FLIGHT PRICE PREDICTION

Developed By-

Mohammad Daim Khan
Manas Chandan Behera

Introduction

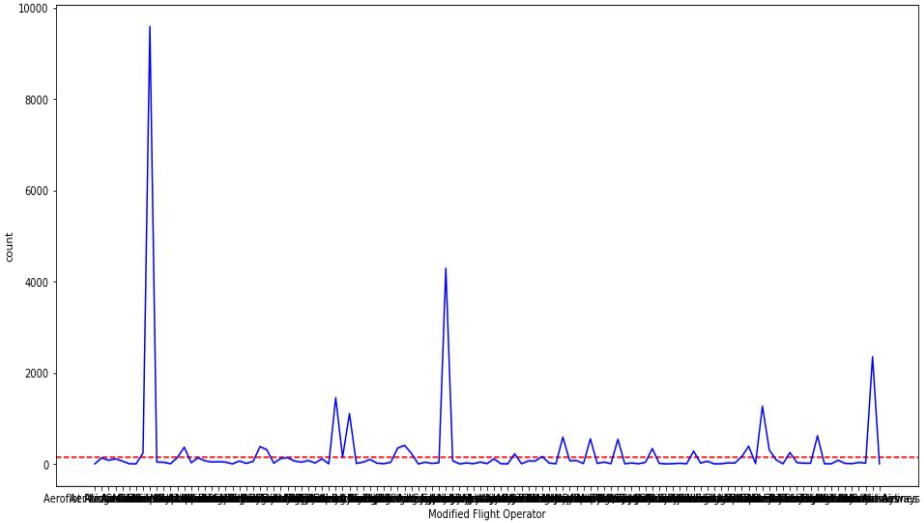
As we know, Flight ticket prices can be something hard to guess,

So we aimed to predict the price for the future travel dates.

To achieve this, we used a web scrapped ease my trip's data for the prediction

Dataset : From Kaggle(Make My Trip Dataset) : Ratio Division 3:1 [Train:Test]

Encoded Airlines

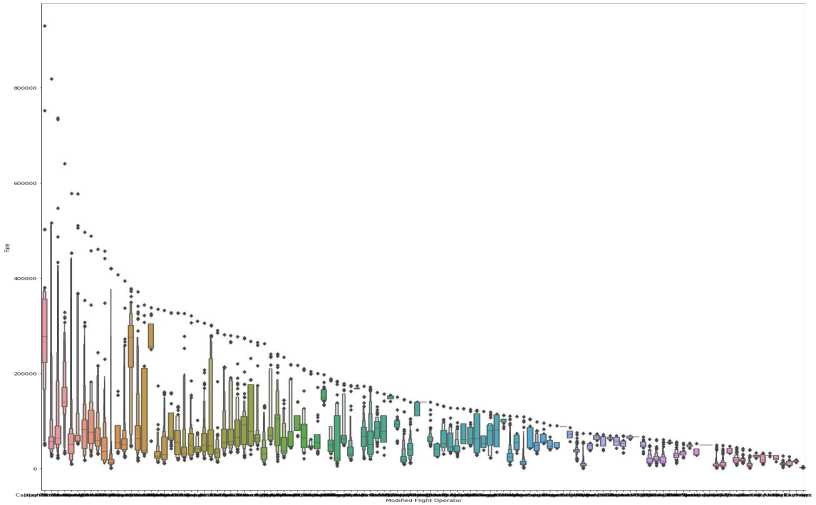


1) Airlines Count Dependency

1st plot= Airlines against count.

Inference= some Airlines has major usage.

Red Line is a median line dividing graph on 0.05 percent of count.



2.) Modified Flight Count

Catplot for checking outliers of airlines asper their price range.

