THE UNIVERSITY OF CHICAGO
MS–APPLIED DATA SCIENCE

Airline Pricing & Revenue Optimization Practice

Predicting Flight Prices: Data-Driven Fare Optimization

Research across airline markets provides insights into ticket price fluctuations, helping stakeholders optimize fare strategies, traveler cost savings, and revenue management through advanced analytics and machine learning.

by Yukai Huang, Zoey Hu, Yueyang Liu, and Wayne Wu

March 2025

Executive Summary

Airfare pricing is highly dynamic, influenced by factors such as demand, airline competition, booking lead time, and flight duration. This project applies machine learning and data analytics to predict flight ticket prices, providing insights that help both travelers and airlines make data-driven decisions.

Key Objectives

- Identify key pricing factors: Analyzing how airlines, departure times, stopovers, and booking lead time affect ticket costs.
- Develop predictive models: Utilizing Linear Regression, Random Forest, XGBoost, and Gradient Boosting to forecast airfare trends.
- Compare Economy vs. Business class pricing: Understanding pricing structures for different travel categories.
- Optimize booking strategies: Identifying the best times to purchase tickets for cost savings.

Business Significance

- Travelers: Provides data-driven insights to book flights at the lowest possible prices.
- For Airlines & Travel Agencies: Enhances dynamic pricing strategies and revenue management.
- For Industry Researchers: Establishes a structured framework for predictive modeling in airfare pricing.

Methodology

- Data Processing & Exploration: Cleaning and analyzing 300,000+ flight records to identify pricing patterns.
- Feature Engineering: Creating key variables such as departure time categories, booking lead time, and route popularity.
- Model Development: Training multiple machine learning models and evaluating them based on Mean Absolute Error (MAE), RMSE, and R^2 Score.
- Insights & Recommendations: Summarizing key takeaways for travelers to book smarter and airlines to refine pricing models.

By leveraging machine learning, this project enhances pricing transparency, optimizes airfare booking decisions, and supports revenue management strategies. The results provide valuable insights for both consumers and industry stakeholders, driving smarter decision-making in airfare pricing.

Data-driven strategies that will transform airfare pricing

Table of Contents

Introduction: Business Background & Significance

Project Background

Airfare pricing is a complex and highly dynamic process influenced by various factors such as demand, airline competition, seasonality, fuel prices, and even geopolitical events. Unlike static pricing in traditional retail, airline ticket prices fluctuate constantly, sometimes changing multiple times a day.

This variability creates challenges for both consumers and airlines:

- *For travelers:* unpredictable pricing makes it difficult to determine the best time to book flights, often resulting in higher costs or missed opportunities for cheaper fares.
- *For airlines:* optimizing pricing is crucial for revenue management, ensuring they maximize profitability while maintaining competitive ticket prices.

To address these challenges, machine learning and data analytics have become essential tools in the airline industry. By leveraging historical flight data and predictive modeling, airlines can refine their dynamic pricing strategies, and consumers can gain data-driven insights on when to book flights at the lowest prices.

This project aims to bridge the gap by utilizing data science to analyze and predict flight ticket prices, offering valuable insights into airfare trends and helping both travelers and airlines make more informed decisions.

Project Objectives

This research focuses on analyzing flight booking data from the "Ease My Trip" platform, containing over 300,000 flight records with key attributes such as: Airline name, Flight duration, Departure and arrival times, Number of stops, Booking lead time (time between booking and departure), Ticket class (Economy, Business).

Using machine learning models, we aim to:

— **Identify key factors influencing airfare fluctuations**

- How significantly do different airlines affect ticket prices?
- How do departure and arrival times impact pricing?
- What role does the number of stops play in determining costs?

— **Develop predictive models to estimate flight ticket prices**

- Using regression-based machine learning models to predict airfare based on historical data.

— **Understand the impact of booking lead time**

- Examining how ticket prices vary when bookings are made at different time intervals before departure (e.g., same-day, one week before, one month before).

— **Analyze pricing differences between Economy and Business class tickets**

- Quantifying and comparing price variations between ticket classes.

— **Provide actionable insights for travelers and airlines**

- Helping consumers find the best booking strategies for lower fares
- Enabling airlines to refine their pricing models using data-driven approaches.

Business Significance

This project holds significant value for multiple stakeholders in the travel and airline industry:

— **For Consumers (Travelers & Business Travelers)**

- Smarter Booking Decisions: Enables travelers to determine the optimal time to book flights at the best rates.
- Cost Savings: Avoid unnecessary expenses by understanding how price fluctuations work.
- Personalized Fare Predictions: Future applications could involve integrating predictive pricing into travel apps.

— **For Airlines & Travel Agencies**

- Optimized Revenue Management: Airlines can fine-tune pricing models to maximize occupancy while maintaining profitability.
- Competitive Pricing Analysis: Insights into how competitors price similar routes can drive better pricing strategies.
- Dynamic Pricing Enhancements: Airlines can leverage predictive models to adjust pricing in real time.

