A logo for a university

AI-generated content may be incorrect.

* **Project Title**: Flight Price Prediction & Insights
* **Name:** Pooja Reddy Sodum
* **UID:** U01090593
* **Email:** sodum.2@wright.edu
* **Instructor:** Bin Li

## Executive Summary

This project uses data-driven methods to deliver an in-depth analysis of airline ticket prices with the goals of maximizing pricing strategies, increasing revenue, and offering practical booking advice. Because airfare pricing is dynamic and complex, our analysis uses advanced reinforcement learning, demand elasticity assessment, anomaly detection, predictive modeling, and exploratory data analysis to identify the main factors influencing flight prices to develop an intelligent pricing engine.

The first step in our journey was a thorough exploratory examination of flight ticket data, which showed that the cost of tickets varies greatly depending on the airline, departure time, class of service, and number of stops. Budget airlines like AirAsia and Indigo offer more consistent and reasonably priced fares than premium carriers like Vistara and Air India. The most important determinant of price was flight class, with business class tickets typically costing 8 times as much as economy. Surprisingly, one-stop flights were the most expensive, while non-stop flights were the least expensive.

We also examined how flight duration and booking dates affected the cost. Our results supported the traditional demand behavior, which states that late reservations lead to much higher costs, especially for economy class visitors who are more cost conscious. Naturally, long-haul flights cost more, and journeys longer than 8 hours have seen a notable price hike. We outperformed conventional linear models in our prediction of flight prices with great accuracy (R2 = 0.8791) using machine learning methods, namely Random Forest regression. Class, duration, and days to departure were the most important factors in determining pricing. About 4.8% of ticket prices were found to be anomalies by an anomaly detection model that used Isolation Forest. These anomalies mostly happened on premium routes and metro connections, indicating a mix of surge-based tactics, pricing errors, or a surge in demand.

Important information on consumer behavior was revealed by the price elasticity of demand (PED) analysis. Due to corporate travel budgets, business class passengers were found to have a high elasticity, or sensitivity, to price fluctuations. On the other hand, inelastic demand was displayed by economic passengers, indicating price-agnostic or necessity-driven behavior. Better segmentation for dynamic pricing strategies is made possible by these insights.   
In order to suggest the best time to book, we used a Monte Carlo simulation to predict the cost of the Delhi–Mumbai route. With a 60–62% chance of price increases over the next 10 days, the model recommended booking right away at ₹5000, emphasizing the significance of timing in airfare strategy.

In summary, this initiative not only pinpoints the variables influencing airfare but also offers a clever, data-supported framework for dynamic pricing, tactical demand division, and the best possible booking suggestions. These results can help travel agencies and airlines increase revenue, boost customer satisfaction, and maintain their competitiveness in a market that is changing.

## Table of Contents

1. Title Page
2. Executive Summary
3. Table of Contents
4. The Problem
5. The Data
6. The Methodology
7. Findings
8. Recommendations
9. Discussion
10. References

## The Problem

The cost of airline tickets varies depending on a number of factors, including competition, class, flight duration, and booking time. In order to enhance income and booking efficiency, this project will analyze these aspects, forecast ticket prices, and suggest optimal pricing options.

## The Data

**Source:** Kaggle ([Flight Price Prediction](https://www.kaggle.com/datasets/shubhambathwal/flight-price-prediction?resource=download))

**Size:** 300,521 records

**Key Variables:**

• Departure/arrival time and city

• Airline and flight class

• Number of stops

• Flight duration

• Ticket price

• Booking window (days before departure)

**Descriptive Analysis Highlights:**

• Vistara and Air India have the highest average prices.

• Economy tickets are ~8x cheaper than Business.

• Direct flights are cheapest; 1-stop flights are costliest.

• Prices increase as departure date approaches.

## Methodology

**Data Cleaning:**

• Standardized time periods into categories.

• Created route and booking window variables.

• Handled missing values and outliers.

