‘Flight Price Prediction in Metro Cities of INDIA’

-An Exploratory Analysis and Regression

Contents:

- ## Executive Summary:

Optimal timing for airline ticket purchasing from the consumer's perspective is challenging principally because buyers have insufficient information for reasoning about future price movements. Anyone who has booked a flight ticket knows how unexpectedly the prices vary. The cheapest available ticket on a given flight gets more and less expensive over time. This usually happens as an attempt to maximize revenue based on -

1. Time of purchase patterns.
2. Keeping the flight as full as they want it (raising prices on a flight which is filling up in order to reduce sales and hold back inventory for those last-minute expensive purchases)

In this project we get to know about the relation between different elements effecting the pricing through visualization plotting. Using hypothesis testing will know the significance between pricing and other effecting elements. We simulate regression model for computing expected future prices and the accuracy behind it.

- ## INTRODUCTION:

Aviation connects people, cultures, and businesses across every continent. It generates economic growth, facilitates international trade, and promotes tourism. With more than 1,300 airlines, that operate 31,717 aircraft at 3,759 airports, it's hard to overstate the impact of the aviation industry. More importantly, aviation is a critical enabler for the broader economy. A safe, secure, efficient, regular and environmentally responsive aviation industry reinforces a range of trade and commerce, tourism, investment activities that contribute significantly to our economic prosperity.

➢ DATA COLLECTION:

'EaseMyTrip' is an internet platform for booking flight tickets, and hence a platform that potential passengers use to buy tickets. Data source was secondary data and was collected from 'EaseMyTrip' website. Data was collected for **38 days** starting from February 11th to March 19th ,2022. Dataset contain information about flight booking options for flight travel between India's top metro cities. There are **987 datapoints** and **10 features** in the cleaned dataset.

Let us first check on the various features of our dataset:

1. **Airlines-** A column with categorical data having 6 different airline companies' names.
2. **Source-city-** City from where the flight takes off. It's a categorical feature having 3 unique cities.
3. **Departure-time-** This is a derived categorical feature obtained, created by grouping time periods into bins. It stores information about the departure time having 6 unique time labels.
4. **Arrival-time-** This is a derived categorical feature obtained, created by grouping time periods into bins. It keeps information about the arrival time having 6 unique time labels.
5. **Destination-city-** City where the flight will land. It is a categorical feature having 4 unique cities.
6. **Class-** A categorical feature that contain information about seat class with 2 distinct values.
7. **Stops-** A numerical feature which shows the number of halts between destination & source city.
8. **Duration-** A numerical feature that display duration of travel time in hours.
9. **Days-left-** A numerical derived characteristic calculated by subtracting trip date by booking date.
10. **Price-**A numerical feature stores information about the ticket price.

- # METHODOLOGY:

- ## Data Cleaning:

The first step of any exploratory data analysis is to clean the dataset. So, we look for any missing value if present and either replace it or remove it (in case of unimportant data). Then we have concatenated 'Source_City', 'Destination_city' and named it 'Travelling_cities. Also have converted 'Duration' from hour to minute, named it 'Duration_mins.

Here this is how our cleaned dataset look like:

```
In [9]: df.head()
```

Out[9]:

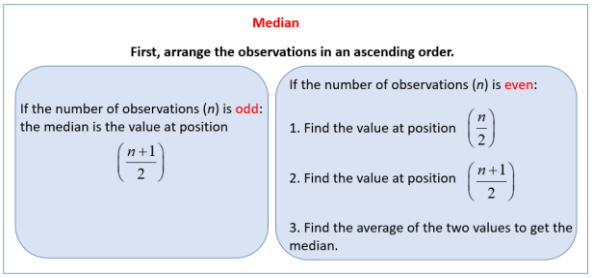
	Airlines	Source_City	Departure_time	Arrival_time	Destination_city	Class	Stops	Duration	Days_left	Price	Travelling_cities	Duration_mins
0	SpiceJet	Delhi	Evening	Night	Mumbai	0	0	2.17	1	5953	Delhi-Mumbai	137
1	SpiceJet	Delhi	Early_Morning	Morning	Mumbai	0	0	2.33	1	5953	Delhi-Mumbai	153
2	AirAsia	Delhi	Early_Morning	Early_Morning	Mumbai	0	0	2.17	1	5956	Delhi-Mumbai	137
3	Vistara	Delhi	Morning	Night	Mumbai	0	1	12.17	1	18923	Delhi-Mumbai	737
4	Vistara	Delhi	Evening	Morning	Mumbai	0	1	15.58	1	19238	Delhi-Mumbai	958

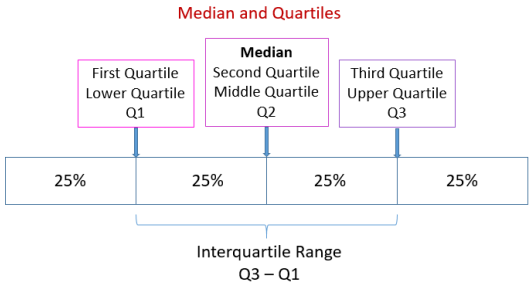
- ## Descriptive Statistics:

Basic descriptive measurement used for numeric features are-

Mean: The arithmetical mean is the average of the given numbers and is calculated by dividing the sum of given numbers by the total number of numbers dividing the sum of given numbers by the total number of numbers.

Median: In a sorted, ascending or descending, list of numbers, the median is the middle number and may be more representative of that data set than the average.





Standard Deviation: A calculation of the amount of variance or dispersion of a set of values is the standard deviation.

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$

σ = population standard deviation
N = the size of the population
x_i = each value from the population
μ = the population mean

- **Diagrammatic Representation:**

- **Bar Graph:** The pictorial representation of grouped data, in the form of vertical or horizontal rectangular bars, where the lengths of the bars are equivalent to the measure of data, are known as bar graphs or bar charts.

- **Line Chart:** A line chart is a type of chart used to show information that changes over time. Line charts are created by plotting a series of several points and connecting them with a straight line. Line charts are used to track changes over short and long periods.

- **Scatter Plot: Scatter plots** are the graphs that present the relationship between two variables in a data-set. It represents data points on a two-dimensional plane or on a **Cartesian system**. The independent variable or attribute is plotted on the X-axis, while the dependent variable is plotted on the Y-axis. These plots are often called **scatter graphs** or **scatter diagrams.**

- **Box Plot:** A box plot is a chart that shows data from a five-number summary including one of the measures of central tendency. It does not show the distribution in particular as much as a stem and leaf plot or histogram does. But it is primarily used to indicate a distribution is skewed or not and if there are potential unusual observations (also called outliers) present in the data set.
  Boxplots are also very beneficial when large numbers of data sets are involved.

- **Pie Chart:** A pie chart is a type of graph that records data in a circular manner that is further divided into sectors for representing the data of that particular part out of the whole part. Each of these sectors or slices represents the proportionate part of the whole. Pie charts, also commonly known as pie diagrams help in interpreting and representing the data more clearly. It is also used to compare the given data.

- **Distribution Plot:** The distribution plot is suitable for comparing range and distribution for groups of numerical data. The distribution plot visualizes the distribution of data. The distribution plot is not relevant for detailed analysis of the data as it deals with a summary of the data distribution.