— **For Data Scientists & Researchers in Aviation**

- Benchmark Dataset for Predictive Modeling: Provides structured data for training and testing machine learning models.
- New Applications in AI-driven Pricing: Potential for real-time fare prediction tools for both consumers and businesses.
- Industry-Level Impact: Supports broader research into demand forecasting, price elasticity, and airline competition dynamics.

Methodology: A Data-Driven Approach

To achieve these objectives, the project will follow a structured machine learning pipeline, leveraging advanced techniques for price prediction:

— **Data Preprocessing & Exploration:**

- Cleaning and structuring data, handling missing values, and formatting inconsistencies.
- Performing exploratory data analysis (EDA) to understand trends and distributions.

— **Feature Engineering:**

- Extracting relevant time-based and route-based features.
- Identifying price outliers and incorporating flight duration metrics.

— **Model Development & Evaluation:**

- Implementing multiple regression-based models: Linear Regression, Random Forest, XGBoost, and Gradient Boosting.
- Evaluating performance using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score.
- Applying the best-performing model to predict future flight prices.

— **Insights & Recommendation:**

- Providing actionable insights in a final report with key takeaways for travelers.

Conclusion: Why This Matters

The ability to predict flight ticket prices accurately can transform the way travelers book flights and how airlines manage pricing. With airfare being a major cost factor in travel, having a data-driven approach to price prediction can lead to substantial savings and better planning for both individuals and businesses.

This study aims to provide a clear, data-backed strategy for:

- Helping consumers book flights at optimal prices
- Empowering airlines with more effective pricing models
- Advancing the use of machine learning in travel analytics

By leveraging predictive analytics, this research paves the way for **greater transparency, smarter decision-making, and innovation in airline revenue management.**

Data Preprocessing & Exploration

Data Overview

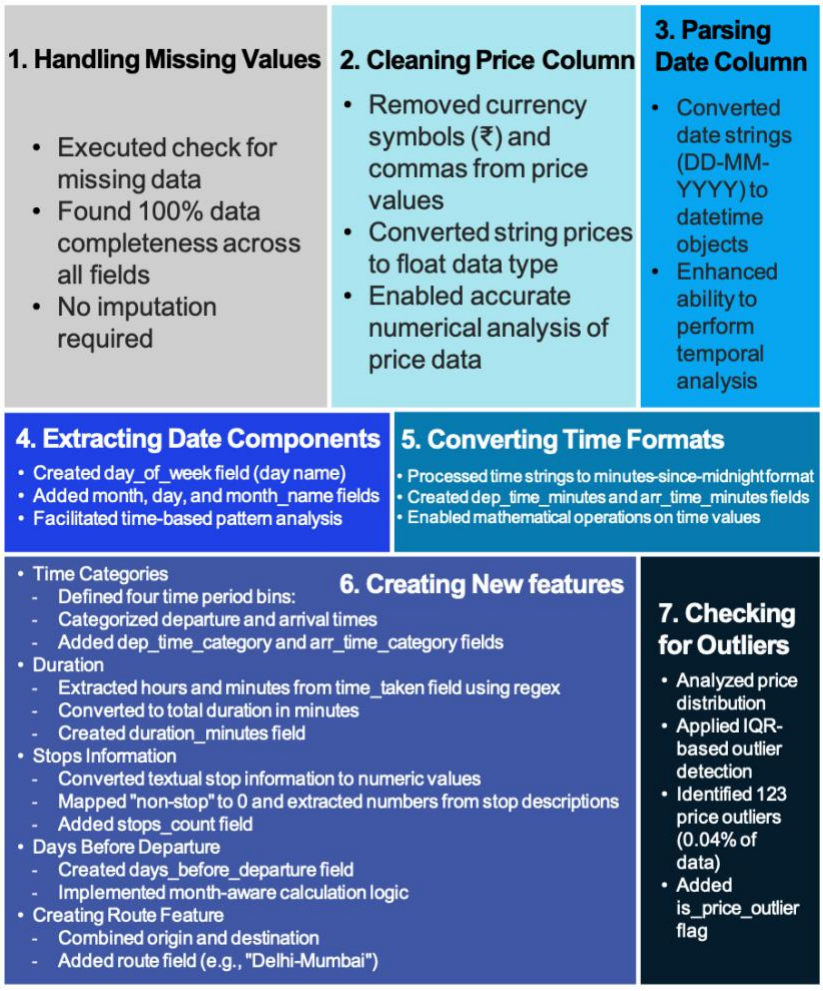
This comprehensive flight dataset consists of 12 key variables that capture essential information about each flight record. The "date" field records when each flight operates, while "airline" and "ch_code" (carrier code) identify the operating carrier, with "num_code" providing the specific flight number. Flight timing is tracked through "dep_time" (departure time), "arr_time" (arrival time), and "time_taken" (duration). Route information is captured by "from" (origin), "to" (destination), and "stop" variables, indicating whether the flight is non-stop or includes layovers.

