Airline Pricing Report

By

Nikhil Chavan
IIT Bombay
nschavan1996@gmail.com

Guide:
Prof. Sameer Mathur,
IIM LUCKNOW

We have given the classified data set and from the given data set we are trying to answer the following crucial question given in the problem statement:
what factors explain the difference in price between an economy ticket and a premium-economy airline ticket?

While doing the analysis of this problem statement, by dividing the task in smaller R codes and running them on Rstudio and based on the results that I have got, I am writing the following conclusions :

1. Positive correlation is found between following data types :
    - flight duration and type of tickets
    - Price_economy, Width_economy, flight duration, quality and international flights
    - Quality, pitch and width premium
    - Price of each class, pitch and width
2. Important predictor for price in both the classes is flight duration and price relative
3. The r squared and the adjusted r squared is relatively high.
4. High r squared values suggest that pitch, width, relative price, quality and flight duration add to the price of a ticket.
5. We get an accuracy of 0.563 using the Random Forest Algorithm on the same set of independent and dependent variables.
6. From the VIF plot we see that Flight Duration and Price Relative are most important factors in predicting Price Economy.
7. Regression tree analysis gives accuracy of 0.467
8. Linear regression model accuracy is 78%
9. The model has an F Statistic of 48.4 which is moderately high the t value of Pitch_premium, width_premium, Price_relative and quality is positive which shows that these predictors are associated with Price_Premium.
10. The MinMax accuracy is relatively high, whereas MAPE is moderate to low, which are good indicators.
11. From the VIF plot we see that Flight Duration and Price Relative are most important factors in predicting Price Economy.
12. the t value of Pitch_economy and quality is positive indicating that these predictors are associated with Price_economy. A larger t-value indicates that that it is less likely that the coefficient is not equal to zero purely by chance. as the p-value for Flight_Duration and Price_Relative is less than 0.05 they are both statistically significant in the multiple linear regression model for Price_Economy response variable.