

## **Practice Problem: Black Friday**

Datahack analyticsvidhya



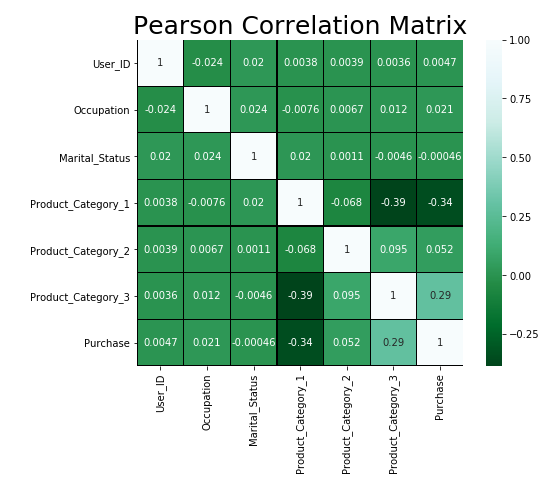
October 5, 2018

By-Yuvraj khanna

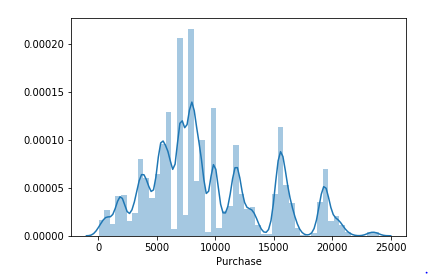
Roll no-17EC30040

# ***Understanding Problem Statement:***

After visualising the data and plots we can clearly see the relations between the features of the given dataset vs purchase cost. This can be easily verified by checking the Correlation matrix given the code file.



After this before we try to process out data we also need to check if our Purchase cost is skewed or not as if it is skewed we need to apply some functions like log(1+x) to remove its skewness.



As We can clearly see this is not a good distribution because of the random increase and decrease but it is clearly not skewed so we can use it for computation.

# ***Method of Analysis and Algorithm used:***

***Possible methods: 1. Linear Regression***

***2. random forest regressor***

***3. XG Boost***

2nd and 3rd methods are too computationally expensive so we are going to use Linear Regression for our model.

***Linear Regression:***

The method used is Linear Regression method.

In linear regression method, we try to map a linear relation between features and the output.

*ho=o0+o1X1+o2X2+......*

For initial o0,o1,o2,etc. we calculate the Mean Square Error which will be function of o0,o1,o2..... and try to minimize it to lowest possible value using methods such as Ordinary Least Squared (OLS) or Gradient Descent method.

The given values of o0,o1,o2... give the required linear regression curve.

We don’t need to write the whole code for this since there are already many python libraries that do it for us very efficiently. One of those examples are sklearn.linear\_model which we will be using in our code.

# ***Result:***

We have used 2 models for this regression problem:

1. Basic linear Regression model with data pre-processing.

***RMSE:*** 4179.18

1. Linear Regression model with data pre-processing and feature engineering.

***RMSE:*** 2807.84

We submit this code and our predictions to the datahack.analyticsvidhya and securing position on the leader board as showed below: