**Airline Pricing Mini Project**

**By**

Name : Mayank Sagar

[mayanksagar.26@gmail.com](mailto:mayanksagar.26@gmail.com)

Phone : 9620834479

College: R.V.College Of Engineering, Bengaluru.

File name : SixAirlines.csv

Introduction : The data set with which we are dealing with here is a classification data set about six airlines companies and their prices of premium economy and economy seats with other parameters .So the main purpose of the data is used to classify between economy prices and premium prices of tickets in a flight.

The analysis and inferences from the SixAirlines data set are as follows:

1. The most expensive airline among all six is the Airfrance airlines.
2. The most affordable and cheap airline among all is Jet airlines.
3. The number of seats in Jet airlines is the least among all six for economy class as well as premium economy class.
4. The most number of economy class seats are offered by Singapore airline
5. The most number of premium class seats are offered by British airline.
6. International flights are more than domestic flights.
7. The price of the ticket for economy class and premium economy class of Airbus aircrafts is more than Boeing aircraft.
8. The price for flights for less duration is less for economy as well as premium economy class.
9. Most number of flights are by British airlines.
10. Low range Economy tickets are more frequent.
11. There is a positive correlation between Price of economy ticket and the flight duration and price of premium economy ticket and flight duration.
12. Positive correlations are seen in case of Price\_Economy between quality and international flight and width\_economy and flight duration.
13. Positive correlations are seen between, quality and pitch and width premium, width\_premium and international, pitch\_premium and international.
14. Positive correlation is seen between prices of each class and pitch and width.
15. Most important predictor for price Economy is Flight duration and price relative.
16. Most important predictor for price Premium is Flight duration and price relative.
17. Hypothesis Test (own analysis)
    1. Average cost of Premium seats in Boing aircraft is less than Airbus aircraft.
       * Our null hypothesis for this is H0:Mu:(P.B – P.A)=0;
       * Our alternate hypothesis is H1:Mu:(P.B - P.A)<0; where P.B=average price in boeing, P.A=average price in Airbus
       * As our p-value after conducting the t test is 0.3309, which is more than 0.05 , so we can’t reject our null hypothesis .
    2. Average cost of Economy seats in Boing aircraft is less than Airbus aircraft
       * Our null hypothesis for this is H0:Mu:(P.B – P.A)=0;
       * Our alternate hypothesis is H1:Mu:(P.B - P.A)<0; where P.B=average price in boeing, P.A=average price in Airbus
       * As our p-value after conducting the t test is 0.2148, which is more than 0.05 , so we can’t reject our null hypothesis .
    3. Average cost of Economy seats in Domestic is less than International.
       * Our null hypothesis for this is H0:Mu:(P.D – P.I)=0;
       * Our alternate hypothesis is H1:Mu:(P.D - P.I)<0; where P.I=average price in International, P.D=average price in Domestic
       * As our p-value after conducting the t test is <2.2e-16, which is less than 0.05 , so we can reject our null hypothesis.
    4. Average cost of Premium seats in Domestic is less than International.
       * Our null hypothesis for this is H0:Mu:(P.D – P.I)=0;
       * Our alternate hypothesis is H1:Mu:(P.D - P.I)<0; where P.I=average price in International, P.D=average price in Domestic
       * As our p-value after conducting the t test is <2.2e-16, which is less than 0.05 , so we can reject our null hypothesis .
18. Multiple regression model 1 for PRICE\_ECONOMY
    1. The first model has PRICE\_ECONOMY as response variable or dependent variable.
    2. Predictor variables are PITCH\_ECONOMY, WIDTH\_ECONOMY,FLIGHT\_DURTION, QUALITY, PRICE\_RELATIVE, INTERNATINAL.
    3. The R^2 value is 0.6512 which is good model
    4. The model’s, p-value: < 2.2e-16 is also lower than the statistical significance level of 0.05, this indicates that we can safely reject the null hypothesis that the value for the coefficient is zero (or in other words, the predictor variable has no explanatory relationship with the response variable).
19. Multiple regression model 2 for PRICE\_PREMIUM
    1. The first model has PRICE\_PREMIUM as response variable or dependent variable.
    2. Predictor variables are PITCH\_PREMIUM, WIDTH\_EPREMIUM, FLIGHT\_DURTION, QUALITY, PRICE\_RELATIVE, INTERNATINAL.
    3. The R^2 value is 0.4732 which is moderately good model
    4. The model’s, p-value: < 2.2e-16 is also lower than the statistical significance level of 0.05, this indicates that we can safely reject the null hypothesis that the value for the coefficient is zero (or in other words, the predictor variable has no explanatory relationship with the response variable).
20. Multiple regression model 3 for PRICE\_RELATIVE
    1. The first model has PRICE\_RELATIVE as response variable or dependent variable.
    2. Predictor variables are PITCH\_PREMIUM, PITCH\_ECONOMY, WIDTH\_EPREMIUM, FLIGHT\_DURTION, QUALITY, INTERNATIONAL, PRICE\_PREMIUM, PRICE\_ECONOMY.
    3. The R^2 value is 0.7306 which is very good model
    4. The model’s, p-value: < 2.2e-16 is also lower than the statistical significance level of 0.05, this indicates that we can safely reject the null hypothesis that the value for the coefficient is zero (or in other words, the predictor variable has no explanatory relationship with the response variable).
21. .After analyzing the data through all parameter , it be inferred that the factors which explain the difference in price of Economy seats and Premium economy seat tickets are FLIGHT\_DURATION, INTERNATIONAL, QUALITY, WIDTH\_ECONOMY and PITCH\_PREMIUM.