

# Airline Flight Fare Analysis

## 1. Project Overview

### Motivation

To provide actionable insights that empower users to make informed decisions related to flight fare strategy. This dashboard also supports broader business goals, including “Optimizing pricing strategies” and “Staying competitive in the airline industry”.

### Objectives

The goals of this project include:

- Analyzing the flight fares data to uncover patterns, trends, and relationships with other variables, such as airline, route, etc.
- Identifying key factors that contribute to fare fluctuations and understanding their impact in determining price changes.
- Creating actionable insights and recommendations based on the analysis to help users in making informed decisions.
- Developing visualizations and reports to effectively communicate the insights to users in a user-friendly manner.

### Scope

The scope of this project include:

- Exploratory data analysis of the flight fares data to uncover insights and trends.
- Data preprocessing to extract relevant information for analysis.
- Applying statistical techniques and machine learning algorithms to identify correlations and patterns.
- Curating actionable insights related to fares fluctuations, booking trends, route analysis, customer segmentation.
- Develop intuitive visualizations and report to present insights in a user-friendly manner.

### Assumptions

This project has following assumptions:

- The available features in the data provide sufficient information to draw meaningful insights.
- The relationship between fare fluctuations and the identified factors remains consistent over time.
- The selected statistical and machine learning techniques can effectively uncover insights from the data.

### Constraints

The limitations of the project include:

- Factor such as seat availability, luggage fees, or airline-specific policies are not considered in the analysis. Additionally, the dataset does not include demand and

supply information for specific flight seats, as this data is not publicly available on the flight booking websites.

- External factors beyond the flight fares data, such as economic indicators, seasonal trends, and aviation fuel prices are not incorporated into the analysis

## 2. Data Set Description

### Data Source:

The dataset consists of flight fares information that can be downloaded from [Kaggle](#).

### Data Collection Method:

The data collection process involved utilizing a Python script that made use of BeautifulSoup and Selenium libraries. The script was designed to extract various flight details from the EaseMy Trip website, an Indian online travel company that provides a range of travel-related services, including flight bookings. These details include the date of booking, date of travel, airline and class, departure time and source, arrival time and destination, duration, total stops, and the price of the flight, which was denominated in Indian Rupee (INR).

### Data Contents:

The dataset comprises flight booking information from India's seven busiest airports. It covers flights scheduled between January 16, 2023 and March 6, 2023. The dataset includes various fields such as date of booking, date of travel, airline and class, departure time and source, arrival time and destination, duration, total stops, and the price of flight.

### Data Relevance

The flight fares data is highly relevant as it provides insights into patterns, trends, and relationships between flight fares and other variables. By studying this data, we can gain a better understanding of the factors influencing fare fluctuations.

## 3. Data Profile

The cleaned dataset, after removing 6,722 duplicates, contains 445,366 rows and 17 columns.

Column	Qualitative /Quantitative	Discrete /Continuous	Nominal /Ordinal
Date of Booking	Quantitative	Continuous	
Booking Day	Qualitative		Ordinal
Date of Journey	Quantitative	Continuous	
Journey Day	Qualitative		Ordinal
Days_left	Quantitative	Discrete	
Airline	Qualitative		Nominal
Flight Code	Qualitative		Nominal
Flight Class	Qualitative		Ordinal
Departure_Time	Qualitative		Nominal
Departure Location	Qualitative		Nominal

<b>Departure_TimeCategory</b>	Qualitative		Ordinal
<b>Arrival_Time</b>	Qualitative		Nominal
<b>Arrival_Location</b>	Qualitative		Nominal
<b>Arrival_TimeCategory</b>	Qualitative		Ordinal
<b>Duration (in min)</b>	Quantitative	Continuous	
<b>Total Stops</b>	Quantitative	Discrete	
<b>Price</b>	Quantitative	Continuous	

	<b>Days_left</b>	<b>Duration (in min)</b>	<b>Total Stops</b>	<b>Price</b>
<b>count</b>	445366	445366	445366	445366
<b>mean</b>	25.62	730.95	0.94	22919.75
<b>std</b>	14.30	437.96	0.42	20394.21
<b>min</b>	1	45	0	1307
<b>max</b>	50	2615	2	143019

### Potential bias

One potential bias in the dataset is sample bias. Since the data is limited to flights booked from India's seven busiest airports, covering flights scheduled between January 16, 2023 and March 6, 2023 with bookings made on January 25, 2023, it may not be fully representative of the entire population and it might not capture seasonal variations in fare movements that occur outside the specified time period.

### Limitations

The limitations of the project include:

- The historical dataset used for analysis is limited to flights departing from the top 7 busiest airports in India and covers only flights scheduled between January 16, 2023 and March 6, 2023, with all booking dates set on 15 January 2023.
- The current dataset focuses solely on flight prices and does not incorporate information about the flight experience quality, such as legroom, in-flight amenities, and customer reviews.

### Ethics

The data contains no Personally Identifiable Information (PII).

## 4. List of Questions

- What are the average flight fares from each of the seven busiest airports in India during the specified time period? How routes and airlines affect the fares?
- How do flight fares vary based on airline and the class of service?
- Are there specific days of the week when the flight fares tend to be lower or higher?
- Are there specific dayparts (times of the day) when the flight fares tend to be lower or higher?
- What are the most and least expensive flight routes based on the departure and destination cities?
- How does the number of stops in flight impact the ticket price?

- Is there a correlation between the flight duration and the ticket price?
- Are there any significant fare fluctuations or price trends during specified time period?
- Are there any patterns in flight fares for specific airlines or routes?