These graphical representations have been used both for univariate and bivariate plotting to find out the relation between different features of our dataset.

- Covariance and correlation:

Covariance is basically a measure of how much two random variables change together. The **correlation coefficient** is the specific measure that quantifies the strength of the linear relationship between two variables in a correlation analysis

A **correlation matrix** is simply a table which displays the correlation coefficients for different variables. The matrix depicts the correlation between all the possible pairs of values in a table. It is a powerful tool to summarize a large dataset and to identify and visualize patterns in the given data.

We have calculated correlation coefficient between different numeric features and represented using 'heatmap' for better understand.

- **Inferential Statistics: -**

Hypothesis Testing-

- Chi Squared Test of Homogeneity: A chi-square test for homogeneity is a test to see if different distributions are similar to each other.

Steps:

1. Define your hypotheses

   H_o: The distributions are the same among all the given populations.

   H_a: The distributions differ among all the given populations.

2. Find the expected counts:

   1. For each cell, multiply the sum of the column it is in and the sum of the row it is in, and then divide by the total in all cells, or the sample size (row total) (column total) sample size.

Here, our testing hypothesis are as followed:

H0: DISTRIBUTION ARE SAME OF BUSINESS CLASS AND ECONOMY CLASS IN DIFFERENT AIRLINES

H1: DISTRIBUTION OF BUSINESS CLASS AND ECONOMY CLASS IN DIFFERENT AIRLINES ARE NOT SAME

- Kruskal-Wallis test:

A **Kruskal-Wallis test** is used to determine whether or not there is a statistically significant difference between the medians of three or more independent groups. This test is the nonparametric equivalent of the one-way ANOVA and is typically used when the normality assumption is violated. The Kruskal-Wallis test does not assume normality in the data and is much less sensitive to outliers than the one-way ANOVA.

Next our testing criteria was to test if there are statistically significance difference between **prices** due to '**stops**' but by 'Shapiro-wilk' test we get to know distributions doesn't follow normal distribution except flights with 2 stops, then we proceed to Kruskal-Wallis as distributions are unknown, though after removing outliers.

Regression Analysis:

Regression analysis is a statistical method used for the elimination of a relationship between a dependent variable and an independent variable. It is useful in accessing the strength of the relationship between variables.
It also helps in modelling the future relationship between the variables.
Regression analysis consists of various types including linear, non-linear, and multiple linear. But the most useful ones are the simple linear and multiple linear.

Multiple Linear Regression:

Linear regression is a statistical method that tries to show a relationship between variables. It looks at different data points and plots a trend line.
The **multiple linear regression** method tries to find the relationship between two or more independent variables and the corresponding dependent variable.

It helps us to know:

- How strong the relationship is between two or more independent variables and one dependent variable (e.g., how rainfall, temperature, and amount of fertilizer added affect crop growth).
- The value of the dependent variable at a certain value of the independent variables (e.g., the expected yield of a crop at certain levels of rainfall, temperature, and fertilizer addition).

Multiple linear regression formula:

The formula for a multiple linear regression is:

$$y = \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n + \epsilon$$

- y = the predicted value of the dependent variable
- B_0 = the y-intercept (value of y when all other parameters are set to 0)
- $B_1 X_1$ = the regression coefficient (B_1) of the first independent variable (X_1) (a.k.a. the effect that increasing the value of the independent variable has on the predicted y value)
- … = do the same for however many independent variables you are testing
- $B_n X_n$ = the regression coefficient of the last independent variable
- ϵ = model error (a.k.a. how much variation there is in our estimate of y)

The MSE, MAE, and RMSE metrics are mainly used to evaluate the prediction error rates and model performance in regression analysis.

- **MAE** (Mean absolute error) represents the difference between the original and predicted values extracted by averaged the absolute difference over the data set.
- **MSE** (Mean Squared Error) represents the difference between the original and predicted values extracted by squared the average difference over the data set.
- **RMSE** (Root Mean Squared Error) is the error rate by the square root of MSE.

R^2 – Coefficient of Determination and Adjusted R Square

In statistics the coefficient of determination denoted by R^2 or r^2 and pronounced "R squared", is the proportion of the variation in the dependent variable that is predicted from the independent variable(s).

Definition:-
A data set has n values marked y_1, y_2,...y_n(denoted as y_i) each associated with a fitted(or modelled, or predicted) value f_1,f_2,...f_n(denoted as f_i or sometimes $\hat{y}_i$)

Define the residuals as $e_i = y_i - f_i$

If $\bar{y}$ is the mean of the observed data:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Then the variability of the dataset can be measured with two sum of squares formulas:

The total sum of squares (proportional to variance of the data):

$$SS_{tot} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

The sum of squares of residuals, also called residual sum of squares:

$$SS_{res} = \sum_{i=1}^{n} (y_i - f_i)^2 = \sum_{i=1}^{n} e_i^2$$

Then the most general definition of the coefficient of determination is

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

R^2 is a statistic that will give some information about the goodness of fit of a model. In regression it measures how well the regression predictions approximate the real data points. The value of R^2 lies between 0 and 1. More the value of R^2 close to 1 indicates a better fit of model. R^2 of 1 indicates that the regression predictions perfectly fit the data.

Adjusted R^2

However each time we add a new predictor variable to the model the R- squared is guarantee to increase even if the predictor variable isn't useful.

The adjusted R-squared or adjusted R^2 is a modified version of R^2 that adjusts for the number of predictors in a regression model. It is calculated as

$$1 - \frac{(1 - R^2)(n - 1)}{n - k - 1}$$

R^2 : The coefficient of determination of the model
n : The number of observations
k : The number of predicted variables

Since R^2 always increases as you add more predictors to a model, adjusted R^2 can serve as a metric that tells you how useful a model is, adjusted for the number of predictors in a model. Therefore the adjusted R^2 tells us the percentage of variation explained by only the independent variables that actually affect the dependent variable. Same as R^2, the value of adjusted R^2 lies between 0 and 1. More the value of adjusted R^2 close to 1 indicates that all the predictor variables in model have better significant effects.

Analysis and Results:

First, we intend to study the summary statistics for the numeric features of our dataset that is 'Duration', 'Days_left', 'Price', 'Duration_mins' i.e., duration in minutes. We will calculate their total number of counts that is frequency, mean, median (50%), standard deviation(std) , quantiles (25%, 75%), maximum, minimum values-

]:

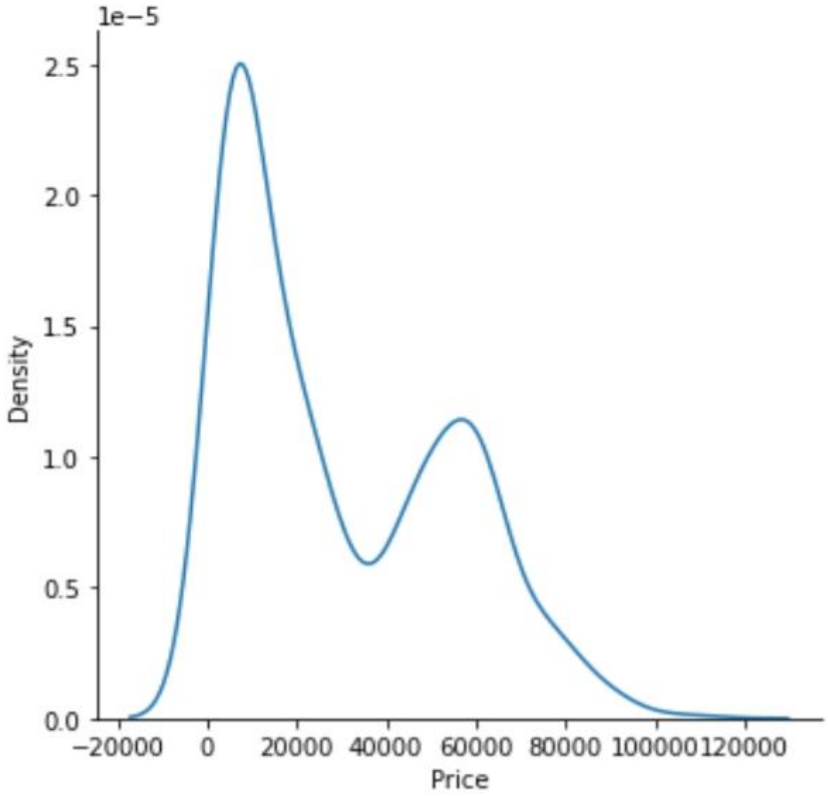
	count	mean	std	min	25%	50%	75%	max
Duration	986.0	12.099909	8.033007	0.83	5.67	10.75	17.42	39.83
Days_left	986.0	16.523327	10.571678	1.00	6.25	16.00	26.00	38.00
Price	986.0	29726.105477	24786.978142	1603.00	6488.00	20771.50	52063.00	110872.00
Duration_mins	986.0	738.910751	484.246550	60.00	360.00	667.00	1062.00	2423.00

Therefore, we can see mean and median values are quite different from each other and mean is **greater than** median, which is a sign of the distributions to be **Positively Skewed.**

Let's look at their distributions as well as skewness and kurtosis:

- *PRICE DISTRIBUTION PLOT-*

```
skew         0.633717
kurtosis    -0.846412
Name: Price, dtype: float64
```

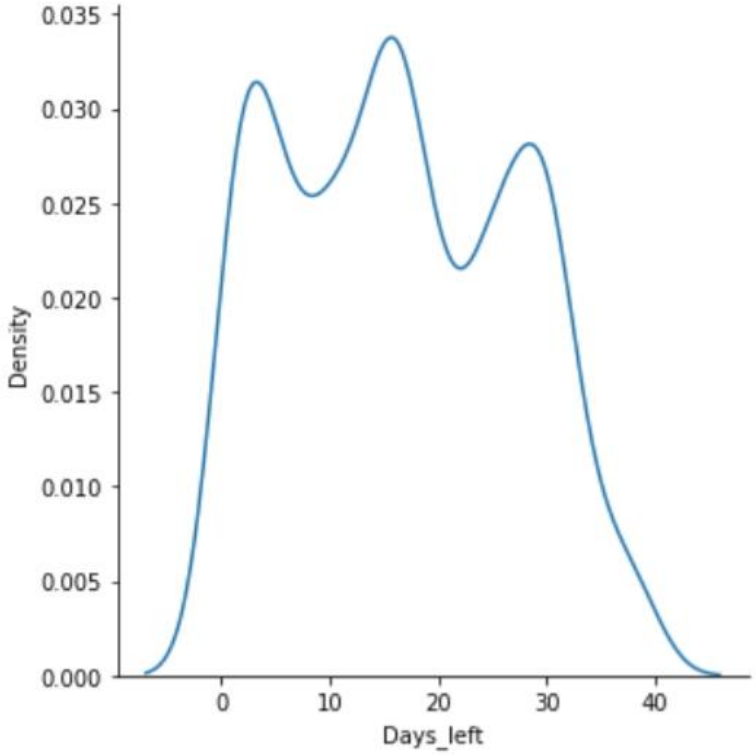


Hence, we can see from the plot Price distribution is positively skewed and its' value is 0.633717

Kurtosis value is -0.846412 as it is negative, we can say it is platykurtic.

- ***Distribution plot for days left to the flight:***

```
skew         0.158882
kurtosis    -1.098389
Name: Days_left, dtype: float64
```

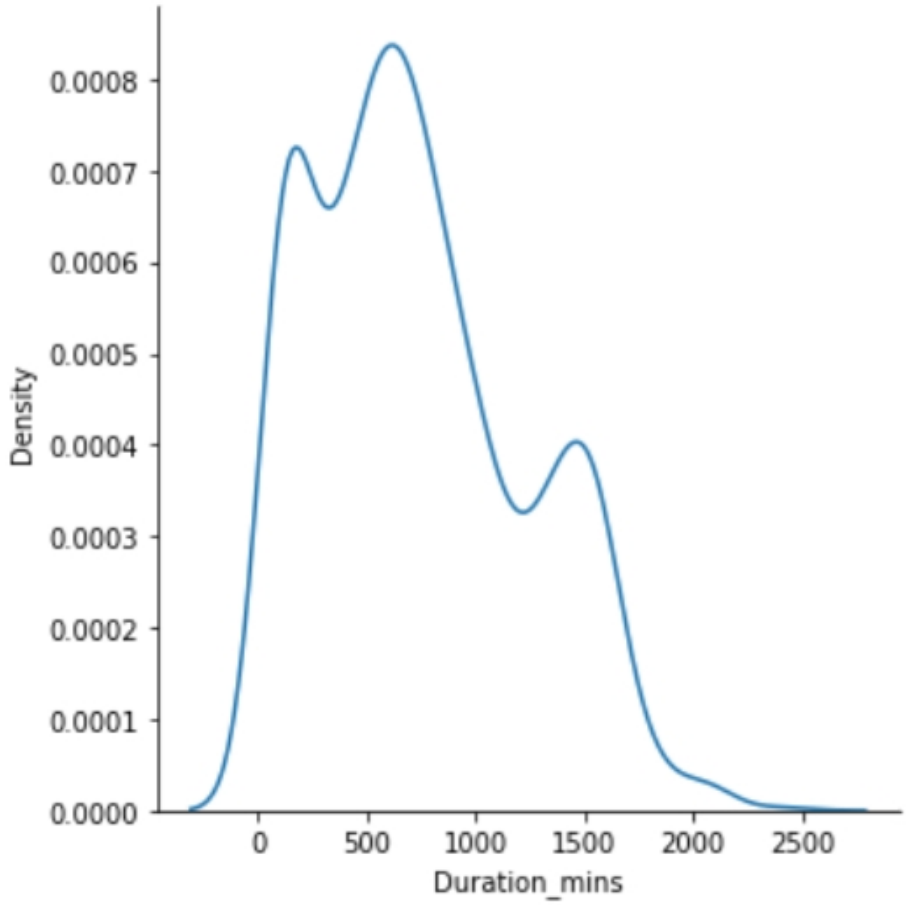


Hence, we can see from the plot distribution for the days left is positively skewed and its' value is 0.158882

Kurtosis value is -1.098389 as it is negative, we can say it is platykurtic.