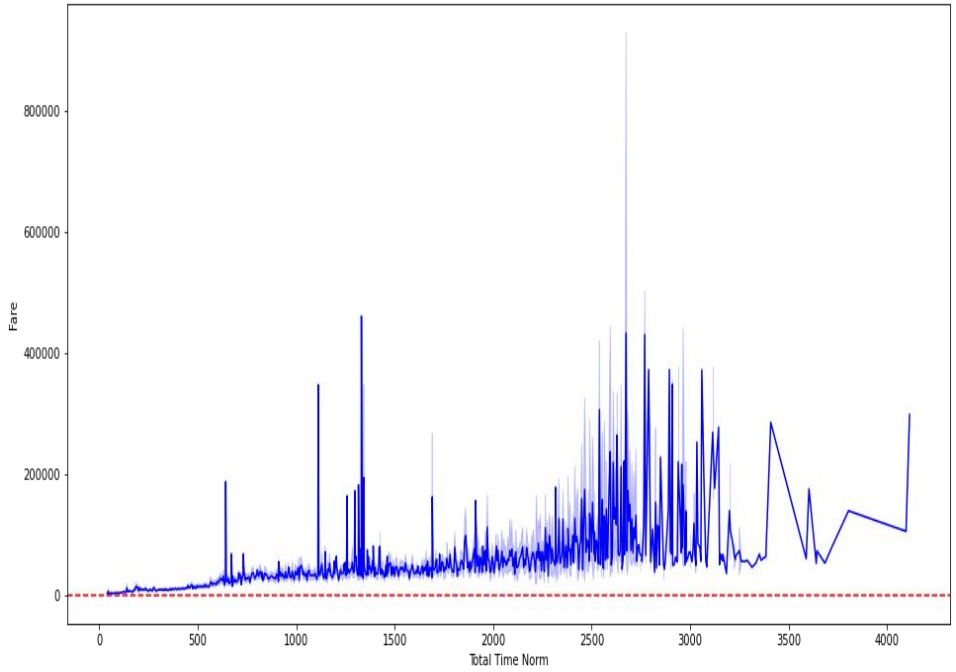
Lower Outlier = Q1 – (1.5 * IQR)
Higher Outlier= Q3 + (1.5 * IQR)

Time Taken Normalization

1) Preprocessing on the basis of the data.
   Converted the ddmmyy to minutes

2) Dependency of the Price on the basis of the time of fly.

