Airline Analysis

Purbasha Chatterjee

[pchatterjee10@gmail.com](mailto:pchatterjee10@gmail.com)

Oregon State University, Corvallis

The dataset SixAirlines.csv includes the variables for six different airlines. The major analysis of the dataset is based on the question *What factors explain the difference in price between an economy ticket and a premium-economy airline ticket?*

Some of the inference that I could draw by analyzing the given dataset are as follows:

* The maximum number of economy seats are sold in the price of around 500 and 2000.
* The maximum number of premium seats are sold at a price of more than 3200
* British airlines has maximum number of seats
* Flight duration are more frequent at an hour of 8 and 10.
* July has the least amount of travel
* International flights are more than domestic flights
* Flights with 122 and 303 seats are maximum
* Only Singapore airlines provides late hours flight whereas for Delta and Jet – there are no flights after 6. British airlines has a wide range in quartiles as compared to any other airlines.
* Price relative is minimal in Air France and Delta airways whereas in Jet, Singapore and Virgin – it has widespread quartiles
* For British airlines upper quartile is higher for premium seats and lower quartile for economy seats
* As per the corrogram, economy seats an premium seats are strongly correlated
* Pitch economy and pitch premium are negatively correlated whereas Price of the economy seats and the price of the premium seats are highly correlated
* The economy seat is positively correlated to the factors – price, width of economy seats and number of seats
* The premium seats are strongly correlated to width of the economy seats, price of the premium seats and total number of seats whereas it is negatively correlated to pitch of the premium, width of the premium and relative price
* The flight duration makes a very strong correlation with the price of both the economy as well as premium seats. It also correlates strongly with the width of the seats
* As per the correlation test pitch of the economy seats and price are correlated, this is also the same for premium seats
* As per the t-test, the airlines and the quality has p-value less than 0.05 so they are statistically significant. Additionally, t-test states that airlines along with price of either economy or premium are statistically significant with p-value < 2.2e-16
* For seats are price of economy as well as premium. It was observed by t-test that are too statistically significant. But all of these test gives a negative value for t
* Applying linear regression over the dataset over width, pitch and price generates for economy seats F-statistic: 30.62 on 3 which is relatively high based on 3 dataset. It maintains positive coefficients for width, pitch and price of the economy seats. It gives p-value: < 2.2e-16 with residual error as 71.38
* Applying linear regression over width, pitch price for premium seats generates F-statistic: 8.259 on 3. It has negative coefficients with premium width and premium pitch but maintains a positive coefficient with premium price. It gives p-value: 2.316e-05 with residual error as 12.95