- ***Distribution plot for Duration in Minutes:***

```
skew         0.547930
kurtosis    -0.518501
Name: Duration_mins, dtype: float64
```



Hence, we can see from the plot distribution for the days left is positively skewed and its' value is 0.547930

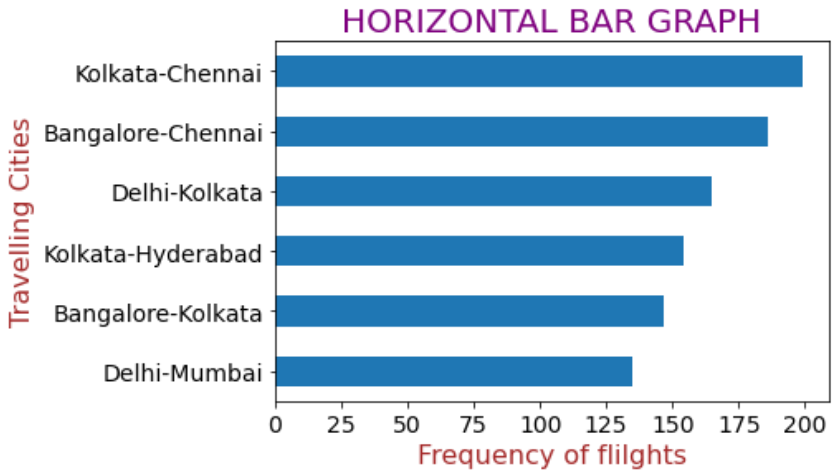
Kurtosis value is -0.518501 as it is negative, we can say it is platykurtic.

Frequency tables and univariate plotting:

The summary techniques we have used so far are suitable only for numeric variables. Categorical variables have values which are typically unordered. Therefore, we need a method to summarize categorical variables based on counts. Frequency tabulation is one way of getting a better idea of the distribution of categorical variables.

Frequency of flights between cities:

Kolkata-Chennai	199
Bangalore-Chennai	186
Delhi-Kolkata	165
Kolkata-Hyderabad	154
Bangalore-Kolkata	147
Delhi-Mumbai	135

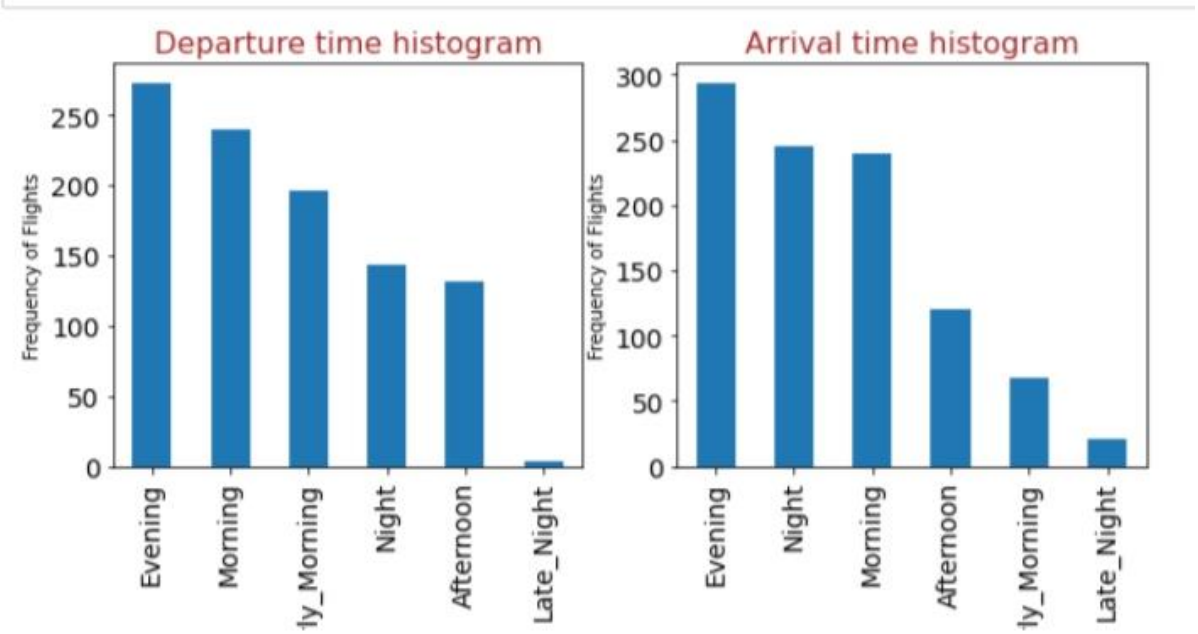


Therefore,

1. **Maximum number of flights are found in between 'Kolkata' and 'Chennai' and the number is 199**
2. **Least number of flights are found in between 'Delhi- Mumbai' and the number is 135**

- ***Frequency of flights according departure and arrival of flights:***

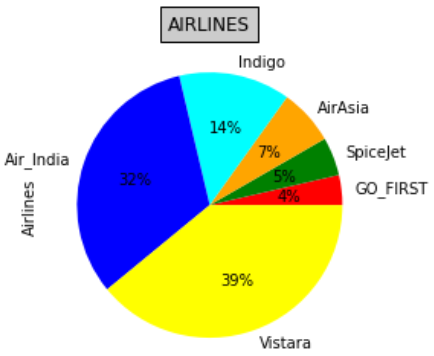
Evening	273	Evening	294
Morning	239	Night	245
Early_Morning	196	Morning	239
Night	143	Afternoon	120
Afternoon	131	Early_Morning	67
Late_Night	4	Late_Night	21
Name: Departure_time,		Name: Arrival_time,	



According to the *bar chart* we can conclude maximum number of flights arrives and departs in the evening. Very few are noticeable at late night.

Frequency of flights according airlines:

Vistara	385
Air_India	318
Indigo	134
AirAsia	67
SpiceJet	46
GO_FIRST	36
Name: Airlines, d	



1. **Here we can see 'VISTARA' airline flies in maximum number. VISTARA is a joint venture between Tata Sons and Singapore Airlines. The name VISTARA is symbolic of its vision of infinite expansion and the desire of the airline to become the Best Airline in INDIA to the people. Headquarters-GURGAON**