The dataset also includes economic information through the "price" variable and service level through the "class" field, distinguishing between Economy and Business offerings. With 300,261 complete records across all variables, this dataset provides a robust foundation for analyzing flight patterns, pricing strategies, and service offerings between various city pairs.

Data Cleaning and Preprocessing

Exhibit 2.1

Preprocessing Steps Implemented



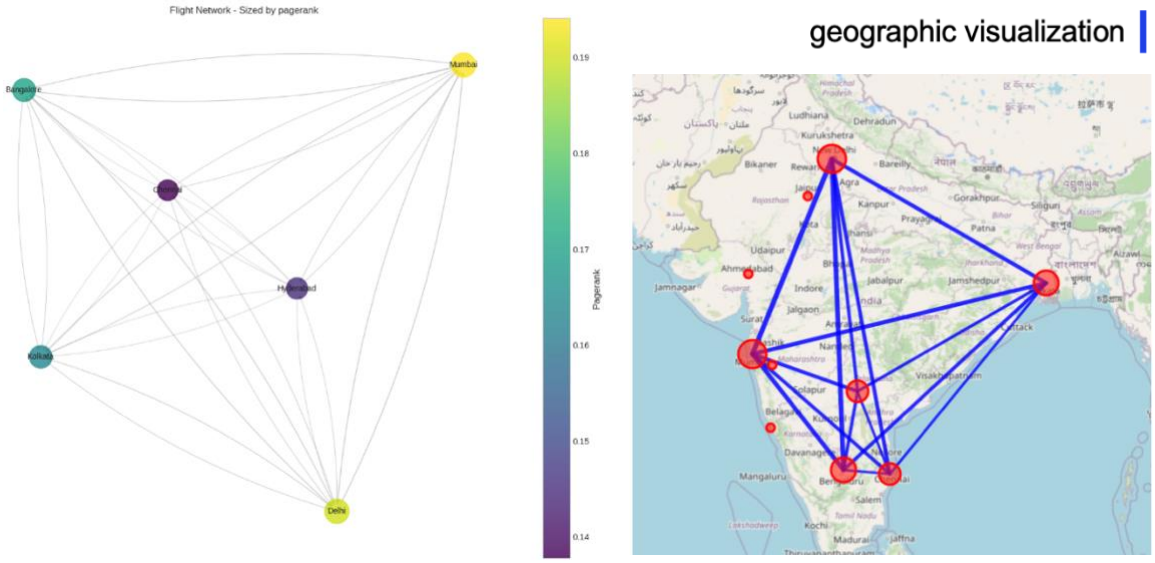
The preprocessing pipeline successfully transformed the dataset:

- Original Data Shape: (300,261, 12)
- Cleaned Data Shape: (300,261, 25)
- Features Created: 13 new analytical fields

Network Analysis

We identified a highly interconnected network of 6 major airports connected by 30 routes, with Mumbai and Delhi emerging as the dominant hubs based on PageRank analysis.

Exhibit 2.2



The color gradient (yellow to purple) representing PageRank values visually confirms the hierarchy from top hubs (Mumbai, Delhi) to secondary hubs (Bangalore, Kolkata) to regional centers (Hyderabad, Chennai).

Exhibit 2.3

Route Analysis

Rank	Route	Frequency	Avg. Price (₹)	Duration (min)	Price/Min (₹)
1	Delhi-Mumbai	15,291	19,354.41	622.15	31.11
2	Mumbai-Delhi	14,809	18,725.32	589.07	31.79
3	Delhi-Bangalore	14,012	17,880.22	621.24	28.78
4	Bangalore-Delhi	13,756	17,723.31	586.79	30.20
5	Bangalore-Mumbai	12,940	23,127.23	654.29	35.35

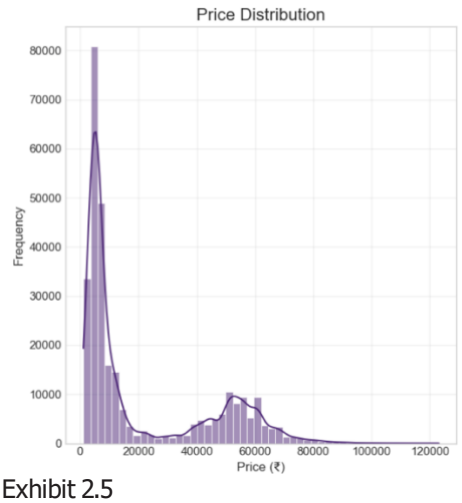
Key observations:

- Delhi-Mumbai corridor is the busiest route with bidirectional traffic exceeding 30,000 flights
- Bangalore-Mumbai route has the highest price per minute (₹35.35), suggesting premium pricing
- Route pricing is generally asymmetric (different prices depending on direction)
- Average flight durations range from 586 to 654 minutes (~9.8 to 10.9 hours)