**Model Development:**

**Exploratory Data Analysis**

To determine how different factors affect flight fares, we looked at departure time, airline, travel class, number of stops, and route in our exploratory data study. Interesting patterns were found in the distribution of flight costs across various departure times. Evening flights were likewise rather pricey (₹21,232.36), but night flights had the highest average price (₹23,062.15) with considerable variety. The average cost of flights in the afternoon was ₹18,179.20, making them a more economical choice. Although late-night flights were the least expensive (₹9,295.30), there were only 1,306 observations of them, indicating that this finding should be regarded with caution. Morning flights were the most popular, as evidenced by their moderate pricing (₹21,002.62) and high sample size.

As premium airlines, Vistara and Air India had the highest average tickets, according to an analysis of average costs per airline. Vistara led in both volume and premium pricing, reinforcing its strong position in the premium market category. In contrast, budget airlines such as AirAsia, Indigo, GO\_FIRST, and SpiceJet provided much cheaper prices. Business class rates were almost 8.4 times more expensive than economy fares, according to a price comparison by travel class. With more than 296,000 entries, economy class continued to be the most popular option among passengers, which is consistent with its reputation for being reasonably priced.

When flight fares were examined according to the number of stops, an unexpected result surfaced. On average, one-stop flights were the priciest, which may have been caused by intricate routing, longer flight times, or last-minute reservations. In contrast, multiple-stop flights had intermediate price, indicating a compromise between cost and convenience, while non-stop flights were not only the most popular but also the least expensive. Flights comprising large metropolises like Chennai, Bangalore, and Kolkata topped the list of the 10 most costly itineraries, according to the survey. As important centers of commerce and technology, these cities frequently see strong demand and last-minute travel, which raises ticket costs.

Overall, the investigation found that departure time, airline brand, travel class, number of stops, and specific routes all have a significant impact on flight pricing variances. Low-cost airlines provide more reasonably priced options with more constrained price distributions, while premium airlines like Vistara and Air India constantly maintain higher costs. One-stop flights surprisingly tend to cost more than non-stop flights, and business class tickets are significantly more expensive than economy class. Additionally, because of the increased demand for business travel, routes between major urban centers are more expensive. Both customer choices and airline pricing models can benefit from these insights, which are essential for comprehending pricing tactics in the airline sector.

A graph of a chart

AI-generated content may be incorrect.

A graph showing different colored rectangular shapes

AI-generated content may be incorrect.

A blue and yellow striped chart

AI-generated content may be incorrect.

A graph of different colored lines

AI-generated content may be incorrect.

**Regression Models (Linear & Random Forest) to predict flight prices.**

A graph with numbers and symbols

AI-generated content may be incorrect.

A graph showing a blue line

AI-generated content may be incorrect.

A thorough examination of a flight price prediction model constructed with Random Forest and Regression approaches is presented in this research. Using the Random Forest model, the study first looks at feature importance. Class turned out to be the most significant predictor, which reflects the huge price difference between Business and Economy prices. In line with established trends that prices increase closer to the departure date, flight duration was the second most important factor, while the number of days to departure had a small effect on prices. The lack of direct influence from other factors, such as source city, destination city, airline, and stops, suggests that their effects are captured indirectly by more dominating variables.

Most of the dots in the scatter plot comparing actual and anticipated prices were grouped together around the ideal diagonal line, suggesting that the model suited the data well. However, the model showed some under- and over-predictions at higher price points (beyond ₹80,000), indicating that the luxury market is more variable.   
Random Forest fared noticeably better than Linear Regression when assessing model metrics. About 68% of the price variance was explained by the linear regression model, which had an R2 of 0.6844 and an RMSE of ₹42,623.08. The Random Forest model, on the other hand, is better at handling nonlinear relationships and feature interactions, as evidenced by its lower RMSE of ₹31,290.99 and R2 of 0.8791, which captures 88% of the variation.

The model performed well in most cases, as evidenced by the low median error (₹739.79) found in an error distribution study. The comparatively high maximum error, however, indicated the existence of outliers or edge instances where the model had trouble. Prediction accuracy was quite high for lower fares (₹0–₹32,000), according to further study by price range. However, mistakes increased dramatically for higher-priced tickets, particularly in the business class long-haul category.

The Random Forest model was found to be a reliable indicator of flight costs, especially for the low-to-mid price range. The main motivating elements were class and duration, with other factors having just slight influence. Building distinct models for various travel classes, paying more attention to high-price outliers, and adding extra features like seasonality or holiday factors to better capture pricing dynamics across all segments are all suggested ways to improve the model's effectiveness.