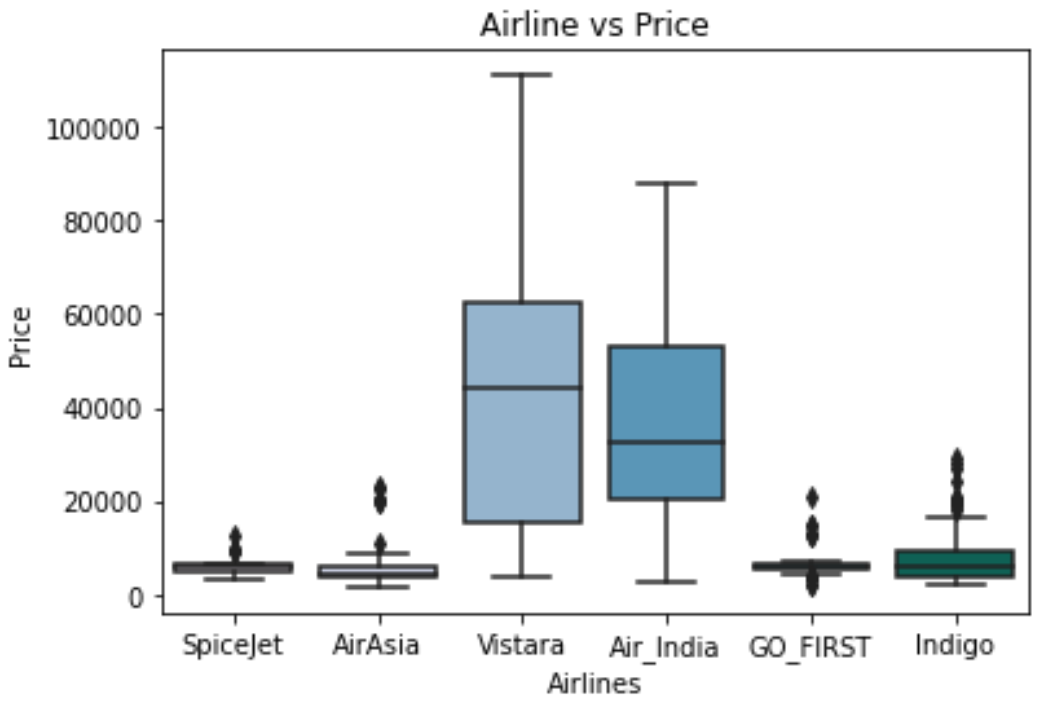
2. **'GO_FIRST' flies the least number. GO_FIRST is owned by conglomerate Wadia Group. Go_First is primarily a budget airline, but it has introduced its 'frequent flyer program' GO CLUB which offers premium services such as access to airport lounges.**

3. **Here as we can see maximum flights are from Air_India and Vistara compare to other Airlines that is 32% and 39% respectively maybe one of the reasons behind this is No other Airlines have business class except this two.**

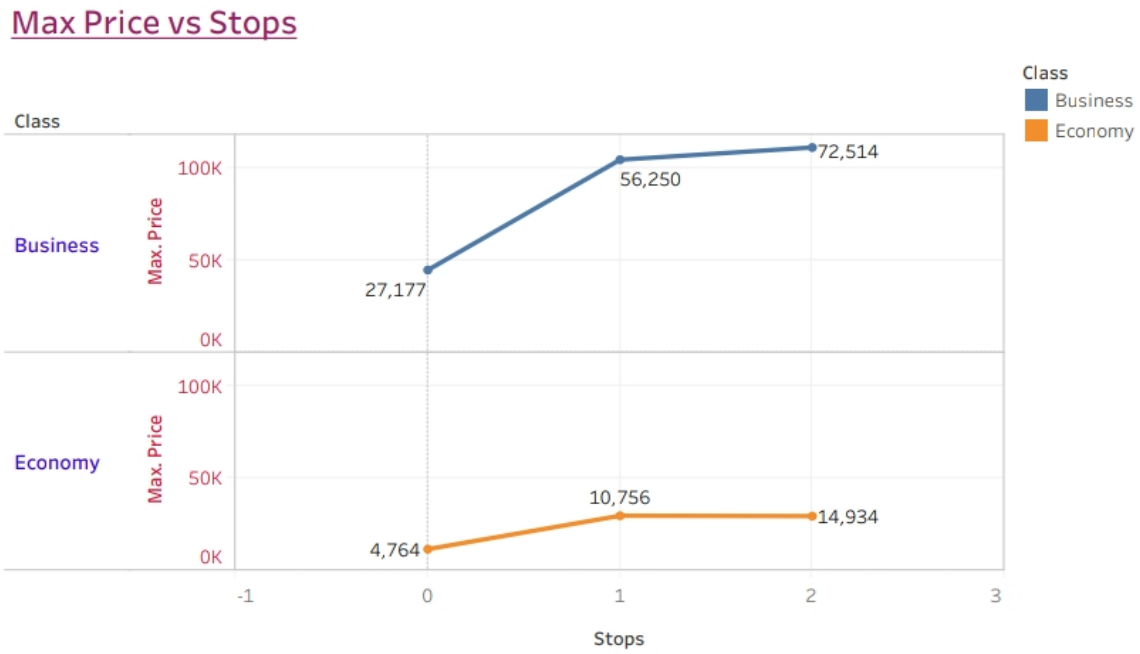
Bivariate plotting:

- ## Average Price vs Airlines:

```
Airlines
AirAsia        6016.925373
Air_India     35682.399371
GO_FIRST       7079.027778
Indigo         7617.820896
SpiceJet       5853.173913
Vistara       41597.205195
Name: Price, dtype: float64
```

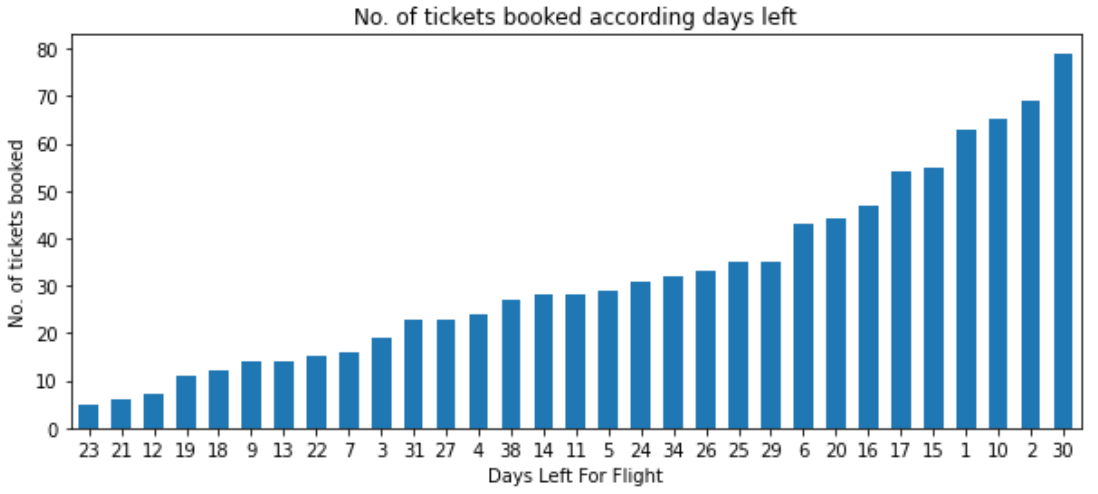


Ticket prices for vistara is real high compare to others Whereas for spice jet is cheapest



