and then does a comparison between these techniques based on their overall performance and accuracy of prediction.

In order to make the most accurate prediction possible, these strategies are put into practise by employing a variety of algorithms and on a variety of platforms. Seven distinct machine learning algorithms were incorporated into our system.

In addition to that, this research delves into the methods of data pre-processing and visualisation that were utilised to achieve the best possible outcomes.

### ACKNOWLEDGEMENTS

We express our humble gratitude to **Dr C. Muthamizhchelvan**, Vice-Chancellor, SRM Institute of Science and Technology, for the facilities extended for the project work and his continued support.

We extend our sincere thanks to Dean-CET, SRM Institute of Science and Technology, **Dr T.V.Gopal**, for his invaluable support.

We wish to thank **Dr Revathi Venkataraman**, Professor & Chairperson, School of Computing, SRM Institute of Science and Technology, for her support throughout the project work.

We are incredibly grateful to our Head of the Department**, Dr M. Lakshmi** Professor, Department of Data Science and Business Systems, SRM Institute of Science and Technology, for her suggestions and encouragement at all the stages of the project work.

We want to convey our thanks to our program coordinators **Dr.E.Sasikala,** Professor, Department of Data Science and Business Systems, SRM Institute of Science and Technology, for her inputs during the project reviews and support.

We register our immeasurable thanks to our Faculty Advisor**, Dr K Shantha Kumari, Ph.D.,** Assistant Professor, Department of Data Science and Business Systems,

SRM Institute of Science and Technology, for leading and helping us to complete our course.

Our inexpressible respect and thanks to my guide, **Dr. Kalpana A V,** Assistant Professor, Department of Data Science and Business Systems, for providing me with an opportunity to pursue my project under his mentorship. He provided us with the freedom and support to explore the research topics of our interest. His passion for solving problems and making a difference in the world has always been inspiring.

We sincerely thank the Data Science and Business Systems staff and students, SRM Institute of Science and Technology, for their help during our project. Finally, we would like to thank parents, family members, and friends for their unconditional love, constant support, and encouragement.

Ponnuri Aniruddha

### TABLE OF CONTENTS

CHAPTER NO. TITLE PAGE

NO.

ABSTRACT 3

ABBREVIATIONS 6

1. INTRODUCTION 7

1. LITERATURE REVIEW 9
2. OBJECTIVES 11
3. BLOCK DIAGRAM 12
4. WORKING MODEL 13

#### 6 PROJECT CODE 14

6.1 PROJECT CODE

#### 7 REQUIREMENTS 25

7.1 HARDWARE REQUIREMENTS 7.2 SOFTWARE REQUIREMENTS

1. PROJECT FINDINGS 24
2. CONCLUSION 33
3. Future Enhancements 34

#### 11 References

## ABBREVIATIONS

**AI** Artificial Intelligence

**IOT** Internet Of Things

**GUI**  Graphical User Interface

**CHAPTER 1**

# INTRODUCTION

#### **1.1 General**

Small businesses were the only retail options available in the past; there were no supermarkets or department stores. The owners of the store were familiar with their clientele and the spending habits, preferences, and preferences of their clients. However, as the little firm expanded into enormous franchises with hundreds of locations around the country, it became nearly difficult to know the clients and their individual tastes. This is because the stores were spread out across the country. Costco, Walmart, and Whole Foods are a few examples of franchises that fall into this category. These stores, who lack an adequate knowledge of their customer base, are having difficulty satisfying the requirements of their customers. Therefore, prediction models are necessary in order to have a deeper understanding of the preferences of customers.

The construction of a prediction model is dependent on several different aspects, including the place and the time. In the United States of America, the day after Thanksgiving known as "Black Friday" is the busiest shopping day of the year.

The day following Thanksgiving, sometimes known as "Black Friday," is traditionally considered to be the beginning of the Christmas shopping season. Because of the dramatic difference in customer spending between a typical day and Black Friday—due to the fact that discounts and price reductions attract a significantly larger number of customers—a prediction model that was designed specifically for Black Friday is only applicable on that day. According to research conducted by the National Retail Federation [1,] over the course of the Black Friday weekend in 2010, 212 million consumers made purchases in-store and online. The fact that the data that was utilized to construct the current prediction model has a number of abnormalities, such as missing values or incorrect information, is the primary issue with the model. In addition to this, choosing the appropriate algorithm is an essential component in the process of generating an accurate model. Last but not least, improved visualization methods are essential for portraying the findings and assisting store owners in developing a deeper understanding of their clientele. In this area, some of the problems that need to be answered include: Can an accurate prediction model be created? Which algorithm should be used because it will be more effective for such a model? Will the accuracy be improved by performing data pre-processing and using a visualization technique?

