**FINAL REPORT**

We had to analyse a data set of six different airlines pertaining to their specifications regarding economy and premium classes which offer better comfort, entertainment, more legroom space etc.. Therefor an analysis was to be done to answer the question – What factors affect the difference in price between economy and premium class.

As we read the data into the R studio, we first analysed each variable in the data set to analyse the trends in various variables. The analysis of each variable is as follows –

* The summary statistics gave us a few details about the mean, median, minimum, maximum, first quartile and third quartile which help us to learn the values present in the data set.
* Boxplot for the Flight duration helps us to understand the range of distribution of the flight duration with respect to each airline. Singapore airlines had the longest median duration and jet airways had the least.
* From the monthly analysis, Virgin, delta, British and Airfrance all had the same monthly distribution and such an analysis is useful to denote which time of the year are people mostly wanting to travel so that hike in prices could take place which could be profitable for the company. Jet airways works in the range august to September.
* Bar charts help us to understand the variation between the domestic and international flights let by these six airlines and the difference is pretty significant.
* The various distribution of economy and premium seats tell us how many of such seats are available in each airline. In terms of economy seats, Airfrance has the maximum, whereas in premium seats are more available. Now with the seat availability there can be a rise in prices offered by the airline. As more number of premium seats will give them more profit than usual and such an analysis can be done.
* With the variation in the pitch and width of the seats we can compare their significance with the prices offered in the particular class. The pitch in economy is greatest in delta although they have three cases that is – 31,32 and the maximum 33 and pitch value of premium seats in greatest in jet airways with a value of 40.
* Width of premium seats are equal in three airlines, delta , jet and virgin although delta offers another size of 18. Width of economy seats is highest in Singapore airlines and equal in all other airlines. These can be the reason for higher price range in Singapore airlines.
* Airfrance provides with the highest median price of economy tickets and that could be because of the duration of flight time being the least. And other factors like economy pitch all come into action here, which is superlative to a few airlines. Jet airways shows the least distribution and this could be because of the least number of seats offered in both economy and premium class.
* Quality is highest in jet airways which again proves as a discrepancy from the prices being more for a good quality flight in both premium and economy class which therefore puts a lot of confusion in the analysis that we should also put human knowledge into the data set instead of blindly believing the data set.

Analysis of the relationships using scatterplots:

* From the first plot, we can see that correlation between the airline and flight duration is longest with british airways and boeing was the flight that was more largely distributed than airbus among the six airlines.
* From the plot it is evident that pitch value of 38 had the largest price values even more than those with 40 values. From the fourth plot, width of economy of 18 was largely distributed with a higher price range.
* Pitch of economy has higher prices for higher pitch value although the mediocre value of 31 has a range of thousand dollars more than the highest price i.e the outliers values.
* Width value of 18 in economy and 19 and 21 in premium gives us a discrepancy in the data set which doesn’t allow us to say that the larger the width value the larger the price, therefore a new hypothesis is that the ranges of 18,19 and 21 do give us the highly priced ones
* The last is a pairs test that I took considering factors that may affect the economy tickets price and that can be explained as follows. Since this contains discrete set of data not a very clear reading can be done, but pitch value of 30 had the largest number of distribution of seats.

Corrgram Ananlysis:

* Basically, a corrgram is a graphical representation of the cells of a matrix of correlations. The idea is to display the pattern of correlations in terms of their signs and magnitudes using visual thinning and correlation-based variable ordering. Moreover, the cells of the matrix can be shaded or colored to show the correlation value.
* The positive correlations are shown in blue, while the negative correlations are shown in red. The darker the hue, the greater the magnitude of the correlation.
* Quality and Pitch premium have the highest positive correlation whereas pitch economy and quality have the highest negative correlation. The pie charts go about in a clockwise direction to show the effects of the factors and their contribution levels.
* The covariance matrix provides with the value of variation.

T Tests:

* T tests were made with a few factors which may have been the factors for the difference in price in economy and premium class. The first T test with the number of seats in premium provides us a p value of less that 0.05 which shows that it is significant and the data is shown with variation with other factors like quality , pitch premium which all give significant relevance to our point of interest.

Regression models:

* From this model we are to find the factors that affect the price of the premium tickets and those include – the airlines and the price of economy tickets along with the other factors. Here, in the formula, the predicted value is the price of premium tickets and the factors of high significance like the airlines and the economy ticket prices can be the b0,b1 and so on values.
* Therefore, if we consider the general formula, then suppose an increase in particular airline number say, britishb airways, the the flight price would increase by 756 dollars from the statistics show. Like wise a number of observations can be made, therefore factors like these affect the price of the premium tickets and is justified by the p values made in the T tests.