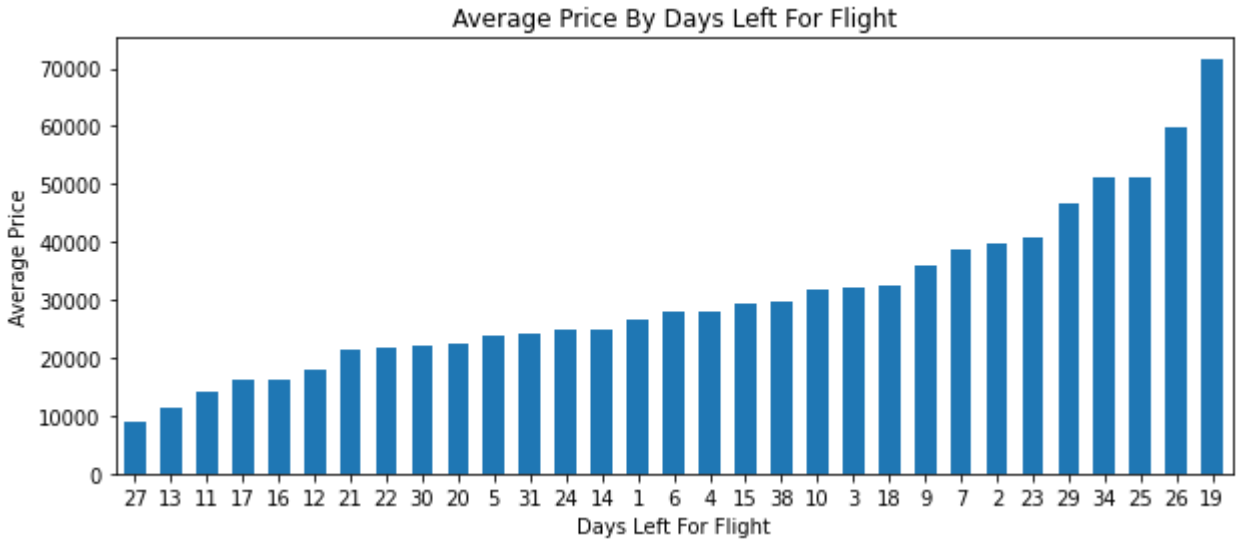
Ticket price of Flights increase according increase in number of stops both for Economy and Business class

Number of tickets booked vs no. of days left for the flight:

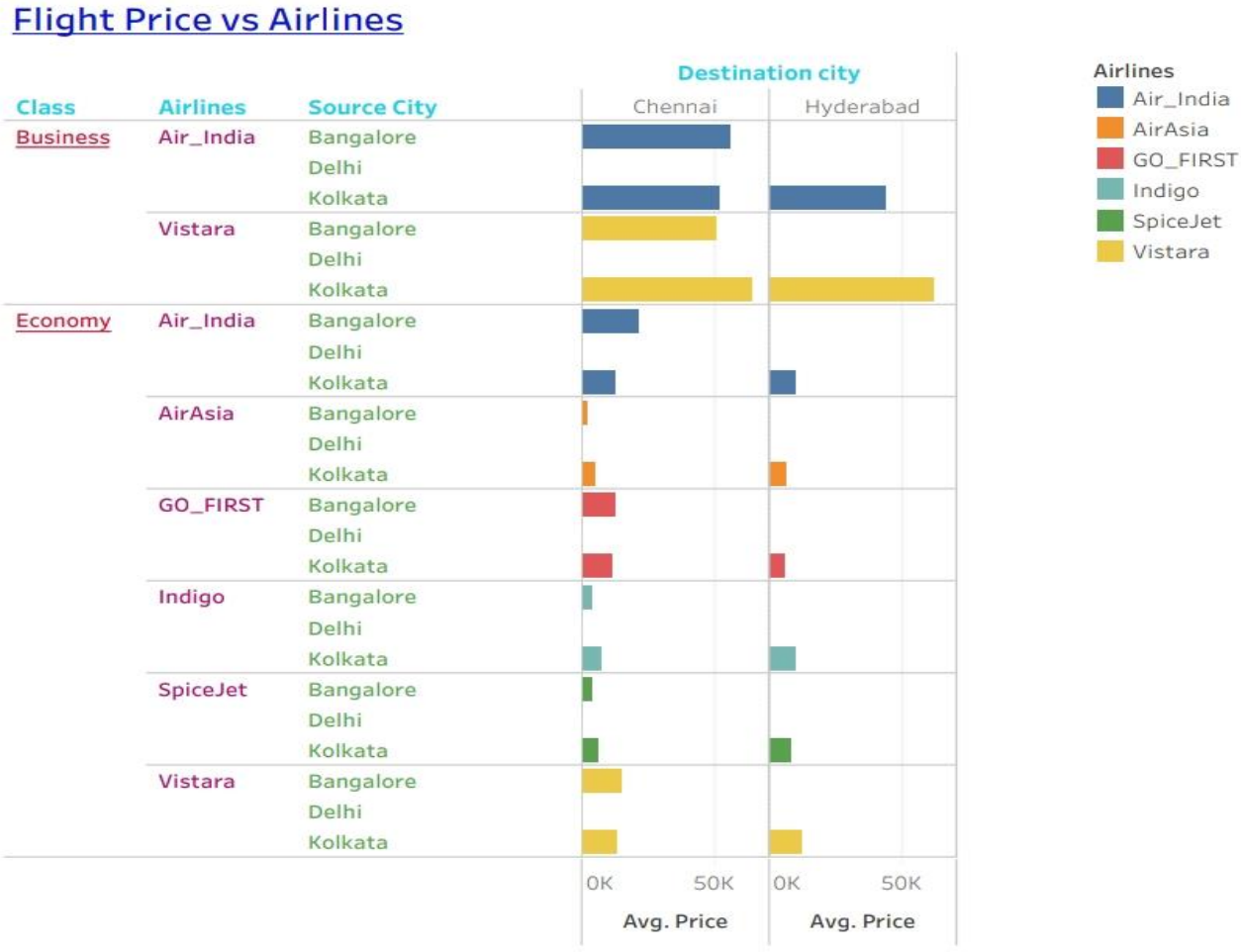


Maximum tickets have been booked 30 days before the flight.

Average price vs Days left for the flight:



Flight prices are almost same in the range 20-24 days before the flight, and high 19 days before.