Exploratory Data Analysis

Exhibit 2.4

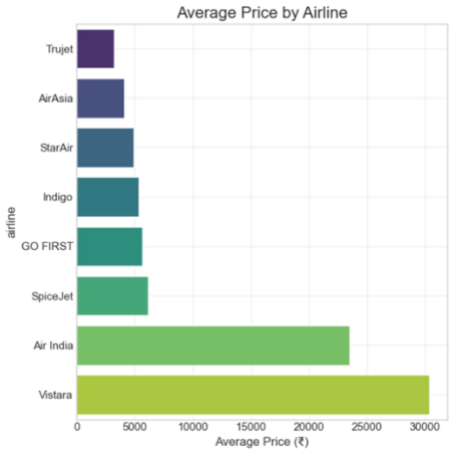
Price Distribution Patterns



The price distribution histogram reveals a distinct bimodal distribution, indicating two primary pricing tiers in the market. The majority of flight prices are concentrated in the lower range (below ₹20,000), creating a prominent peak. A secondary, smaller peak appears around the ₹40,000-60,000 range, likely representing business class fares. This distribution suggests that the market serves distinctly different customer segments with limited middle-ground pricing.

Exhibit 2.5

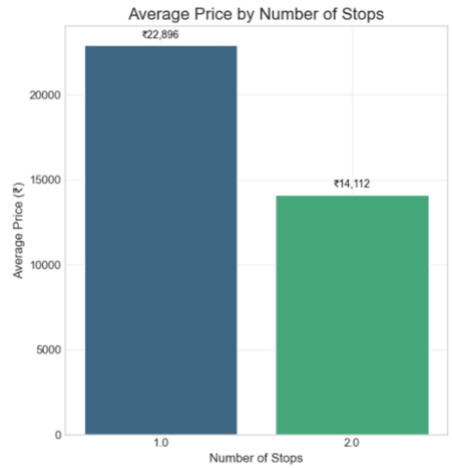
Airline Pricing Strategies



The price distribution histogram reveals a distinct bimodal distribution, indicating two primary pricing tiers in the market. The majority of flight prices are concentrated in the lower range (below ₹20,000), creating a prominent peak. A secondary, smaller peak appears around the ₹40,000-60,000 range, likely representing business class fares. This distribution suggests that the market serves distinctly different customer segments with limited middle-ground pricing.

Exhibit 2.6

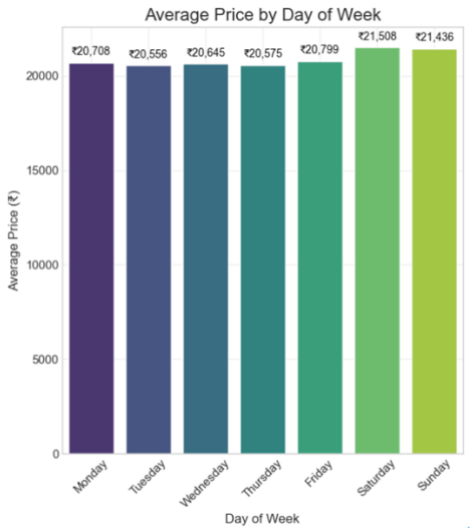
Impact of Stops on Pricing



The analysis clearly demonstrates that non-stop flights (0 stops) command a substantial premium, with an average price of approximately ₹22,596. Flights with one stop show a marked reduction, averaging around ₹14,112. This represents a nearly 38% price decrease for accepting one connection, suggesting that connections are significantly factored into pricing algorithms. This relationship provides valuable information for cost-conscious travelers willing to accept longer journeys for better value.

Exhibit 2.7

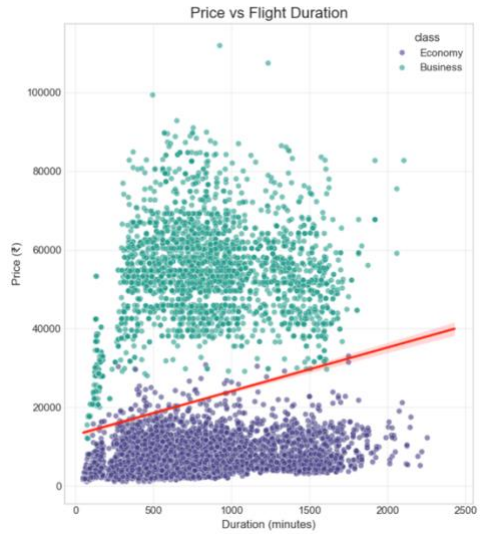
Day of Week Pricing Patterns



The visualization of average prices by day of week reveals moderate but meaningful price fluctuations throughout the week. Saturday and Sunday demonstrate the highest average fares (₹21,808 and ₹21,436 respectively), confirming the weekend premium typically associated with leisure travel. Prices remain relatively stable from Monday through Friday, with only minor variations between ₹20,556 and ₹20,799. The weekend premium, while present, is relatively modest at approximately 5% above weekday rates.