#### **1.2 MOTIVATION**

As merchants are concerned with how consumers purchase across several channels, an examination of shopping habits during the holiday shopping season may shed light on multichannel shopping behaviour. This study's objectives are to comprehend variances in Black Friday buyers' attitudes and intentions and to propose marketing strategies that capitalise on these disparities. Retailers face a hurdle with multichannel consumers because it is unclear why consumers utilise particular channels. We contribute to the literature by focusing on the most specific purchasing activity throughout the year in the expectation that using this particular microcosm of increased buying activity would aid in comprehending multichannel shopping activity throughout the remainder of the year.

25–40% of yearly U.S. retail sales are comprised of winter holiday buying (NRF, 2011). In order to achieve excellent financial results for the year, shops eagerly anticipate the holiday shopping season, during which consumers spend substantially. During the holiday season, merchants rely heavily on advertising and consumer word-of-mouth to draw people into shops and online. To maximise their efforts, businesses are eager to comprehend consumer purchasing decisions that will help them achieve earnings during this season. The goal of this study is to gain insight into consumers' intentions to shop on these two sale days so that retailers can make more informed decisions regarding Christmas marketing. Understanding the shopping behaviours of holiday shoppers will enable retailers to comprehend why some consumers prefer mall shopping and others prefer internet shopping. The value of this study is an understanding of consumer motivations during the two most crucial shopping days throughout the holiday season, as well as the reasons underlying these motivations. The purpose is to comprehend these reasons on a theoretical level in order to offer retailers with the answers they require about holiday marketing spending. With this information, merchants can create campaigns that target their audiences depending on the holiday shopping channel preferences of consumers.

**CHAPTER 2**

# LITERATURE REVIEW

* **Vishal Shrivastava**, the author proposed, “A study of various clustering algorithms on retail sales data”, this paper discusses the four major clustering algorithms kmeans, density based, filtered, farthest first clustering algorithm and comparing the performances of these principal clustering algorithm on the aspect of correctly class wise cluster building ability of algorithm. The results are listed on datasets of retail sales using weak interface and compute the correctly cluster building instances in proportion with incorrectly formed cluster. A comparison of these four algorithms is given on the basis percentage of incorrectly classified instances.
* **Akshay Krishna** and **Akhilesh V**, the authors proposed, “Sales – forecasting for retail stores using machine learning techniques”, this paper tries to predict the sales of a retail store using different machine learning techniques and tries to determine the best algorithm suited to a problem statement. They implemented normal regression techniques and as well as boosting techniques. Finally, they conclude that boosting algorithm have better results than the regular regression algorithms.

#### • **Gopalakrishnan T**, **Ritesh Choudhary** and **Sarada Prasad**, the authors proposed,

“Prediction of Sales Value in Online shopping using Linear Regression”, [8] the aim of this paper is to analyze the sales of a big superstore, and predict future sales for helping them to increase their profits and make their brand even better and competitive as per the market trends by generating customer satisfaction as well. The technique used for prediction of sales is the Linear Regression Algorithm, which is a famous algorithm in the field of Machine Learning.

* **Velmurugan Ramasamy** The author proposed, “CONSUMERS PREFERENCE TOWARDS ORGANIC FOOD PRODUCTS.” The market of organic products is growing as the number of people willing to consume organic food and consumer preference towards organic food products is ever increasing. The promotion of organic food products constitutes an important option not only for producers, government and consumers but also to respond to societies ‘desire for higher food quality and food production that is less damaging to environment systems and improve the quality of life; this makes the study of consumer preferences highly important. The main purpose of this study is to identify the factors influencing consumer preferences toward organic products. The target population in this research includes consumers of Kozhikode district of Kerala state of India
* In 2015, **Elvander,** a master’s student from Uppsala University created a recommender system for Plick [4] using user-based collaborative filtering. This method is largely used in the recommender system field used in renowned systems such as Youtube and Amazon. Elvander states that due to the unstable nature of the item data in the system, it is more beneficial to use users to create recommendations rather than using items.

**CHAPTER 3**

# OBJECTIVES

For the purpose of this study, the dataset consisted of information regarding various sales transactions. Anyone is able to access it by going to the following URL in their web browser.