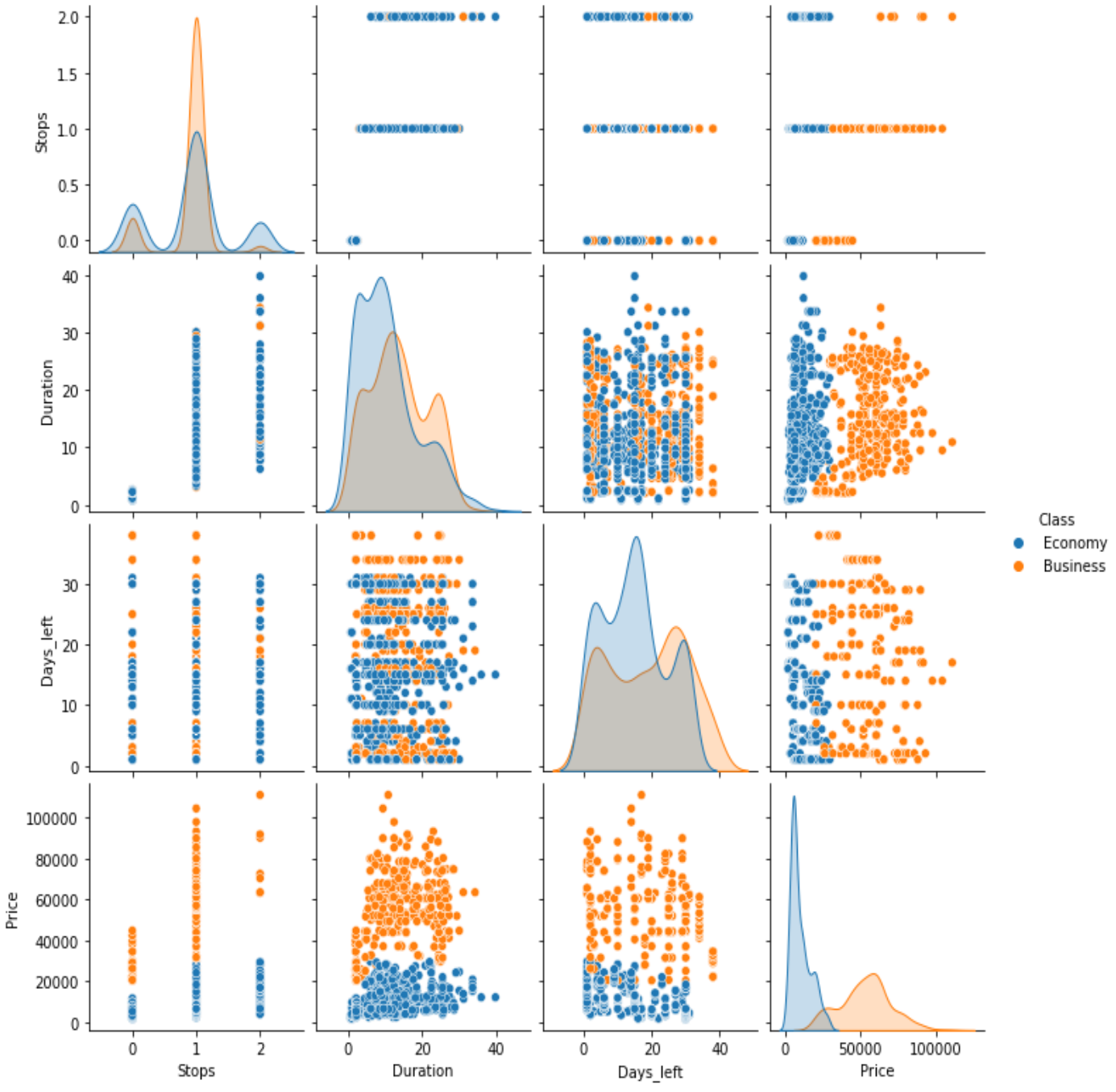


Correlation Matrix:

The correlation matrix is a bit easier to understand, since the values are normalized. The diagonal values of the correlation matrix are all 1, as a result of the normalization. The magnitudes of the correlation values between the variables range between -1 and +1. A magnitude closer to +1 or -1 indicates a high correlation.

Some pairs of variables are weakly correlated, with small magnitudes. For example, ***Price*** *and* ***Days_left*** *are weakly correlated, whereas* ***Stops*** *and* ***Duration*** *are strongly correlated.*

Pairplot:



The pairplot () function of seaborn library of python language helps in creating an axes grid through which each numeric variable present in data is shared across y-axes in the form of rows and across x-axes in form of a column. Scatter plots are created to show pairwise relationships and, in the diagonal, the distribution plot is created to show the distribution of the data in the column.

Analysis of Hypothesis Testing:

- *Now we proceed to few testing problems for better under standing of our data:*

H0: DISTRIBUTION ARE SAME OF BUSINESS CLASS AND ECONOMY CLASS IN DIFFERENT AIRLINES
H1: DISTRIBUTION OF BUSINESS CLASS AND ECONOMY CLASS IN DIFFERENT AIRLINES ARE NOT SAME

To proceed with the test, we see our crosstab to be:

Class	Business	Economy
Airlines		
AirAsia	0	67
Air_India	221	97
GO_FIRST	0	36
Indigo	0	134
SpiceJet	0	46
Vistara	232	153

As, we can see value counts of AirAsia, GO_FIRST, Indigo, SpiceJet are 0 or null for business class therefore these values are not necessary for our testing so we removing this given rows. Hence, we get our 2*2 contingency table as follows:

Class	Business	Economy
Airlines		
Air_India	221	97
Vistara	232	153

Now, proceeding with **chi-square test of homogeneity** at 5% level of significance:

```
Test statistic: 6.122272005344314,
 P value:0.013348912236178838,
 Degree of freedom:1
```

Therefore, as our p value <0.05 we reject null hypothesis H0.

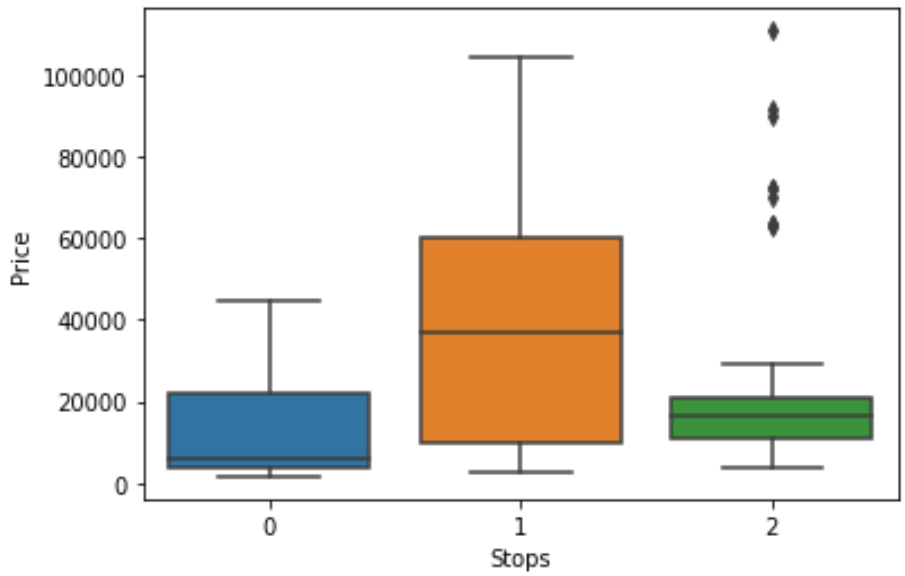
Observed value: `([[204.91322902, 113.08677098],`
`[248.08677098, 136.91322902]])`

Our next hypothesis is to find significance between stops and pricing:

H0: There has no significance between pricing of flights and number of stops.

H1: There has significance between pricing of flights and number of stops

To test this at first, we have divided the dataset in three sub part according their number of stops between the journey and got to find there has some outliers in the data with 2 stops in between source city and destination city.



Though, after removing the outliers we are unable to find any known distributi on of the three datasets. That's why proceeding with non-parametric test.

Here, for this test Kruskal-Wallis test is appropriate as it is comparison between three data type which is more than two an extended version of Mann-Whitney test.

We are proceeding with test at 5% level of significance.

After testing we get the p value **`7.445339051955684e-40 which is less than 0.05.`**

So, we reject Null Hypothesis I.e., there has significance difference in pricing between flights with 1 stop, 2 stops and non-stop.

Regression Analysis:

As we have seen our dataset have 6 categorical variables, we need to convert it to either ordinal or nominal data first to proceed with any kind of regression.

Here, after encoding categorical features for example (**Airlines_Air_India, Airlines_Indigo, Departure_time_Evening etc)** will be either 0 or 1.

Now after converting all to numeric nature, we will check 'Mutual info regressor' which estimates mutual information for fixed categories like in a classification problem or a continuous target variable in regression problems. Mutual Information works on the entropy of the variables. We will have better idea about applying multiple linear regression.