Exhibit 2.8

Price vs. Flight Duration Relationship



The scatter plot examining price versus flight duration reveals several important patterns:

- There is a clear class-based separation, with business class fares (green points) consistently higher than economy fares (purple points) regardless of duration.
- The red regression line indicates a general positive correlation between duration and price, confirming that longer flights tend to cost more.
- The wide vertical spread of points at similar durations suggests that factors beyond flight time significantly influence pricing within each class.
- The relationship appears stronger within the economy class segment, where duration seems to more consistently predict price increases.

Practical Implications:

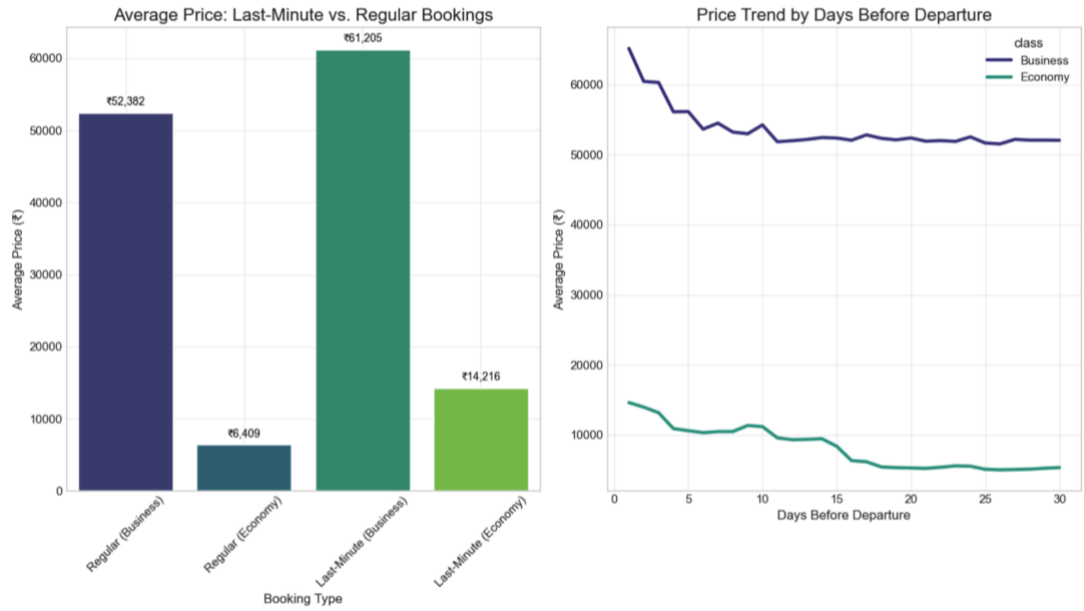
1. Airlines appear to be employing sophisticated price discrimination strategies across carriers, classes, and booking windows to maximize revenue while serving diverse market segments.
2. The pricing premium for non-stop flights suggests travelers place significant value on convenience, which carriers can leverage in their pricing strategies.
3. The relatively modest correlation between price and other factors suggests that market positioning and brand perception may play substantial roles in pricing beyond purely operational factors.

This comprehensive analysis provides valuable insights into the determinants of flight pricing in the Indian market, highlighting the complex interplay between carrier choice, class selection, scheduling, and booking strategy in determining the final fare.

Deeper Research

Last-Minute Booking Effect on Flight Prices: How is Price Affected When Tickets are Bought Just 1-2 Days Before Departure?

Exhibit 2.9



The analysis of last-minute bookings reveals a substantial price premium for tickets purchased 1-2 days before departure. The tabular data presents compelling evidence of pricing differentiation based on both booking timing and travel class.

Regular business class tickets demonstrate a robust pricing structure with a mean price of ₹52,382, a median of ₹53,164, and a standard deviation of ₹12,853. This relatively narrow standard deviation indicates consistent pricing practices for advanced business class bookings. The price range extends from ₹12,000 to ₹123,071, suggesting that while baseline prices remain stable, route-specific factors can significantly influence the upper boundary of pricing.

When examining last-minute business class bookings, a marked price escalation is evident, with the mean price increasing to ₹61,205 - representing a 16.8% premium over regular bookings. The median rises to ₹59,948, and notably, the standard deviation increases substantially to ₹16,025, indicating greater price volatility for urgent travel needs. The price floor increases by 28% to ₹15,360, demonstrating that even the most economical last-minute business options command a significant premium.

