

## Description of Airfare Data

This dataset contains flight fare data that was collected from the EaseMyTrip website using web scraping techniques. The data was collected with the goal of providing users with information that could help them make informed decisions about when and where to purchase flight tickets. By analyzing patterns in flight fares over time, users can identify the best times to book tickets and potentially save money.

### Sources:

1. Data collected using Python script with BeautifulSoup and Selenium libraries.
2. Script collected data on various flight details such as Date of booking, Date of travel, Airline and class, Departure time and source, Arrival time and destination, Duration, Total stops, Price.
3. The scraping process was designed to collect data for flights departing from a specific set of airports (Top 7 busiest airports in India).  
Note that the Departure Time feature also includes the Source airport, and the Arrival Time feature also includes the Destination airport. Which is later extracted in Cleaned\_dataset. Also both cleaned and scraped datasets have provided so that one can use dataset as per their requirement and convenience.

### Inspiration:

1. Dataset created to provide users with valuable resource for analyzing flight fares in India.
2. Detailed information on flight fares over time can be used to develop more accurate pricing models and inform users about best times to book tickets.
3. Data can also be used to study trends and patterns in the travel industry through air can act as a valuable resource for researchers and analysts.

### Limitations:

1. This dataset only covers flights departing from specific airports and limited to a certain time period.
2. To perform time series analysis one have gather data for at least top 10 busiest airports for 365 days.
3. This does not cover variations in aviation fuel prices as this is the one of influencing factor for deciding fare, hence the same dataset might not be useful for next year, but I will try to update it twice in an year.
4. Also demand and supply for the particular flight seat is not available in the dataset as this data is not publicly available on any flight booking web site.

### Scope of Improvement:

1. The dataset could be enhanced by including additional features such as current aviation fuel prices and the distance between the source and destination in terms of longitude and latitude.
2. The data could also be expanded to include more airlines and more airports, providing a more comprehensive view of the flight market.
3. Additionally, it may be helpful to include data on flight cancellations, delays, and other factors that can impact the price and availability of flights.

4. Finally, while the current dataset provides information on flight prices, it does not include information on the quality of the flight experience, such as legroom, in-flight amenities, and customer reviews. Including this type of data could provide a more complete picture of the flight market and help travelers make more informed decisions.