**BLACK FRIDAY SALES PREDICTION**

**INTRODUCTION:**

Black Friday is an institution that was introduced in the United States and it is celebrated the day after Thanksgiving for several decades ago, with initial reports of the term found in the early 19th century (Bell et al. 2014). This day is considered to be a trade celebration marking the beginning of the Christmas holiday season. For several decades, Black Friday has been widely recognized as the largest shopping day of the year in the US. For most retailers, it is the busiest day of the year. According to Statista, $534 million were spent on Black Friday in 2008. In 2013, US consumers spent more than $50 billion and over 92 million people visited brick-and-mortar stores (Grannis, 2014). That number has since skyrocketed to $1.5 billion in 2014. Black Friday is traditionally the day with the highest purchasing volume of the year which accounts for 30% of the annual turnover of retail stores and in some product categories such as jewelry, this percentage reaches 40% (www.fundivo.com). Moreover, Black Friday has been considered as a unique consumption ritual that blends elements of traditional shopping with holiday rituals (Thomas and Peters, 2011).

In anticipation of this day, retailers usually add employees, increase their inventories, prepare new promotions and change store layouts. Retailers have implemented numerous strategies for drawing customers to the stores. Many stores open their doors at 5 a.m., or earlier, to accommodate eager shoppers; some even remain open all night. Taken together, these factors combine to create a unique shopping event that, despite its popularity, has received surprisingly limited research attention. Retailers advertise heavily and promote special sales to attract customers. Many businesses provide incentives, such as ‘door buster’ and ‘early bird’ sales, to motivate consumers to visit their store—and to come early (Horovitz 2009). Most large retailers post their Black Friday advertisements, coupons and offers online well in advance, giving shoppers insight about upcoming sales so they can plan their shopping and Black Friday strategies (Katz 2008).

On Friday, November 25, 2016 for the first time the Black Friday practice was introduced in Greece. According to the Greek Trade Union (ESEE), retailers responded differently to the event. Thus, 1040% of retailers participated in Black Friday depending on the city. Mostly large chain stores participated by offering sales from 10-40%. Although its global widespread, there has been a lack of scientific research attention to consumers' attitudes toward and behaviors during Black Friday. Mostly there is a lack of quantitative research on the topic. Most recent available studies take a qualitative approach investigating the phenomenon from a ritualistic scope (Bell, Weathers, Hastings and Peterson, 2014; Thomas and Peters, 2011) and to the authors’ knowledge no quantitative study is available that examines the attitudes and the behavior of consumers toward Black Friday, especially when this institution is introduced for the first time in a country. Therefore, the purpose of the study was to study the behavior, attitudes and intentions of Greek consumers towards Black Friday. Moreover, the objective of this research was to understand differences between shoppers on Black Friday and non-shoppers and to suggest how retailers can harness these differences in terms of marketing.

**OBJECTIVES OF RESEARCH:**

**Predicting purchase**  Build simple machine learning model that can be predict how much a customer is likely to spend on the eve of Black Friday.

**Pattern recognition**

Reveal and understand the most important factors from predictors such as Age,gender,stay in years etc,that influence the spending of a customer.

Establish a quantitative impact of the revealed factors and how they influence purchase by a customer on a personal leveli.e whether they have a positive or negative contribution on the purchase.

**Problem Statement:**

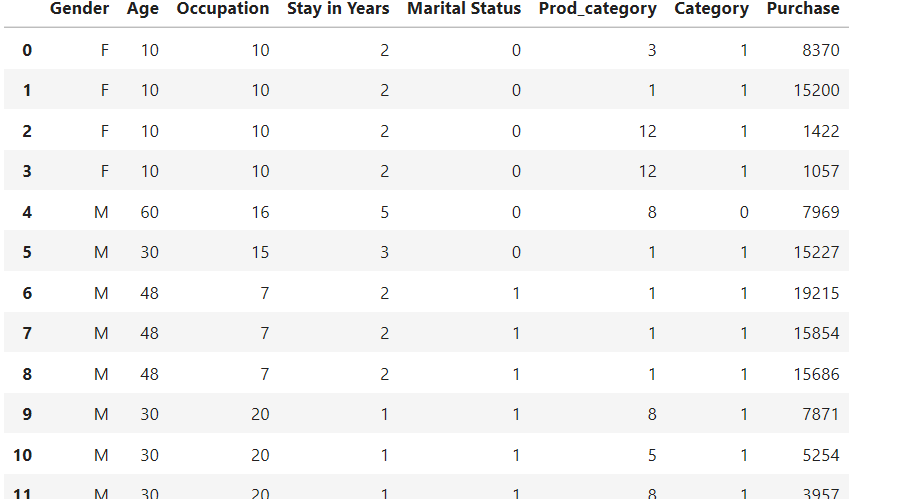
The challenge was to predict purchase prices of various products purchased by customers based on historical purchase patterns. The data contained features like age, gender, marital status, categories of products purchased, city demographics etc.

**Review of Literature:**

Black Friday sales is trade celebration marking the beginning of the Christmas holiday season. The goal of black friday sales prediction is to predict the purchasing amount of products by customers. So for this we need the previous data records of sales conducted .From that data it is possible to predict the purchasing amount of products by machine learning concepts using multiple models. In this case we prefer decision tree model in which we can get accurate results. Finally it is somewhat problematic with black friday by some misbehaviours of customers and emotions and many more.The predicting analysis may be changed by some major strategies such as merchandising,customer segmentation, resource identification . Black friday is a memorable day of the year because of its greatness acheieved by greatest deals of all products.

