

## CAPSTONE PROJECT - 3

### FLIGHT DATA ANALYSIS :

#### Dataset Description Report

The dataset contains flight booking options from the website Easemytrip for travel between India's top 6 metro cities. It includes 300,261 datapoints and 11 features. The features include:

**Date\_of\_Journey:** The date of the journey.

**Source:** The source city of the flight.

**Destination:** The destination city of the flight.

**Route:** The route taken by the flight.

**Dep\_Airline: Time:** The departure time of the flight.

**Arrival\_Time:** The arrival time of the flight.

**Duration:** The duration of the flight.

**Total\_Stops:** The number of stops the flight makes.

**Additional\_Info:** Additional information about the flight.

**Price:** The price of the flight ticket.

**Modules needed:** Numpy, pandas, matplotlib, seaborn

You need to do the following analysis actions:

1. Load the file
2. Print first 5 rows of data
3. Print last 5 rows of data
4. Cleaning the data for missing values, null values etc..

5. Remove IN-DATA `index` column
6. Get some info about the data
7. Get some description about the data

## **Visualization**

1. What are the airlines in the dataset, accompanied by their frequencies?
2. Departure time against Arrival time using barplot.
3. Source city against Destination city.
4. Does price vary with Airlines?
5. Does ticket price change based on the departure time and arrival time using line plot?
6. How the price changes with change in Source and Destination?
7. Duration of travel vs city
8. Show the high price with class type for city.

This dataset can be used to analyze various factors affecting flight prices, such as the airline, departure and arrival times, source and destination cities, and the duration of the flight. The visualizations help in understanding the distribution and relationships between these factors and the ticket prices.