

REPORT

Flight Price EDA



PROBLEM STATEMENT

Flight ticket prices are renowned for their volatility, influenced by a multitude of factors including flight timing, destination, and travel duration. However, in the era of machine learning algorithms, we can harness historical flight data to create a predictive model that assists consumers in making informed decisions about when to purchase tickets.

The lack of transparency regarding future price movements in the airline industry often leaves travelers uncertain about the optimal time to buy tickets. This project seeks to unravel the underlying patterns in flight prices within India by analyzing historical data and recommending the best timing for ticket purchases. Through a comparative analysis of different machine learning models, this project aims to either confirm or debunk common myths surrounding the airline industry. Additionally, it intends to quantify the potential savings achievable by purchasing tickets at the most opportune moment.

Interestingly, flight prices are highly responsive to a myriad of factors, including the specific route, the month and day of departure, the time of the flight, and the airline company. For instance, on competitive routes like Mumbai to Delhi, ticket prices tend to rise as the departure date approaches, while on other routes like Delhi to Guwahati, there's a specific timeframe when ticket prices hit their lowest point. Furthermore, the dataset reveals two categories of airline carriers operating in India - the economical group and the luxurious group, with cheaper flights typically falling within the former category. The data also confirms that certain times of the day are associated with higher ticket prices. This project has the potential to expand its scope to encompass various routes, enabling travelers to make substantial savings when purchasing flight tickets in the Indian domestic airline market. By utilizing machine learning to navigate the complex web of pricing dynamics, this initiative

empowers consumers to make more informed decisions and optimize their travel expenses.

DATASET

The primary aim of this research is to scrutinize a comprehensive flight booking dataset acquired from Ease My Trip, a prominent online platform facilitating flight ticket reservations. Ease My Trip serves as a popular medium through which prospective travelers purchase airline tickets. The dataset at hand encompasses a wealth of information regarding the various flight booking choices accessible via the Ease My Trip website, specifically catering to air travel between the six major metropolitan cities in India.

This meticulously curated dataset comprises an impressive 300,261 data points, each of which is accompanied by 11 distinct features. These 11 features encapsulate a diverse array of information related to flight bookings, ranging from passenger preferences and choices to flight details and pricing information.

Furthermore, this dataset provides a unique window into the dynamics of the aviation industry, shedding light on patterns, pricing strategies, and demand fluctuations.

FEATURES

The dataset is composed of 11 distinct features, each offering valuable insights into the world of flight bookings and air travel. Here's a detailed explanation of these features:

- 1) Airline: This categorical feature records the name of the airline company associated with each flight option. With six unique airlines, it provides information on the carriers available for booking, which is crucial for passengers choosing their preferred airline.
- 2) Flight: Another categorical feature, "Flight," contains the flight code, offering a unique identifier for each flight. This information helps distinguish between different flights and is vital for tracking specific travel options.

-
- 3) Source City: This categorical feature identifies the city from which the flight departs. With six distinct cities, it offers insights into the departure locations and allows passengers to select flights originating from their desired city.
- 4) Departure Time: Derived from the actual departure time, this categorical feature categorizes time periods into bins, resulting in six unique time labels. It aids in understanding flight schedules and the convenience of departure times.
- 5) Stops: A categorical feature with three possible values, "Stops" reveals the number of stopovers between the source and destination cities. This is essential information for travelers looking for non-stop or connecting flights.
- 6) Arrival Time: Similar to "Departure Time," this derived categorical feature groups time intervals into six distinct labels, providing insights into arrival schedules at the destination city.
- 7) Destination City: This categorical feature identifies the city where the flight lands, offering a choice of six unique cities for passengers planning their trips.
- 8) Class: A categorical feature with two values, "business" and "economy," indicating the seat class. It helps passengers select their preferred class, which often corresponds to different levels of comfort and pricing.
- 9) Duration: A continuous feature that quantifies the total travel time in hours. This is vital for passengers who want to know the duration of their journey when selecting a flight.
- 10) Days Left: This derived feature calculates the time gap between the booking date and the trip date. It helps passengers assess the advance booking period, which can impact ticket availability and pricing.
- 11) Price: As the target variable, "Price" stores the ticket price. This is the key variable of interest, as it reflects the cost associated with each flight option, enabling passengers to make informed decisions based on their budget and preferences.

Data Collection and Preparation

1. **Data Sources** To gather historical flight data, we'll source information from various reliable sources, ensuring that we collect data related to flight bookings via Ease My Trip. These sources could include databases, APIs, or historical records provided by the platform.
2. **Data Quality Assurance** Given that our dataset contains 300,261 data points and 11 features, we must meticulously examine the data for quality and consistency. This involves performing data validation checks, ensuring that all records are complete, and resolving any potential data anomalies.
3. **Handling Missing Values** In the event of missing data, we'll apply appropriate strategies, such as imputation, to maintain data integrity. It's essential to ensure that we have complete information for all relevant features.
4. **Outlier Management** Detecting and addressing outliers is critical to maintaining the accuracy of our analysis. We'll employ statistical techniques to identify and, if necessary, handle outliers that may affect our results.
5. **Feature Engineering** Creating new features, as mentioned in the dataset description, is a crucial step. For instance, we'll calculate the "Days Left" by subtracting the trip date from the booking date. We'll also categorize "departure time" and "arrival time" to provide better insights into the relationships between these time periods and ticket prices.