**DATA COLLECTION:**

Collection of data is the most important task in machine learning.It is the key component which is used to predict the result what we want. One of the most data providing source ,www.kagggle.com is the place where we can get ant dataset. This is the place where we got this dataset. It includes the information about user id, age, occupation , product category, gender and purchase amount. Purchase is the dependent variable or output which depends on above following factors.

****

**METHODOLOGY:**

Before looking at the data it is important to understand how does the company expect to use and benefit from this model? This first brainstorming helps to determine how to frame the problem, what algorithms to select and measure the performance of each one.

We can categorize our Machine Learning (ML) system as:

**Supervised Learning task:** we are given labeled training data (e.g. we already know how much a customer spent on a specific product);

**Regression task:** our algorithm is expected to predict the purchase amount a client is expected to spend on this day.

**Plain batch learning:**since there is no continuous flow of data coming into our system, there is no particular need to adjust to changing data rapidly, and the data is small enough to fit in memory, so plain batch learning should work.

The decision tree is a machine learning model which gives us accurate results.so we are using this for our data. Decision Trees are a type of Supervised Machine Learning (that is you explain what the input is and what the corresponding output is in the training data) where the data is continuously split according to a certain parameter. The **tree** can be explained by two entities, namely **decision** nodes and leaves.

**Exploratory Data Analysis (EDA):**

We’ve made our first assumptions on the data and now we are ready to perform some basic data exploration and come up with some inference. Hence, the goal for this section is to take a glimpse on the data as well as any irregularities so that we can correct on the next section, **Data Pre-Processing.**

Univariate Analysis:

To get an idea of the distribution of numerical variables, histograms are an excellent starting point. Let’s begin by generating one for Purchase, our target variable.

Numerical Predictors:

Now that we’ve analysed our target variable, let’s consider our predictors. Let’s start by seeing which of our features are numeric.

**Distribution of the variable: Occupation**

As seen in the beginning, Occupation has at least 20 different values. Since we do not known to each occupation each number corresponds, is difficult to make any analysis. Furthermore, it seems we have no alternative but to use since there is no way to reduce this number as we did on  with Item\_Type.

**Distribution of the variable : Marital\_Status**

As expected there are more single people buying products on Black Friday than married people, but do they spend more?

**Distribution of the variable : Product\_Category\_1**

From the distribution for products from category one, it is clear that three products stand out, number 1, 5 and 8. Unfortunately, we do not know which product each number represents.

**Data Preprocessing:**

During our EDA we were able to take some conclusions regarding our first assumptions and the available data:

**Age :** should be treated as numerical. It presents age groups.

**City\_Category:**We can convert this to numerical as well, with dummy variables. Should take a look at the frequency of the values.

**Occupation**: It seems like it has at least 16 different values, should see frequency and try to decrease this value.

**Gender:** There are possibly two gender, we can make this binary.

**Product\_ID:** Should see if the string “P” means something and if there are other values.

**Stay\_In\_Current\_City\_Years:**We should deal with the ‘+’ symbol.

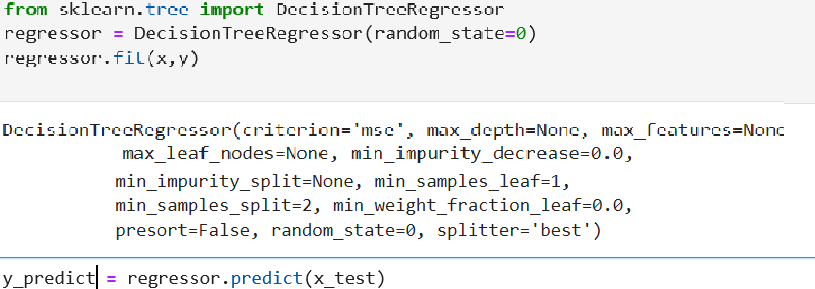
**Product\_Category\_1**and **Category**: Have NaN values.

Usually, datasets for every challenge such as those presented in Analytics Vidhya or Kaggle come seperated as a train.csv and a test.csv. It is generally a good idea to combine both sets into one, in order to perform data cleaning and feature engineering and later divide them again. With this step we do not have to go through the trouble of repeting twice the same code, for both datasets. Let’ s combine them into a dataframe datawith a sourcecolumn specifying where each observation belongs.

**Statistical Techniques:**

 we will import the train\_test\_split function to divide all our data into two sets: training and testing set. The training set will be used to fit our model. Training data is always used for learning, adjusting parameters of a model and minimizing an error on the output. The rest of the data (the Test set) will be used to evaluate performances.

**Data modelling:**



Decision tree regression observes features of an object and trains a model in the structure of a tree to predict data in the future to produce meaningful continuous output. Continuous output means that the output/result is not discrete, i.e., it is not represented just by a discrete, known set of numbers or values.

By fitting decision tree regressor to the dataset we can predict the value of purchase.

**Conclusion:**

Machine learning can be used for a variety of tasks. In this article, we used a machine learning algorithm to predict the amount that a customer is likely to spend on Black Friday. We also performed exploratory data analysis to find interesting trends from the dataset. For the sake of practice, I will suggest that you try to predict the Product that the customer is more likely to purchase, depending upon his gender, age, and occupation.