Hence, our Mutual info regressor are:

Days left	1.638352
Duration	0.663298
Class	0.657075
Stops	0.387963
Source_City_Kolkata	0.361054
Airlines_Vistara	0.354626
Airlines_Air_India	0.335507
Source_City_Delhi	0.329128
Destination_city_Kolkata	0.324633
Destination_city_Hyderabad	0.241178
Destination_city_Mumbai	0.229284
Airlines_Indigo	0.205412
Airlines_SpiceJet	0.105498
Departure_time_Morning	0.097862
Arrival_time_Early_Morning	0.088506
Airlines_GO_FIRST	0.081912
Departure_time_Night	0.081407
Departure_time_Evening	0.074861
Arrival_time_Night	0.065387
Arrival_time_Morning	0.060487
Arrival_time_Evening	0.060131
Arrival_time_Late_Night	0.038306
Departure_time_Early_Morning	0.038073
Departure_time_Late_Night	0.000600

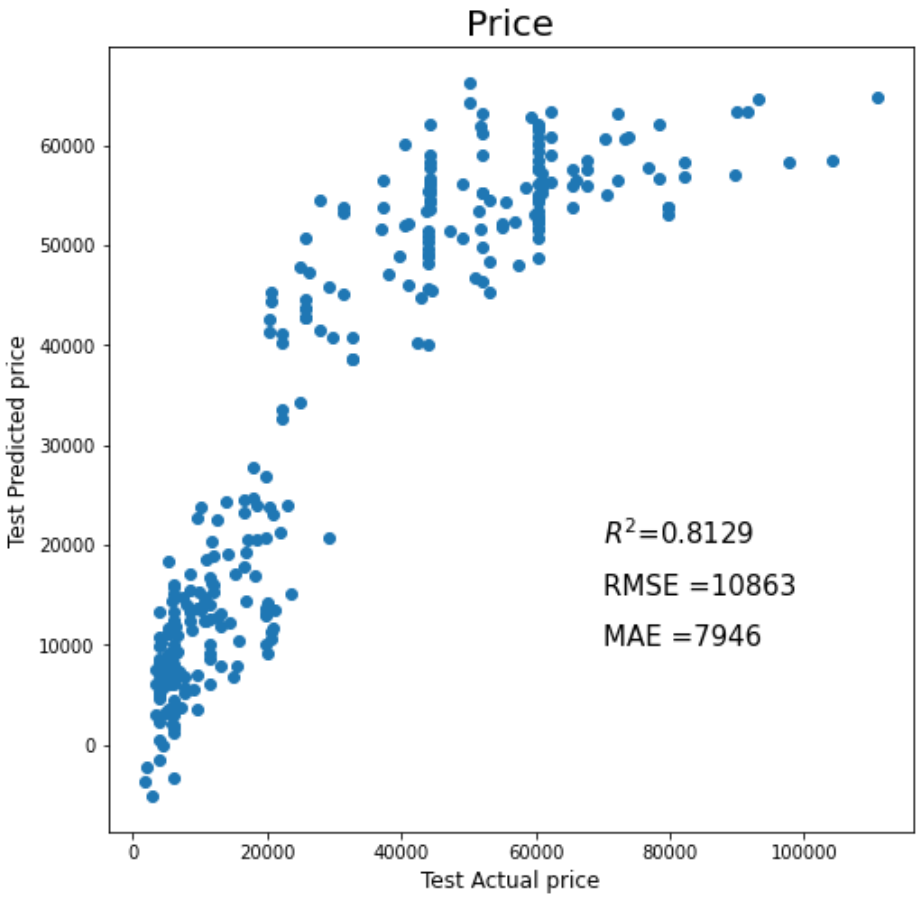
We know higher the regressor value better the effect on prediction. So, number of days left, stops, duration, class effects the pricing of flights more. Now, proceeding with Multiple linear regression:

Scatter plots are used to observe relationships between variables and uses dots to represent the relationship between them. Here points are nearly aligned in a line.

Evaluating the model accuracy is an essential part of the process of creating machine learning models to describe how well the model is performing in its predictions. The MSE and RMSE metrics are mainly used to evaluate the prediction error rates and model performance in regression analysis.

For our dataset those values are:

```
              r2 score:  0.812924533365962
    mean squared error:  118002255.48213412
root mean squared error: 10862.884307684313
```



****HENCE, OUR PREDICTION IS 81.29% ACCURATE****

- ***CONCLUSION:***

- *Maximum number of flights are flies in between 'Kolkata' and 'Chennai' and the number is 199*
- *Least number of flights are flies in between 'Delhi- Mumbai' and the number is 135*
- *Maximum flights lands in the evening, least are noticeable at late night.*
- *Here as we can see maximum flights are from Air_India and Vistara compare to other Airlines that is 32% and 39% respectively maybe one of the reasons behind this is No other Airlines have business class except this two.*
- *Flight prices are way higher in Business class and in between Vistara , Air India- the two airlines with business class, ticket prices in Air India is cheaper.*
- *Flight price increases according increasing in number of stops in between travelling cities.*
- *It is better to buy tickets 20-23 days ago or just 2-3 days ago as prices are comparatively cheaper then.*
- *For Business class ticket prices are highest while travelling from Kolkata to Chennai or Hyderabad.*
- *For Economy class ticket prices are highest while travelling from Bangalore to Chennai.*
- *From testing we can say distribution of Business and economy class is different in various airlines.*
- *Using Multiple Linear Regression, we can predict pricing with 81.29% accuracy.*

- ***SOFTWARE AND PROGRAMMING LANGUAGE USED:***

  - *Python*
  - *R Studio*
  - *Microsoft Excel*
  - *Microsoft Word*
  - *Tableau*

Appendix:

The dataset used for the project is with dimension (987,10)
The first 10 rows of the dataset are as followed:

Out[2]:

	Airlines	Source_City	Departure_time	Arrival_time	Destination_city	Class	Stops	Duration	Days_left	Price
0	SpiceJet	Delhi	Evening	Night	Mumbai	Economy	0	2.17	1	5953
1	SpiceJet	Delhi	Early_Morning	Morning	Mumbai	Economy	0	2.33	1	5953
2	AirAsia	Delhi	Early_Morning	Early_Morning	Mumbai	Economy	0	2.17	1	5956
3	Vistara	Delhi	Morning	Night	Mumbai	Economy	1	12.17	1	18923
4	Vistara	Delhi	Evening	Morning	Mumbai	Economy	1	15.58	1	19238
5	Vistara	Delhi	Evening	Morning	Mumbai	Economy	1	15.75	1	19920
6	Air_India	Delhi	Afternoon	Night	Mumbai	Economy	1	7.92	1	19710
7	Air_India	Delhi	Early_Morning	Night	Mumbai	Economy	1	13.25	1	19710
8	Vistara	Delhi	Early_Morning	Night	Mumbai	Economy	1	16.00	1	20130
9	Air_India	Delhi	Evening	Evening	Mumbai	Economy	1	22.75	1	19815
10	Vistara	Delhi	Evening	Night	Mumbai	Economy	1	6.00	1	21075

❖ **Source Code:**

https://drive.google.com/file/d/1oQKolG4CS74I7Att48o2Dx2TqGjxcHuS/view?usp=drivesdk

Flight Price - INDIA(Domestic) | Tableau Public