The economy class segment displays even more dramatic pricing differentials. Regular economy bookings average ₹6,409 with a median of ₹5,699, while last-minute economy bookings surge to an average of ₹14,216 with a median of ₹13,614. This represents a remarkable 121.8% increase in mean price - more than double the regular rate. The standard deviation similarly increases from ₹3,521 for regular bookings to ₹5,530 for last-minute bookings, confirming heightened price variability.

The bar chart visualization reinforces these findings by clearly illustrating the price escalation across both classes. The most dramatic price differential appears in the economy class segment, where the relative increase is substantially greater than in the business class segment. This suggests that airlines implement more aggressive dynamic pricing for economy tickets as departure approaches, potentially capitalizing on less flexible travel requirements.

The price trend line by days before departure provides additional context by demonstrating the progressive price increases as the departure date approaches. Business class fares exhibit a steep initial incline in the final 5 days before departure, followed by more moderate increases between 5-10 days out. Economy class shows a similar pattern but with more pronounced percentage increases in the final days.

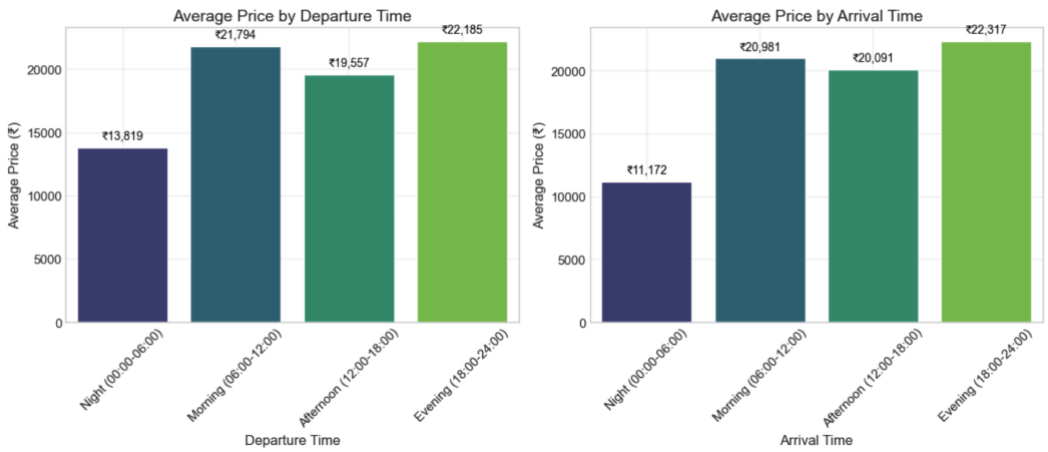
This pricing structure indicates a sophisticated yield management system that maximizes revenue by identifying and charging premium rates to travelers with less scheduling flexibility. The data suggests that booking at least 7-10 days in advance can yield substantial savings, with the most favorable pricing available 15+ days before departure.

Price Variations by Time-of-Day Analysis: Does Ticket Price Change Based on Departure and Arrival Time?

Exhibit 2.10

Price by Departure and Arrival Time Categories:

	Departure Time	Arrival Time	Mean Price (₹)	Median Price (₹)	Count
14	Evening (18:00-24:00)	Afternoon (12:00-18:00)	28459.063129065387	12792.0	14605
9	Afternoon (12:00-18:00)	Morning (06:00-12:00)	26695.274054529462	11310.0	17055
4	Morning (06:00-12:00)	Night (00:00-06:00)	26070.137432188065	9627.0	2212
7	Morning (06:00-12:00)	Evening (18:00-24:00)	26052.683760960826	10643.0	61697
13	Evening (18:00-24:00)	Morning (06:00-12:00)	24005.7007405532	7766.5	31598
15	Evening (18:00-24:00)	Evening (18:00-24:00)	20233.381579422057	7413.0	27719
3	Night (00:00-06:00)	Evening (18:00-24:00)	19961.87227124942	7666.5	4306
11	Afternoon (12:00-18:00)	Evening (18:00-24:00)	18616.087142694454	6651.0	43905
6	Morning (06:00-12:00)	Afternoon (12:00-18:00)	18378.74400344197	6933.0	37188
10	Afternoon (12:00-18:00)	Afternoon (12:00-18:00)	17949.192185067102	7220.0	13487
5	Morning (06:00-12:00)	Morning (06:00-12:00)	15380.822099447514	6105.0	22625
2	Night (00:00-06:00)	Afternoon (12:00-18:00)	14337.589559479915	6221.0	5153
8	Afternoon (12:00-18:00)	Night (00:00-06:00)	10218.103612923762	5102.0	6283
1	Night (00:00-06:00)	Morning (06:00-12:00)	9093.034785727094	4558.0	5577
0	Night (00:00-06:00)	Night (00:00-06:00)	8762.259597806216	4498.0	547
12	Evening (18:00-24:00)	Night (00:00-06:00)	7105.03711928934	4121.0	6304



The comprehensive analysis of departure and arrival time influences on ticket pricing reveals significant price differentials based on the time of day travelers choose to fly. The data demonstrates that both departure and arrival times function as critical variables in airline pricing algorithms.