Longer Duration $\propto$ Higher Price



Source Count

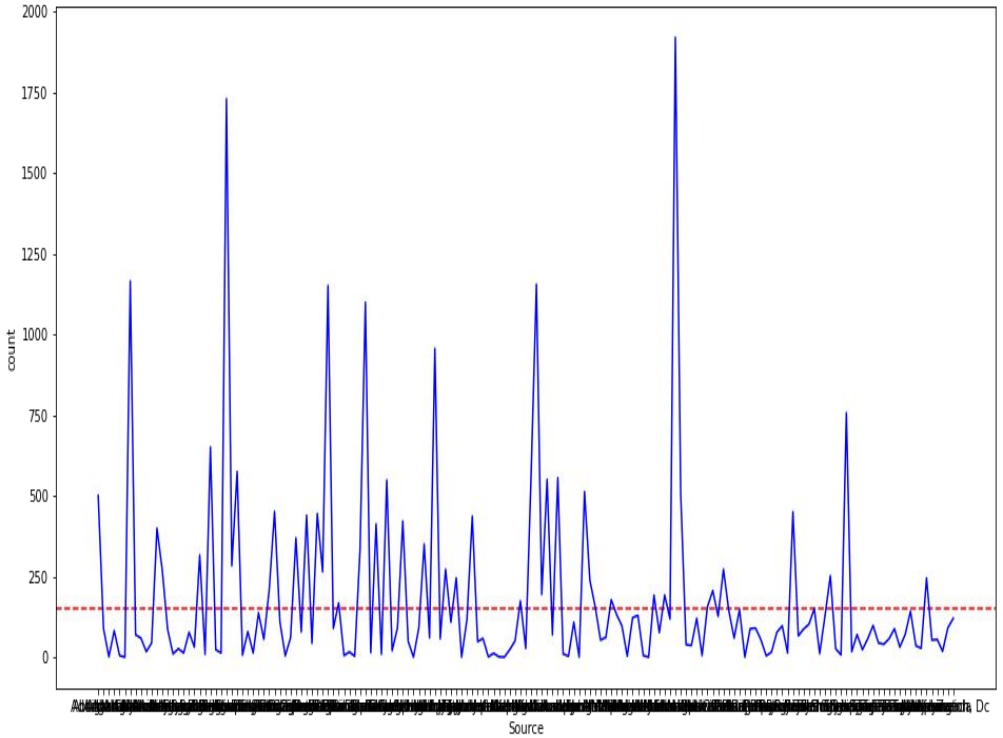
Line Plot of Source vs Count

Red line is divider on basis of count having more than 1 percent.

Inference:

Some Source have major usage. And are high above in one percent range.