ANALYSIS

In the analysis phase of our project, we utilized visualization techniques and libraries such as Matplotlib and Seaborn to answer a series of crucial business questions. These questions are designed to extract valuable insights from the dataset and inform various aspects of the airline industry, including ticket pricing, passenger preferences, and strategies for enhancing the passenger experience. Here's a comprehensive elaboration of the questions and the insights derived from our visualizations:

1. Ratio Between Business and Economy

Visualized the distribution of business and economy class passengers. The results revealed the ratio between these two classes, providing an understanding of the class preferences of travelers.

2. Most Profitable Class

By plotting the relationship between class (business and economy) and ticket prices, identify the more profitable class, which is vital for airlines to maximize revenue.

3. Ratio Between Different Airlines

Visualizing the distribution of passengers across various airlines helped to discern the market share and popularity of each airline among travelers.

4. Ratio Between Different Arrival and Departure Times

Examining the ratio of passengers for different arrival and departure times helps identify peak travel periods and offers insights for airlines to schedule more flights during high-demand times.

5. Ratio Between Source and Destination Cities

Visualizing the distribution of passengers based on source and destination cities provides insights into the popularity of different routes, which can influence pricing and route planning.

6. Ratio of Number of Stops

- Visualizing the number of stops and their distribution helps understand passenger preferences for non-stop or connecting flights.

7. Business-class passengers with Two Stops

identified how many business class passengers choose flights with two stops, allowing airlines to cater to this specific group.

8. Most Popular Destination for Business Class

Analyzing the destination cities for business-class passengers revealed which cities are most popular among this passenger group.

9. Airlines with Late-Night Departure

Visualizing the departure times for different airlines highlighted which carriers often operate late-night flights, which is valuable information for passengers seeking specific schedules.

10. Destination City for Late-Night Departures

determined the destination cities that late-night departure flights are most commonly associated with, assisting passengers in planning their itineraries.

11. Correlation Analysis

By constructing a correlation matrix and visualizing the relationships between variables, you can identify which factors have a strong influence on each other. This aids in understanding the interconnected dynamics of the airline industry.

12. Relationship Between Duration and Price

Established the relationship between the duration of flights and ticket prices, enabling airlines to consider pricing strategies based on travel time.

13. Relationship Between Price and Days Left

Visualizing the connection between ticket prices and the number of days left before departure allows for the determination of how pricing changes over time, which is valuable for passengers planning in advance.

14. Offers Based on Duration

Through visualization, it is suggested to offer incentives or promotions for flights with specific durations to attract passengers with varying travel preferences.

15. Offers Based on Arrival Time and Destination Cities

Visualizing data on arrival times and destination cities can help airlines design promotions and offers that encourage passengers to choose specific flight times and destinations.

The analysis phase used visualizations to answer key business questions, providing actionable insights for airlines to optimize their pricing strategies, improve passenger experiences, and make data-driven decisions. These findings can lead to more informed choices for both passengers and service providers in the Indian domestic airline market.

Insights and Conclusion

In conclusion, the analysis of the dataset has revealed several key findings and actionable insights that can significantly impact the strategies and decisions made within the Indian domestic airline market.

1. Class Preferences and Profitability

The majority of passengers tend to book economy class tickets. However, the analysis demonstrated that the business class is more profitable for airlines, highlighting the importance of optimizing this segment.

2. Vistara and Air India Focus on Business Class

It is found that Vistara and Air India primarily target the business class segment, offering relatively expensive tickets. Despite their higher prices, these airlines attract more business-class passengers and, consequently, enjoy higher profits.

3. Targeting Early-morning and Late-Night Travelers

Passengers with early morning and late-night arrival times emerged as distinct groups of interest. It is recommended to focus on these travelers and offer them incentives or promotions to enhance their travel experiences.

4. Encouraging Travel to Different Destination Cities

To diversify travel patterns and promote less-traveled destinations, we propose offering incentives and promotions to encourage passengers to explore alternative cities. This approach can stimulate demand and open up new opportunities for airlines.

5. VIP Passengers Prefer Air India

The data revealed that Air India caters to VIP passengers with exceptional services. Most of these VIP travelers opt for business class, indicating a strong alignment between premium services and class preference.

6. Relationship Between Stops and Duration

There is a significant positive relationship between the number of stops and the flight duration. To cater to this segment, we recommend creating special offers or packages that address the needs of travelers seeking longer flights with more stops.

7. Duration and Price Relationship

Contrary to expectations, it was found that there isn't a strong relationship between the duration of the flight and ticket prices. This suggests that airlines have flexibility in pricing flights of varying durations.

8. Days Left and Price Dynamics

Our analysis highlighted a key relationship between the days left before departure and ticket prices. As the days left decrease, prices tend to increase. We advise passengers to book their tickets well in advance, ideally more than 10 days before departure, to secure better pricing.

Recommendations

- For passengers, we suggest the following:
- Book tickets at least 10 days before departure for better pricing.
- If the flight duration is over 10 hours, look out for duration-based offers.
- Traveling to Chennai or Hyderabad? Consider destination-based offers.
- If you're heading to other destination cities with early morning or late-night departures, explore arrival-based offers.
- By focusing on specific traveler segments, offering tailored incentives, and promoting less-traveled destinations, airlines can better serve their customers and improve their competitive position in the Indian domestic airline market. These insights can benefit both passengers and airline providers by facilitating more informed travel choices and effective service strategies.