The tabular data categorizes flight times into four distinct periods: Night (00:00-06:00), Morning (06:00-12:00), Afternoon (12:00-18:00), and Evening (18:00-24:00). Among the 16 possible departure-arrival time combinations, flights departing during the Evening and arriving in the Afternoon command the highest average price (₹28,459), significantly exceeding the median (₹12,792). This substantial differential between mean and median indicates a right-skewed distribution with premium pricing for select routes or carriers during these time slots.

Conversely, the lowest-priced combination involves Night departures with Night arrivals, averaging ₹8,762 with a median of ₹4,498. This represents a striking 69.2% reduction compared to the highest-priced category. The substantial difference indicates that passengers willing to accept the inconvenience of overnight travel can secure significantly lower fares.

When examining departure time in isolation, Evening departures command the highest average price (₹22,185), followed closely by Morning departures (₹21,794). Afternoon departures are priced moderately (₹19,557), while Night departures offer the most economical options (₹13,819). This pricing structure reflects passenger preference for daytime travel, with a particular premium placed on evening departures that may allow travelers to maintain productive workdays before travel.

For arrival times, a similar pattern emerges with Evening arrivals commanding the highest premium (₹22,317), followed by Morning arrivals (₹20,981) and Afternoon arrivals (₹20,091). Night arrivals are significantly discounted (₹11,172), representing a 49.9% reduction compared to evening arrivals. This substantial difference highlights passengers' strong preference to avoid overnight arrivals, which airlines accommodate through reduced pricing.

The bar chart visualization effectively illustrates these pricing differentials, showing a consistent premium for Morning and Evening travel periods across both departures and arrivals. The most striking observation is the substantially lower pricing for Night-time operations, which appears consistent regardless of whether it applies to the departure or arrival segment of the journey.

Examining the data more granularly reveals interesting patterns in specific time combinations. The Evening departure to Morning arrival combination, often representing overnight long-haul flights, maintains relatively high pricing (₹24,005), suggesting that for certain route types, the convenience of overnight travel without losing productive daytime hours commands a premium rather than a discount.

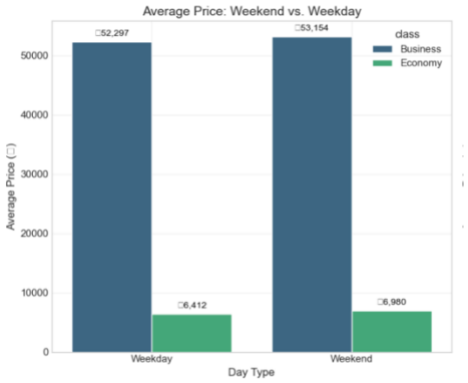
This time-based pricing structure demonstrates airlines' sophisticated segmentation strategy, effectively distinguishing between business and leisure travelers based on their scheduling preferences and willingness to pay. The data indicates that price-sensitive travelers can secure substantial savings by selecting Night departures or arrivals, while those prioritizing convenience will pay premiums for daytime and especially evening travel options.

Feature Engineering

There are **19** selected features in total to build the model, **8** are newly added engineered features:

Exhibit 3.1

1 is_weekend



— **Definition**

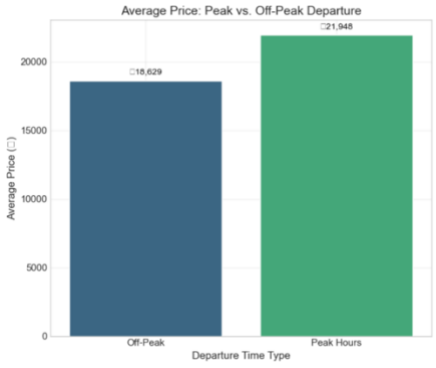
A binary indicator (0 or 1) showing whether the flight day is a weekend (Saturday or Sunday)

— **Why It's Useful**

Weekend flights may have different demand patterns and pricing dynamics compared to weekday flights. Plot shows weekend has higher prices than weekday for both business and economy

Exhibit 3.2

2 is_peak_departure



— **Definition**

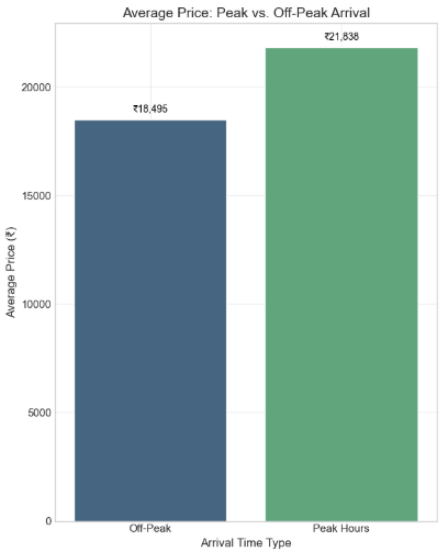
A binary indicator that flags whether the flight's departure time falls within peak hours (Morning: 06:00-12:00 or Evening: 18:00-24:00)

