Data Analytics

BLACK FRIDAY

(A study of sales through consumer behaviors on Black Friday)

CWID	First name	Last Name	IIT Email
A20401921	Anusha	Satish	athattehalli@hawk.iit.edu
A20436206	Sivaranjani	Prabasankar	sprabasankar@hawk.iit.edu

Type of Project: Multiple linear regression

Introduction

Black Friday is an informal name for the Friday following Thanksgiving Day in the United States, which is celebrated on the fourth Thursday of November. It has routinely been the busiest shopping day of the year for a much longer time in the United States.

Many stores offer highly promoted sales on Black Friday and open very early, such as at midnight, or may even start their sales at some time on Thanksgiving. Many non-retail employees and schools have both Thanksgiving and the following Friday off, which, along with the following regular weekend, makes it a four-day weekend, thereby increasing the number of potential shoppers.

Black Friday relies on a few simple retail strategies that, with tons of customer data and forecasting software, have become precise and planning for Black Friday is key for many retailers, particularly in predicting consumer interest in product ranges, which many retailers got wrong last year. It must be carefully planned for every year to ensure orders can be fulfilled without compromising on the level of customer service and seamless delivery. This includes additional staff across shops and accounting for the increased demand.

To analyze and discover the various aspects that determine the purchase based on some factors like products and their categories, location, the age of customers which may benefit the retailers, induced us for selecting black Friday dataset.

Data Sets

The dataset has sample of the transactions made in a retail store. The store wants to know better the customer purchase behavior against different products where we are trying to predict the dependent variable (the amount of purchase) with the help of the information contained in the other variables.

To work on this project, we have chosen the dataset provided by Mehdi Dagdoug to predict Black Friday sales based on various parameters such as Product category, Customer age, gender and location etc.

There are more than half a million (550 000) records available to train the models which we would be using to test and predict the sales with a 95% confidence level.

Data Analytics

This dataset has been retrieved from the link: https://www.kaggle.com/mehdidag/black-friday

Research Problems

1. We have columns represented in both categorical and numerical features. We need to confirm if all the predictors have a significant effect on the Purchase
2. Product Category 2 &3 has lot of missing values which needs to cleanse by using appropriate method.
3. It might be useful if we represent the Stay_In_Current_City_Years category numerically rather than categorically.
4. Identify multicollinearity problem among the determining factors if any.
5. Identify and eliminate influential points which might affect our model in predicting the purchase.

Potential Solutions

1. Confirm if all the predictors (categorical and numerical features) have a significant effect on the Purchase using linear regression.
2. Fill the missing data automatically in Product Category 2 &3 by using Global constant fill, Attribute mean value or Attribute mean value for all the samples belonging to same class.
3. Convert 'Stay_In_Current_City_Years' from categorical to numeric using R built-in functions.
4. Identify multicollinearity problem using VIF (variance Inflation Factor) and solve by Post processing
5. Use Cooks distance and eliminate influential points which may affect our model.

Evaluations

Considering the size of data set, we would choose hold-out evaluation to split the data instead of N-fold cross evaluation method. We build multiple models as part of this project and identify the most accurate model using RMSE value and use other techniques like feature selection and goodness of fit to improve our model.

Expected Outcomes

The expected outcome would be the customer behavior on purchases to benefit the retailers during black Friday sales.