**Price Elasticity Analysis to study demand sensitivity.**

A graph with red lines

AI-generated content may be incorrect.

Using both standard and alternative point elasticity calculations, the Price Elasticity of Demand (PED) analysis of aircraft routes was conducted per class (Economy and Business). The usual method was used to assess 53 routes, and the results showed that 23 of them were extremely elastic (demand changes more than proportionately with price) and 30 were inelastic (demand changes less than proportionately with price). With a mean elasticity of roughly -4.75, the data reveals that business class routes are extremely elastic when broken down by class, meaning that demand for business class declines significantly as costs rise. Routes in economy class, on the other hand, are largely inelastic, indicating that passengers are less susceptible to price fluctuations, most likely as a result of financial constraints or necessary travel requirements.

These results were further corroborated by 54 routes using the alternative point elasticity approach. The elasticity breakdown in this case revealed eight inelastic, eight very inelastic, fourteen elastic, and twenty-four highly elastic routes. whereas economy routes were marginally more elastic than in the usual technique, they remained largely fell between inelastic and elastic, whereas business class routes continued to show noticeably higher elasticity.

The data also revealed the Top 10 most price-sensitive (very elastic) itineraries, the majority of which were in business class. This further supports the idea that business travelers are extremely sensitive to price fluctuations, most likely as a result of more flexible scheduling or work travel regulations. The availability of less expensive options may have contributed to the high elasticity of a select budget routes, such as Mumbai to Kolkata. The Top 10 least price-sensitive (inelastic) routes, on the other hand, were mostly in economy class and usually covered short-haul, necessity-driven travel where demand is constant regardless of price changes.

Based on organizational or personal cost considerations, the study concludes that business class passengers are significantly more sensitive to rate hikes than economy passengers. In the meanwhile, despite slight fare adjustments, economy travelers typically continue to make reservations. With the use of this elasticity research, airlines may better segment their pricing strategies across various passenger categories, forecast changes in demand with fare modifications, and improve revenue by implementing dynamic pricing methods.

**Anomaly Detection (Isolation Forest) to identify extreme price spikes.**

A graph showing a red and blue line

AI-generated content may be incorrect.

Z-Score and Isolation Forest techniques are used to analyze flight pricing data for anomalies. Approximately 4.81% of the dataset, or 14,431 abnormalities, were found out of the 300,521 records that were analyzed. The price of anomalies was ₹76,078.08 on average, ₹113,871.08 at the highest, and ₹13,528.08 at the lowest. The majority of anomalies were high-priced outliers, which could be a sign of price gouging, data entry problems, or unusual demand surges.

The majority of anomalies were found to be caused by Vistara and Air India when anomalies were broken down by airline, most likely as a result of their premium pricing policies. Budget airlines like Indigo and AirAsia, on the other hand, displayed less anomalies, indicating that their pricing strategies are more stable. Routes connecting big cities also had the highest frequency of pricing anomalies, according to the data. This could be because to supply-demand imbalances, last-minute business travel, or peak travel seasons.

A graph of a graph

AI-generated content may be incorrect.

Visualizations supported these findings: the top plot displayed anomalous prices as red dots scattered widely across the dataset, with a notable cluster above the ₹60,000 range, while normal prices (blue dots) formed a more consistent distribution. The bottom plot, a price distribution histogram, showed that normal prices peaked in the ₹5,000–₹20,000 range, whereas anomalies were more frequent at much higher price points, highlighting the skew caused by premium or irregular fares.

In conclusion, the anomaly detection analysis found that about 4.8% of flight price data points were irregular, predominantly involving premium airlines and busy routes. These anomalies likely reflect dynamic pricing adjustments, data entry mistakes, or high-demand scenarios such as festivals and emergencies. The use of the Isolation Forest model proved effective in identifying these outliers for further review, investigation, or data cleaning to improve pricing analysis accuracy.

**Monte Carlo Simulation for booking timing.**

A graph of a graph showing the growth of the economy

AI-generated content may be incorrect.