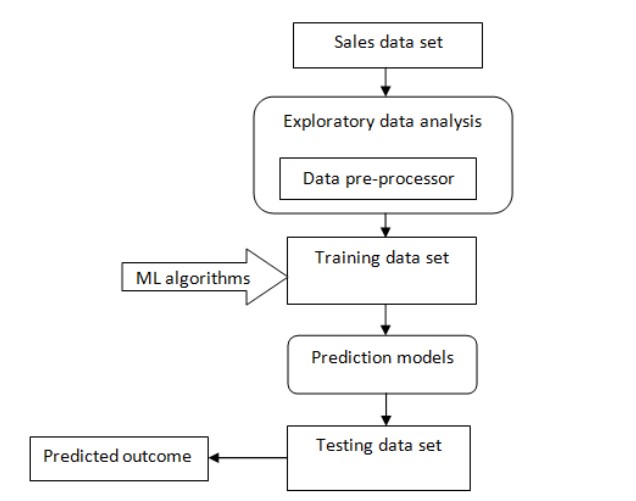
The following URL provides access to Black Friday sales: https://datahackanalyticsvidhya.com/contest/black-friday/#data-dictionary.

It has around 550 thousand records of transactions related to sales. The twelve distinguishing features of each record are outlined in Figure 1, which may be found here. An organisation that deals in retail goods is interested in gaining an understanding of the purchasing patterns of customers in relation to a wide range of products that can be categorised in a variety of different ways. They have supplied a rundown of the high-volume product purchases that were made by a variety of customers over the course of the preceding month for a variety of different products. Additionally included in this data set are client demographics such as age, gender, marital status, city type, and stayjn current city, in addition to product characteristics such as product id and product category, as well as the total amount of money spent during the month prior to the current one.

Now that we have this dataset, we can use it to train an algorithm for supervised machine learning in order to forecast the amount of money consumers will spend on a variety of products. This would make it easier for the company to provide a personalised offer for customers in relation to the different products.

**CHAPTER 4**

# BLOCK DIAGRAM



**CHAPTER 5**

# Working Model

1.XGBoost

The XGBoost model internally implements the stepwise, ridge regression which dynamically selects the features and removes the multi-collinearity with the features. This implementation gave the bet results of this dataset. It uses ensemble model to learn from the weak predictors and eliminate the less important features to develop a strong model. We got a RMSE of 2729.

2.Linear Regression

Linear Regression establishes a relationship between dependent variable (Y) and one or more independent variables (X) using a best fit straight line (also known as regression line). It is represented by an equation Y=a+b\*X + e, where a is intercept, b is slope of the line and e is error term. This equation can be used to predict the value of target variable based on given predictor variable(s).

3.Random Forest

The Random Forest algorithm is a type of Supervised Machine Learning that has found widespread application in the areas of both classification and regression. It constructs decision trees using a variety of samples, then uses the majority vote of those samples to determine categorization and uses an average to determine regression. It is one of the most important features of the Random Forest Algorithm that it can handle the data set containing continuous variables as in the case of regression and categorical variables as in the case of classification. This capability is one of the most important features of the Random Forest Algorithm. It produces superior outcomes when applied to classification tasks.

**CHAPTER 6**

# PROJECT CODE

*# Black Friday Sales Prediction ## Import Statements import numpy as np import pandas as pd import seaborn as sns import matplotlib.pyplot as plt*

*%matplotlib inline*

*from google.colab import drive drive.mount('/content/drive') p1 = '/content/drive/MyDrive/Colab Notebooks/FDS/train.csv'*

*## Reading Data From Training Set data = pd.read\_csv(p1)*

*dataset = data[['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']] dataset.head()*

*## Data Analysis & Visualization*

*### Purchase sns.distplot(dataset['Purchase'],color='maroon',bins=25) plt.ylabel('No. Of Customers') plt.xlabel("Amount spent in purchase") print('Skewness of data is :',dataset['Purchase'].skew()) print('Kurtosis of data is :',dataset['Purchase'].kurtosis())*

*#### Univariate analysis of target variarble ( Purchase ) shows Gaussian distribution with skewness of 0.56 and kurtosis of -0.42 which is pretty good.*

*sns.boxplot(dataset['Purchase'],color='green')*

*#### Box Plot of target variable shows presence of outliers which need to be removed from the data. ### Gender sns.countplot(dataset['Gender'])*

*#### It can be seen from data that male buyers are relatively more.*

*### Age sns.countplot(dataset['Age'])*

*#### As expected, most purchases are made by people between 18 to 45 years old. ### City Category sns.countplot(dataset['City\_Category'])*

*#### City B has relatively more buyers than A & C*

*### Marital Status sns.countplot(dataset['Marital\_Status'])*

*#### Unmarried buyers are relatively greater.*

*### Stay In Current City Years sns.countplot(dataset['Stay\_In\_Current\_City\_Years'])*

