Capstone Project 3

Flight Price Prediction Analysis Report

Introduction

This report presents an analysis of a dataset containing flight booking options from the website Easemytrip for travel between India's top 6 metro cities. The dataset includes 300,261 data points and 11 features. The analysis aims to uncover insights about flight prices, airlines, travel durations, and other relevant factors.

Data Loading and Initial Inspection

The dataset was loaded using pandas, and the first and last five rows were inspected to understand its structure. The dataset was then cleaned to handle missing values and remove any unnecessary columns.

Data Cleaning

Missing Values: Rows with missing values were dropped to ensure data integrity.

Index Column: The index column was removed as it was not needed for the analysis.

Data Overview

Info: The dataset contains 11 features, including Airline, Source, Destination, Price, Dep_Time, Arrival_Time, Duration, and Class.

Description: Statistical summaries of the data were generated to understand the distribution of numerical features.

Visualizations and Insights

Airlines Frequencies:

The dataset includes multiple airlines, with their frequencies visualized using a count plot.

Insight: The most frequent airlines in the dataset can be identified, which helps in understanding market dominance.

Departure Time vs Arrival Time:

A bar plot was used to compare departure and arrival times.

Insight: This visualization helps in identifying peak travel times and potential delays.

Source City vs Destination City:

The relationship between source and destination cities was visualized using a count plot.

Insight: Popular travel routes can be identified, which is useful for route optimization and marketing strategies.

Price Variation with Airlines:

A box plot showed how prices vary across different airlines.

Insight: This helps in understanding which airlines offer more expensive or cheaper flights, aiding in budget planning.

Price vs Departure and Arrival Time:

Line plots illustrated how ticket prices change based on departure and arrival times.

Insight: This can help travelers choose the best times to book flights for cost savings.

Price Variation with Source and Destination:

A box plot showed how prices change with different source and destination cities.

Insight: Identifying expensive and cheap routes can help in planning travel budgets.

Duration of Travel vs Source City:A box plot visualized the duration of travel against source cities.

Insight: This helps in understanding the average travel time from different cities, which is useful for scheduling.

High Price with Class Type for City:

A box plot showed the high prices with class types for different cities.

Insight: This helps in understanding the premium pricing for different classes and cities, aiding in luxury travel planning.

Conclusion

This analysis provides valuable insights into flight prices, travel durations, and airline frequencies. The visualizations help in understanding the factors affecting flight prices and can be useful for making informed decisions regarding flight bookings. By leveraging these insights, travelers and businesses can optimize their travel plans and strategies.