**Report On Airline Pricing Project**

By

Name: Ankit Kashyap

[ankitkashyap.te15@rvce.edu.in](mailto:ankitkashyap.te15@rvce.edu.in)

Phone: 9582854497 College: R.V College Of Engineering

Introduction : The data set with which we are working over here is a classification data set. It is used to classify between economy prices and premium prices of tickets in a flight. The following data set has data available corresponding to different Airlines and several other parameters.

From my analysis of the data I have concluded the following inferences. These inferences are made by executing the R code snipped one after the other for easier understanding:

1. Air France is the most expensive airline among all the given six.
2. The most affordable and cheap among all the airlines is the Jet Airlines.
3. Jet Airlines offer the least number of seats in both PREMIUM ECONOMY and ECONOMY class.
4. Singapore Airlines offer the most number of ECONOMY class seats.
5. British Airlines offer the most number of PREMIUM ECONOMY class seats.
6. International flights are more than domestic flights.
7. Most number of flights are by British airlines.
8. The price of ECONOMY and PREMIUM ECONOMY of Airbus is more than Boeing.
9. Low range Economy tickets are more frequent.
10. Flights having less price and less duration is more for both ECONOMY as well as PREMIUM ECONOMY.
11. There is a positive correlation between Price Premium and the flight duration and Price Economy and the 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
    1. Average Cost of Premium seats in Boeing Aircraft is less than Airbus.
       1. Our null Hypothesis is h0:P.B-P.A=0
       2. Our Alternate hypothesis is h1:P.B-P.A<0, where P.B is average price in Boeing aircraft and P.A is average price in Airbus aircraft.
       3. As our p value after conducting 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 Boeing Aircraft is less than Airbus.
       1. Our null Hypothesis is h0:P.B-P.A=0
       2. Our Alternate hypothesis is h1:P.B-P.A<0, where P.B is average price in Boeing aircraft and P.A is average price in Airbus aircraft.
       3. As our p value after conducting t test is 0.2148 which is more than 0.05 so we can’t reject our null hypothesis.
    3. Average Cost of Premium seats in International is more than Domestic.
       1. Our null Hypothesis is h0:P.B-P.A=0
       2. Our Alternate hypothesis is h1:P.B-P.A<0, where P.B is average price in Boeing aircraft and P.A is average price in Airbus aircraft.
       3. As our p value after conducting t test is 2.2e-16 which is less than 0.05 so we can reject our null hypothesis.
    4. Average Cost of Economy seats in International is more than Domestic.
       1. Our null Hypothesis is h0:P.B-P.A=0
       2. Our Alternate hypothesis is h1:P.B-P.A<0, where P.B is average price in Boeing aircraft and P.A is average price in Airbus aircraft.
       3. As our p value after conducting t test is 2.2e-16which is more than 0.05 so we can reject our null hypothesis.
18. Multiple Regression Model1 for Price Economy:
    1. The first model has Price Economy as response variable
    2. Predictor variables are PITCH\_ECONOMY + WIDTH\_ECONOMY + FLIGHT\_DURATION + QUALITY + PRICE\_RELATIVE + INTERNATIONAL
    3. The R square value is 0.6512 which is a good model. 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 Model2 for Price Premium:
    1. The first model has Price Premium as response variable
    2. Predictor variables are PITCH\_PREMIUM + WIDTH\_PREMIUM + FLIGHT\_DURATION + QUALITY + PRICE\_RELATIVE + INTERNATIONAL
    3. The R square value is 0.4732 which is a moderately good model. 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 Model3 for Price Relative:
    1. The first model has Price Premium as response variable
    2. Predictor variables are PITCH\_PREMIUM + WIDTH\_PREMIUM + PITCH\_ECONOMY + QUALITY + INTERNATIONAL + FLIGHT\_DURATION+PRICE\_ECONOMY+ PRICE\_PREMIUM
    3. The R square value is 0.7306 which is a very good model. 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 and all the parameters it can be concluded that the factors which explain the difference in the Pricing system of Economy and Premium Economy seats tickets are:**
    1. **Flight Duration**
    2. **International**
    3. **Quality**
    4. **Width\_Economy**
    5. **Pitch\_Premium**