*#### People staying in city for year are more keen to buy the product.*

*### Occupation plt.figure(figsize=(14,5)) sns.countplot(dataset['Occupation'])*

*## Handling Categorical Values ### OneHotEncoding gen\_onehot\_features = pd.get\_dummies(dataset['Gender'])*

*dataset = pd.concat([dataset[['User\_ID','Product\_ID','Age', 'Occupation', 'City\_Category',*

*'Stay\_In\_Current\_City\_Years', 'Marital\_Status', 'Product\_Category\_1', 'Product\_Category\_2', 'Product\_Category\_3',*

*'Purchase']],pd.DataFrame(gen\_onehot\_features)],axis=1) gen\_onehot\_features.head() gen\_onehot\_features\_city = pd.get\_dummies(dataset['City\_Category']) dataset = pd.concat([dataset[['User\_ID','Product\_ID','Age', 'Occupation',*

*'Stay\_In\_Current\_City\_Years', 'Marital\_Status', 'Product\_Category\_1',*

*'Product\_Category\_2', 'Product\_Category\_3',*

*'Purchase','M','F']],pd.DataFrame(gen\_onehot\_features\_city)],axis=1) gen\_onehot\_features\_city.head() dataset.head() ### Missing Values dataset.isnull() dataset['Product\_Category\_2'] = dataset['Product\_Category\_2'].fillna(999) dataset['Product\_Category\_3'] = dataset['Product\_Category\_3'].fillna(999) dataset['Product\_Category\_2'] = dataset['Product\_Category\_2'].astype(int) dataset['Product\_Category\_3'] = dataset['Product\_Category\_3'].astype(int) dataset.head()*

*### Mapping Ordered Data gen\_ord\_map = {'0-17': 0, '18-25': 1, '26-35': 2,*

*'36-45': 3, '46-50': 4, '51-55': 5,'55+':6}*

*dataset['Age'] = dataset['Age'].map(gen\_ord\_map) dataset.head() ### LabelEncoding from sklearn.preprocessing import LabelEncoder*

*gle = LabelEncoder() genre\_labels = gle.fit\_transform(dataset['Stay\_In\_Current\_City\_Years']) genre\_mappings = {index: label for index, label in enumerate(gle.classes\_)} dataset['Stay\_In\_Current\_City\_Years'] = genre\_labels gle = LabelEncoder() genre\_labels = gle.fit\_transform(dataset['User\_ID']) genre\_mappings = {index: label for index, label in enumerate(gle.classes\_)} dataset['User\_ID'] = genre\_labels dataset.head() ## Removing Outliers from scipy import stats z = np.abs(stats.zscore(dataset['Purchase']))*

*threshold = 2.33*

*np.where(z > 2.33)*

*dataset = dataset[(z<2.33)] sns.boxplot(dataset['Purchase'])*

*## Splitting Data*

*X = dataset[['User\_ID','Age', 'Occupation', 'Stay\_In\_Current\_City\_Years', 'Marital\_Status',*

*'Product\_Category\_1', 'Product\_Category\_2', 'Product\_Category\_3', 'M', 'A', 'B']]*

*y = dataset['Purchase']*

*from sklearn.model\_selection import train\_test\_split*

*X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=0) X\_train*

*## Training Model ### Linear Regression from sklearn.linear\_model import LinearRegression*

*regressor = LinearRegression()*

*regressor.fit(X\_train,y\_train)*

*print("Intercept:",regressor.intercept\_) print("\nSlope:",regressor.coef\_)*

*y\_pred = regressor.predict(X\_test)*

*acc = regressor.score(X\_train,y\_train)\*100*

*from sklearn import metrics print('Mean Absolute Error:', metrics.mean\_absolute\_error(y\_test, y\_pred)) print('Mean Squared Error:', metrics.mean\_squared\_error(y\_test, y\_pred))*

*print('Root Mean Squared Error:', np.sqrt(metrics.mean\_squared\_error(y\_test, y\_pred))) print("Accurarcy :",acc)*

*### XGBoost %%time*

*import xgboost as xgb*

*xg\_reg = xgb.XGBRegressor(objective ='reg:squarederror', colsample\_bytree =*

*0.3, learning\_rate = 0.2, max\_depth = 10, alpha = 15, n\_estimators = 1000)*

*xg\_reg.fit(X\_train,y\_train)*

*y\_pred = xg\_reg.predict(X\_test)*

*from sklearn import metrics print('Mean Absolute Error:', metrics.mean\_absolute\_error(y\_test, y\_pred)) print('Mean Squared Error:', metrics.mean\_squared\_error(y\_test, y\_pred))*