The booking recommendation analysis for a flight from Delhi to Mumbai (Economy Class) suggests that travelers should book now at ₹5000, as waiting is unlikely to offer any price advantage. The system forecasts that prices are likely to rise, with a 60–62% probability of increase each day over the next 10 days, reinforcing the urgency to secure the current fare. A detailed price forecast graph shows a clear upward trend in expected ticket prices, with the lowest median and lower bound price appearing today. The forecast also highlights that the uncertainty in pricing grows the longer one waits, making future prices less predictable and likely higher. A summary table further confirms that today is the optimal day to book, and delaying the purchase would expose travelers to both higher prices and greater risk. In conclusion, for this specific Delhi to Mumbai flight, the current fare of ₹5000 is optimal, and it is strongly recommended to book immediately to lock in the best price and avoid future fare increases.

**Reinforcement Learning (Q-Learning Agent) for dynamic pricing strategy.**

A graph and a chart

AI-generated content may be incorrect.

This compares Reinforcement Learning (RL)-based dynamic pricing with fixed pricing for maximizing flight booking revenues.

* **Training Progress:**  
  Revenue increased from ~₹534,000 to ~₹563,000 over 1000 training episodes, while the exploration rate decreased as the model shifted from trying new strategies to exploiting the best ones.  
  The RL model generated over 11% higher revenue compared to fixed pricing, showing that adaptive pricing better captures value from varying booking behaviors.
* **Visual Insights:**

**Learning Curve:** Revenue initially fluctuated but became more stable and consistently higher as training progressed.

**Revenue Comparison:** The RL agent clearly outperformed fixed pricing in average revenue.

Reinforcement Learning effectively learned an optimal dynamic pricing strategy, boosting revenue by ~11.3%, and is highly valuable for dynamic markets like flight bookings.

**Alternative Models Considered:**

• Decision Tree Regressor

• ARIMA for time-series-based fare forecasting

• Clustering for route segmentation

## Findings

**Exploratory Analysis**

• Afternoon flights are cheapest; night flights are most expensive.

• Premium carriers dominate higher fare segments.

• Booking earlier leads to ~27–58% savings (depending on class).

**Regression & Random Forest Models**

• Random Forest achieved R² = 0.8791 (vs. 0.6844 for Linear Regression).

• Top predictors: Class, Duration, and Days Left.

• Median absolute error = ₹739.79; higher errors in expensive segments.

**Anomaly Detection**

• 4.81% of prices flagged as anomalies.

• Anomalies concentrated in premium airlines and metro city routes.

**Price Elasticity**

• Business routes are highly elastic (avg. -4.75).

• Economy routes mostly inelastic (avg. -0.47).

• Price-sensitive routes include Delhi → Hyderabad and Kolkata → Bangalore.

**Monte Carlo Simulation**

• Predicted price trend showed a 60–62% chance of price rise within 10 days.

• Recommended “Book Now” for optimal price capture.

**Reinforcement Learning (RL)**

• RL agent improved average revenue by 11.3% over fixed pricing.

• Learning curve showed performance stabilization after ~600 episodes.

## Recommendation

1. Dynamic Pricing Strategy: Implement RL-based pricing to maximize revenue under varying demand.

2. Early Booking Incentives: Offer discounts for early Economy bookings to capture price-sensitive consumers.

3. Segmented Targeting: Customize fares based on traveler type (business vs. leisure).

4. Real-time Anomaly Monitoring: Flag and validate high-price outliers to maintain pricing credibility.

## Discussion

Limitations:

• Simulated dataset; real-world deployment would require integration with live fare APIs.

• Did not factor in holidays, competitor pricing, or weather.

• RL agent tested on historical simulations only.

Future Work:

• Incorporate real-time search data and seasonality.

• Enhance RL model using deep Q-networks (DQN).

• Deploy in a web-based fare recommendation tool.

## References

• [Flight Price Prediction](https://www.kaggle.com/datasets/shubhambathwal/flight-price-prediction?resource=download)

• Sutton, R.S., & Barto, A.G. (2018). Reinforcement Learning: An Introduction

• Sklearn Documentation

• Airline Revenue Management Research Papers

• Pandas/Seaborn/Scikit-learn Libraries