Possible Feature



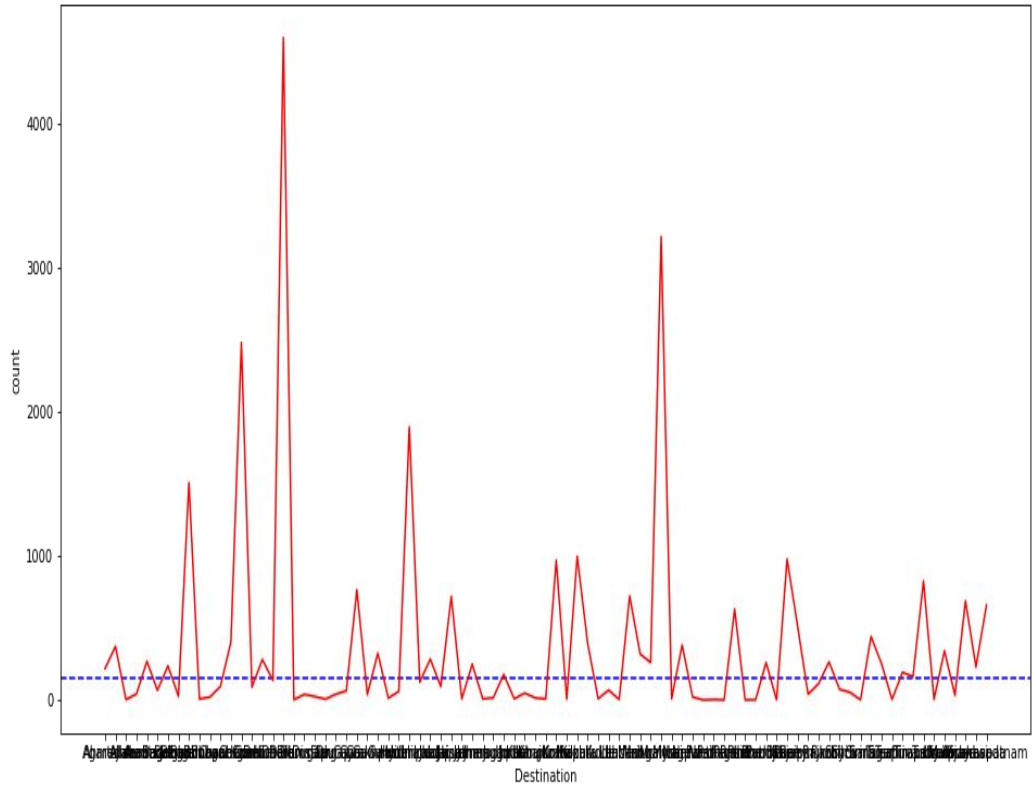
Destination Of The Fly

Line Plot of Destination vs Count

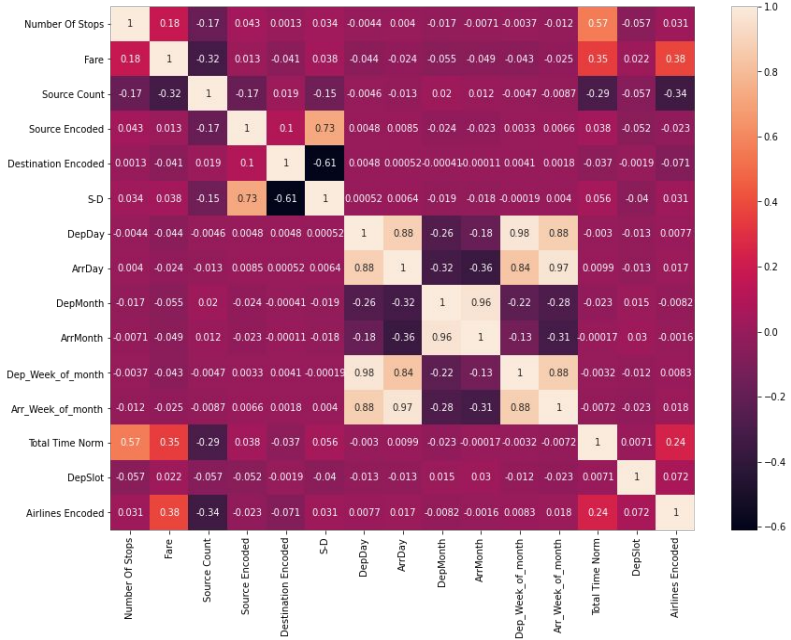
Red line is divider on basis of count having more than 1 percent.

Inference:

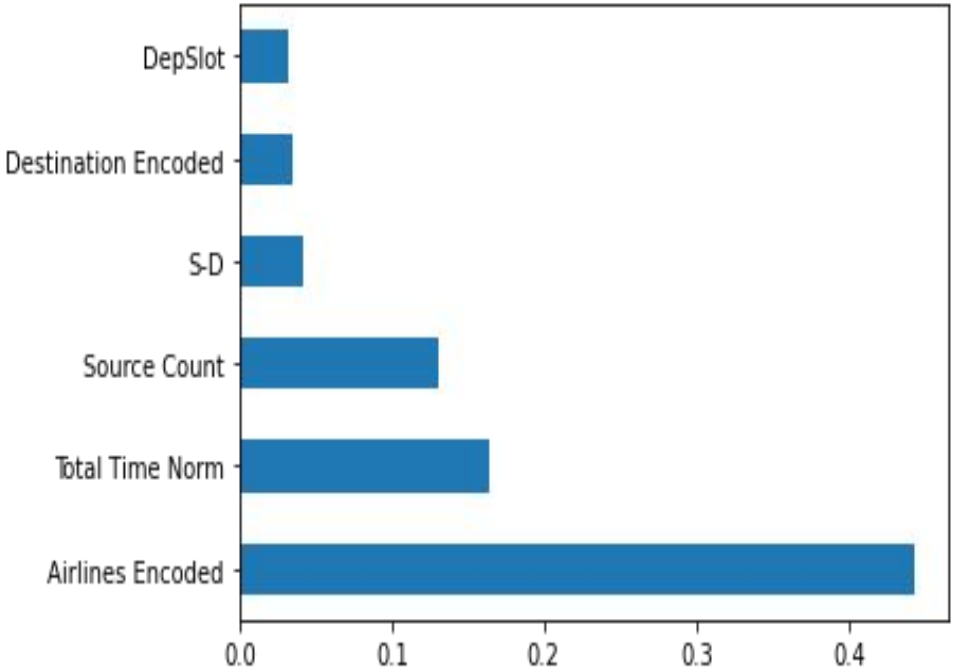
Possible pair formation of higher source count with their destination count.



Feature Selection



FILTER SELECTION(CORRELATION)



EMBEDDED METHOD(TREE BASED)

Algorithm Analysis

ALGORITHM	OBTAINED R2 SCORE
Linear Regression	0.21
Support Vector Regressor	0.10
XGBoost (Tree based)	0.48
Random Forest	0.61

Thank You

Mohammad Daim Khan

Manas Chandan Behera

Daim's Github :
https://github.com/Daim-Nickel-Penny

Manas's Github
https://github.com/ManasChandan