*print('Root Mean Squared Error:', np.sqrt(metrics.mean\_squared\_error(y\_test, y\_pred))) plt.scatter(y\_test,y\_pred,alpha=0.5) plt.plot(y\_test,y\_test,color='red')*

*### Random Forest from sklearn.ensemble import RandomForestRegressor*

*regressor = RandomForestRegressor(n\_estimators=70, random\_state=0) regressor.fit(X\_train, y\_train)*

*y\_pred = regressor.predict(X\_test) acc = regressor.score(X\_test,y\_test)\*100*

*from sklearn import metrics print('Mean Absolute Error:', metrics.mean\_absolute\_error(y\_test, y\_pred)) print('Mean Squared Error:', metrics.mean\_squared\_error(y\_test, y\_pred))*

*print('Root Mean Squared Error:', np.sqrt(metrics.mean\_squared\_error(y\_test, y\_pred))) print("Accurarcy :",acc)*

**CHAPTER 7**

**REQUIREMENTS**

### 7.1 Hardware requirements

Processor: Intel Multicore Processor (i3 or i5 or i7)

RAM: 4GB or Above

Hard Disk: 100GB or Above

## 7.2 Software Requirements

**Programming Language:** Python

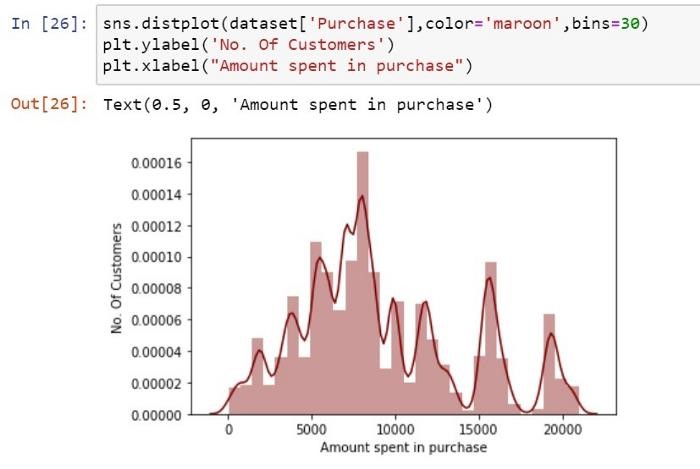
**Operating System:** Windows or Linux **Tools:** Anaconda Navigator, Seaborn, sklearn

**CHAPTER 8**

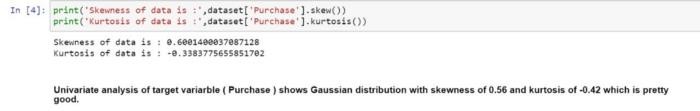
# PROJECT FINDINGS

**Purchase — Target Variable**

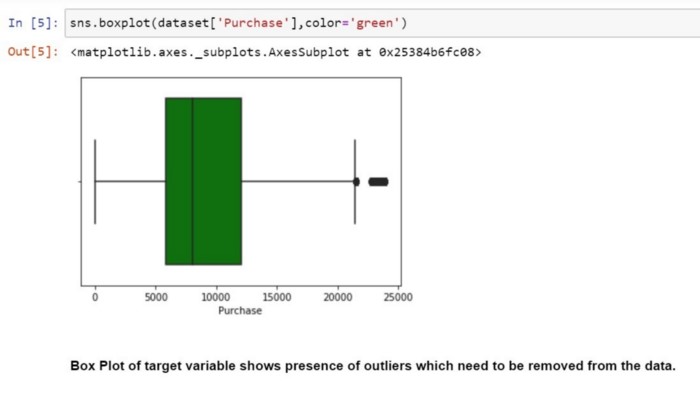
To analyze the distribution we constructed a distplot of the data:



From the Distplot it looks like the data follows Gaussian(Normal) Distribution. We then perform the univariate analysis:



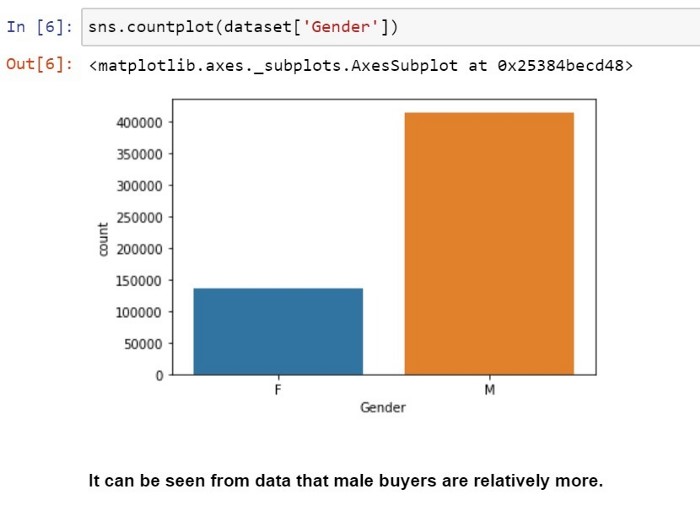
For Outlier Analysis we constructed a Box Plot:



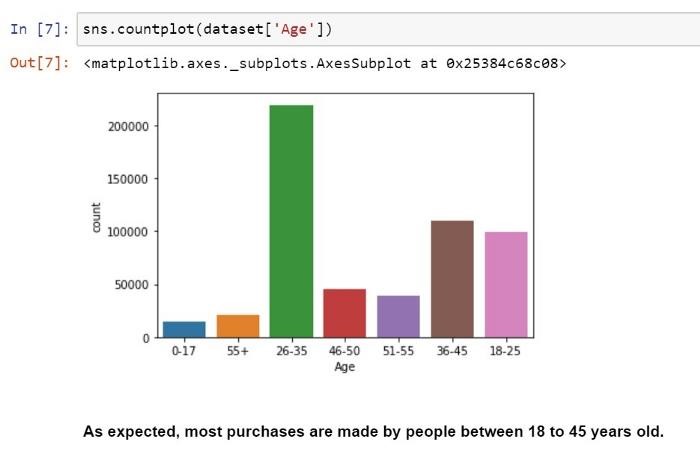
Since the Target Variable Purchase is a continous valued attribute. In order to predict it we have to use the various Regression Models of Machine Learning.

### Gender

Constructing a Countplot of Gender attribute:



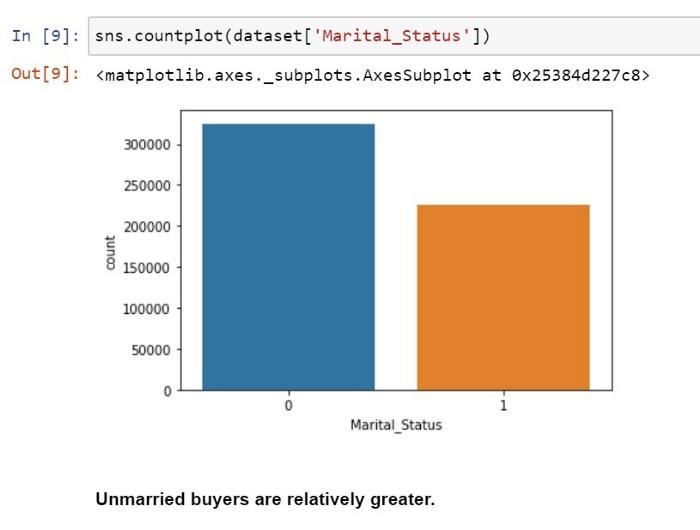
### Age



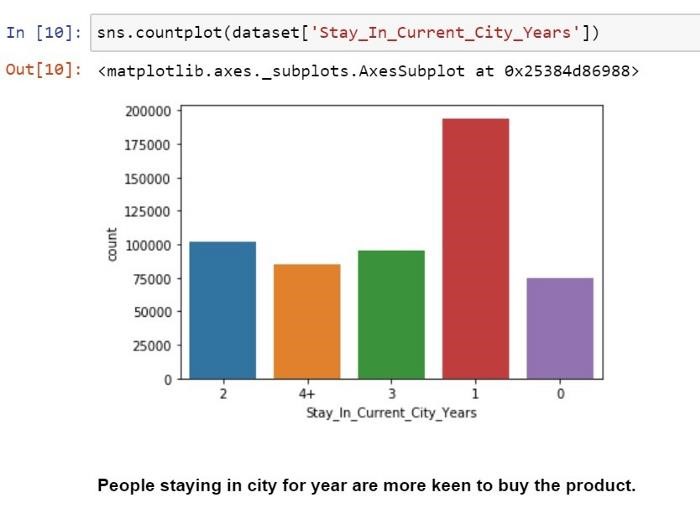
### City Category



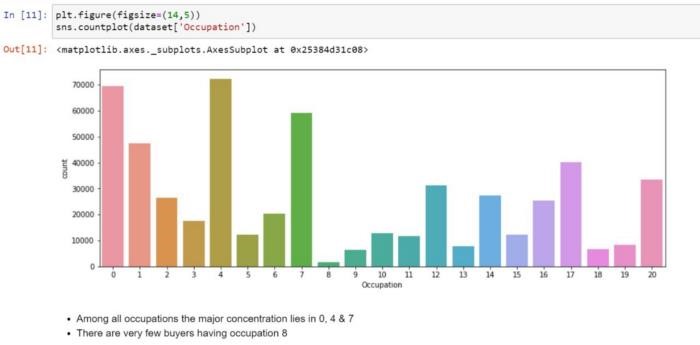
### Marital Status



### Stay in Current City



### Occupation



The analysis indicates that the attributes Occupation, Age and Gender will have a significant impact on the predicted values.

### Model Selection

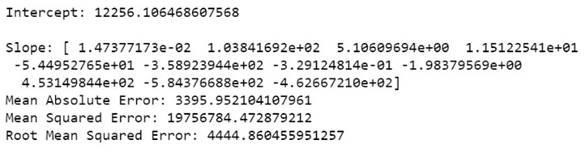
We tried 3 Regression Models for this dataset. The model that gives the highest Accuracy is the most suitable model. **Criteria for Accuracy:**

For regression models we calculate accuracy by finding out

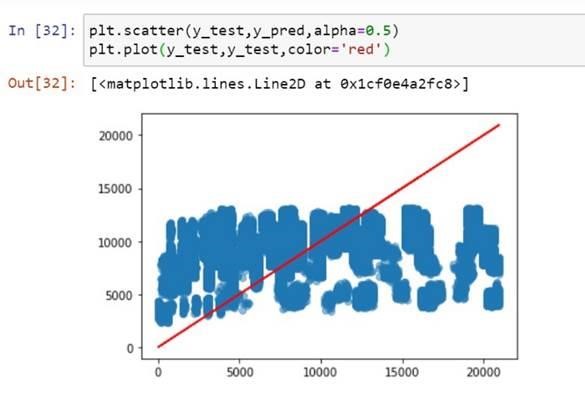
i. Mean Absolute errorii.Mean Squared erroriii. Root mean Squared error (RMSE) iv. Accuracy

Hence RMSE is the final deciding factor. For visualization we plot a graph of y\_actual vs y\_predicted. The red line shows how the scatter plot would have been if the accuracy was perfect(RMSE). The blue points is the actual scatter plot shows that depicts deviation from actual results. The farther the spread from the red line worse the model.

### 1. Multiple Linear Regression

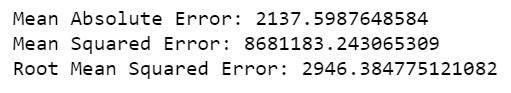


RMSE: 4444

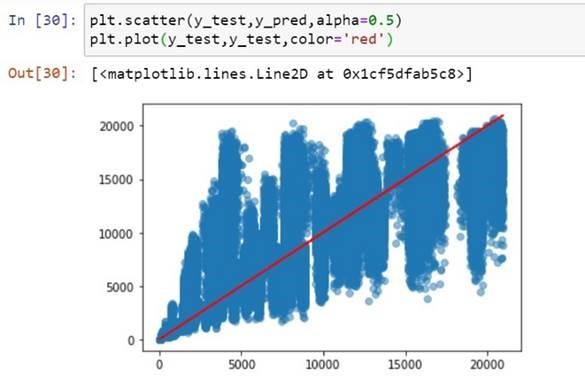


As we can see this model gives RMSE above 4000. Even from the graph we can see that this model provides very poor accuracy. In the cases where Purchase value is above 15000 the model fails completely.

### 2. Random Forest Regressor

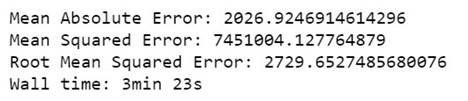


RMSE: 2946

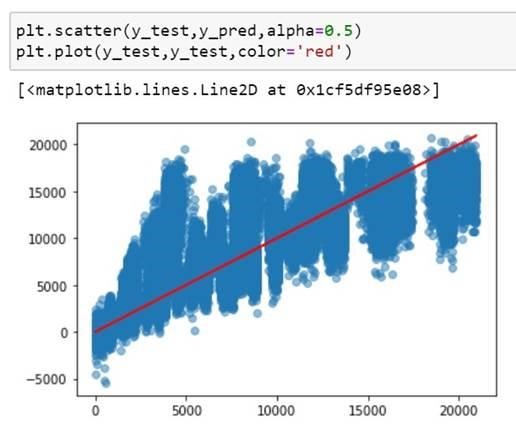


Random forest model gives results significantly better than Linear Regression. Also it does not fail for values above 15000. The RMSE value is approx 2900.

