Objective-

I love flying and always look forward to it when I plan trips. After the Covid-19 pandemic, the aviation industry started recovering from the extremely low-ticket prices we saw during the peak of the pandemic. However, with the war in Ukraine and the increase in ATF (Aviation Turbine Fuel) prices, ticket prices have soared to new highs. Because of this, I decided to analyze ticket prices in India to understand the factors affecting them better. I want to find out things like how many flights are available across India, the availability of tickets in different classes, and the price range for each class.

Problem Statement-

In this project, we study the data which is in tabular format using various Python libraries like Pandas, NumPy, Matplotlib and Seaborn.

We study different columns of the table and try to co-relate them with others and find a relation between those two.

We try to find and analyze those key factors like class of travel, duration of fight, etc. which helps us understand the pricing of tickets to plan and schedule our air travel in efficient way.

Dataset Information-

We got the information from Kaggle, which is considered secondary data. This data is about the options for booking tickets on the "Easemytrip" website for flights between the six biggest cities in India. After cleaning, the dataset has 11 different pieces of information and includes 300,261 data points. We collected the data in two parts: once for business class tickets and once for economy class tickets. The website offered a total of 300,261 different flight booking choices. We collected data for 50 days, from February 11 to March 31, 2022.

The Link to the Dataset is [here](https://www.kaggle.com/datasets/shubhambathwal/flight-price-prediction).

The various features of the cleaned dataset are explained below:

1. Airline: The airline column contains the name of the airline firm. There are six different airlines, making it a category trait.
2. Flight: The flight code of the aircraft is stored in flight.
3. Source City: City where the flight departs from is a classification feature with 6 distinctive cities.
4. Departure Time: This is a categorical feature that was deduced from time periods being divided into bins. It has six different time labels and stores information about the departure time.
5. Stops: A category feature that holds the number of stops between the source and destination cities and has 3 different values.
6. Arrival Time: This derived categorical feature was developed by binning time intervals. It maintains information regarding the arrival time and has six different time labels.
7. Destination City: The location of the aircraft's landing. It is a classification feature with 6 distinctive cities.
8. Class: A permanent feature that shows the total number of hours needed to travel between cities.
9. Duration: A permanent feature that shows the total number of hours needed to travel between cities.
10. Days Left: The trip date is subtracted from the booking date to arrive at this derived feature.
11. Price: Information about the ticket price is stored in the target variable.