### 3.XGBoost Regressor



RMSE: 2729



XG Boost produces results similar to Random Forest however the RMSE is slightly better than Random Forest with an RMSE of 2700. Hence it is the most suitable model.

Out of the three models XGBoost provides the highest accuracy with an RMSE of 2729. This a pretty good accuracy if we compare it to the leaderboard on Analytics Vidya. This accuracy get us into the top 37%. With a little more effort an accuracy of 2400 can be achieved via Hyperparameter tuning.

**CHAPTER 9**

# CONCLUSION

We conclude that the complex models like neural network are an overkill for simple problems like regression. And simpler models along with proper data cleaning perform well for the regression. Also, based on the current trend, the number of shoppers on the Black Friday is only going to increase. The study agrees that machine learning techniques produce better prediction models that can be used at stores and the store owners can analyze their customer base to better target the customers and increase the sales on a Black Friday. The study also agrees that the data must be pre-processed to attain an effective dataset for developing the prediction model. Several techniques were discussed in this study to attain the best model. However, there is still no definite solution as to what the correct technique is to attain a model with high accuracy. To improve the results, a dataset with sufficient features and increase in quantity must be obtained. Further research must be conducted in enhancing the existing machine learning techniques to work in real time and develop an efficient model. Also, the models developed must be tested on data with different volumes to test its scalability and performance. In future work, the result of regression on balanced dataset can be studied by changing the data distribution. This can be done by selecting a sample of dataset or removing certain records to balance the type of data.

Out of the three models XGBoost provides the highest accuracy with an RMSE of 2729. This a pretty good accuracy if we compare it to the leader board on Analytics Vidya. This accuracy get us into the top 37%. With a little more effort an accuracy of 2400 can be achieved via Hyperparameter tuning.

**CHAPTER 10**

# FUTURE ENHANCEMENTS

The fundamentals of machine learning as well as the associated data processing and modelling algorithms have been discussed in this paper. Following this discussion is an application of these algorithms to the task of forecasting sales for Black Friday at a number of different locations. As a result of the implementation, the prediction results show a correlation between the various factors that were taken into consideration and how a specific location of medium size recorded the highest sales, which suggests that other shopping locations should follow similar patterns in order to improve their sales.

This method of sales forecasting can be made to be more forward-thinking and fruitful by making use of a number of different instances, parameters, and factors. When it comes to systems that are based on predictions, accuracy is one of the most important factors, and it is possible to significantly improve it by increasing the number of parameters that are used. Additionally, an investigation into the operation of the submodels can result in an increase in the overall productivity of the system. Because of the Internet of Things (IoT), further project collaboration is possible in the form of a web-based application or in any device supported with an in-built intelligence. Both of these options make the project more practically applicable. In addition, the various stakeholders who are concerned with the sales information can provide additional inputs to assist in the generation of hypotheses, and additional instances can be taken into consideration in order to generate results that are more precise and which are closer to reflecting real-world circumstances. It is possible that the traditional means, when coupled with efficient data mining methods and properties, will have a higher and more positive impact on the overall development of the tasks that are performed by the corporation as a whole. The more expressive regression outputs, which are also more understandable and are bounded with some degree of accuracy, are one of the primary highlights. In addition, the flexibility of the approach that has been proposed can be increased with variants at an extremely pertinent stage of the process of building regression models. In order to accurately assess and optimise, there is a further requirement for experiments to ensure accurate measurements of both the accuracy and the resource efficiency.

**CHAPTER 11**

**REFERENCES**

1. <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8663760&tag=1>

1. <https://link.springer.com/chapter/10.1007/978-3-030-90055-7_2>

1. <https://www.mdpi.com/0718-1876/16/5/83/htm>

1. [https://www.businessmanagementideas.com/project-report/consumer-behaviourproject-report/project-report-on-consumer-behaviour-marketing/8216](https://www.businessmanagementideas.com/project-report/consumer-behaviour-project-report/project-report-on-consumer-behaviour-marketing/8216)

1. [https://towardsdatascience.com/customer-segmentation-analysis-with-python6afa16a38d9e](https://towardsdatascience.com/customer-segmentation-analysis-with-python-6afa16a38d9e)

1. <https://medium.com/@omigirish1999/black-friday-sales-prediciton-c4f2d141f5fa>