— **Why It's Useful**

Flights during peak departure times may experience higher demand, affecting pricing and seat availability. Plot shows peak has higher flight prices

Exhibit 3.3

3 is_peak_arrival



— **Definition**

A binary indicator that identifies if the flight's arrival time is during peak hours (Morning: 06:00-12:00 or Evening: 18:00-24:00)

— **Why It's Useful**

Similar to departure, peak arrival times might influence operational factors and pricing due to passenger volume and airport congestion. Plot shows peak has higher flight prices

Exhibit 3.4

4 days_to_flight_factor

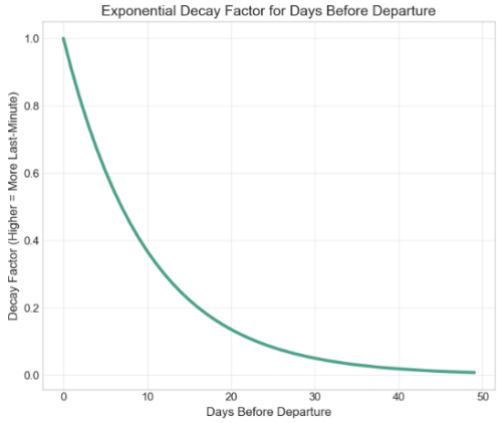


Exhibit 3.5

5 route_popularity

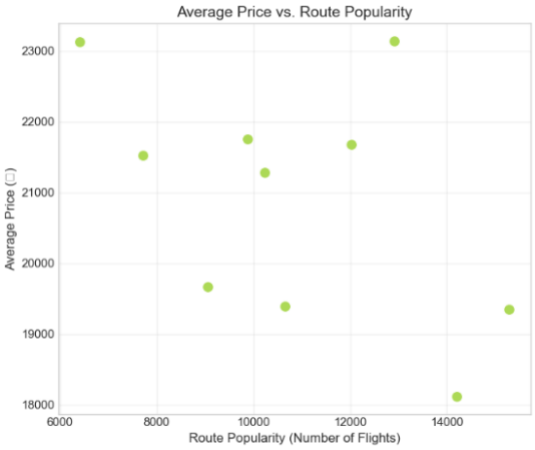
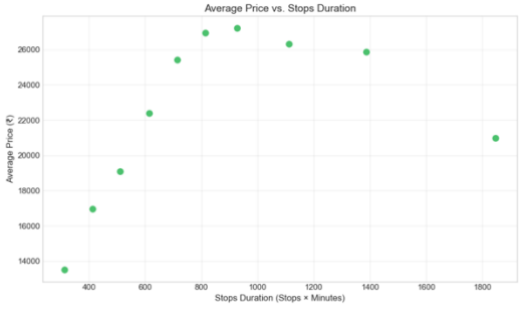


Exhibit 3.6

6 stops_duration



— **Definition**

A exp decay factor that captures the effect of the number of days before departure on flight price, assuming that the influence decays exponentially with time

— **Why It's Useful**

This feature can model the diminishing impact of advance booking on price; as the number of days before departure increases, the influence on price decreases exponentially

— **Definition**

A measure of how frequently a specific route appears in the dataset, we assume it the overall popularity

— **Why It's Useful**

Popular routes may have higher competition or demand, influencing pricing strategies and capacity planning. The plot shows that the more popular routes have lower average price which is interesting

— **How Calculated**

```
route_popularity =
engineered_df['route'].value_counts().to_dict()
engineered_df['route_popularity'] =
engineered_df['route'].map(route_popularity)
```

— **Definition**

An interaction feature that combines the number of stops and the flight's duration, capturing the compounded effect that multiple stops have on the overall travel time

— **Why It's Useful**

This feature helps quantify how additional stops may extend the travel time beyond just the flight duration, potentially affecting flight price and customer satisfaction

— **How calculated**

```
engineered_df['stops_duration'] =
engineered_df['stops_count'] *
engineered_df['duration_minutes']
```

7 route_avg_duration

— **Definition**

The average flight duration for each specific route, providing a baseline expectation for how long flights on that route typically take.

— **Why It's Useful**

By comparing an individual flight's duration to the average duration for its route, you can identify anomalies (e.g., unusually long or short flights), which might influence pricing decisions or reflect operational efficiency

— **How calculated**

```
route_avg_duration = engineered_df.groupby('route')['duration_minutes'].mean().to_dict()
engineered_df['route_avg_duration'] = engineered_df['route'].map(route_avg_duration)
```

8 duration_ratio

— **Definition**

The average flight duration for each specific route, providing a baseline expectation for how long flights on that route typically take.

— **Why It's Useful**

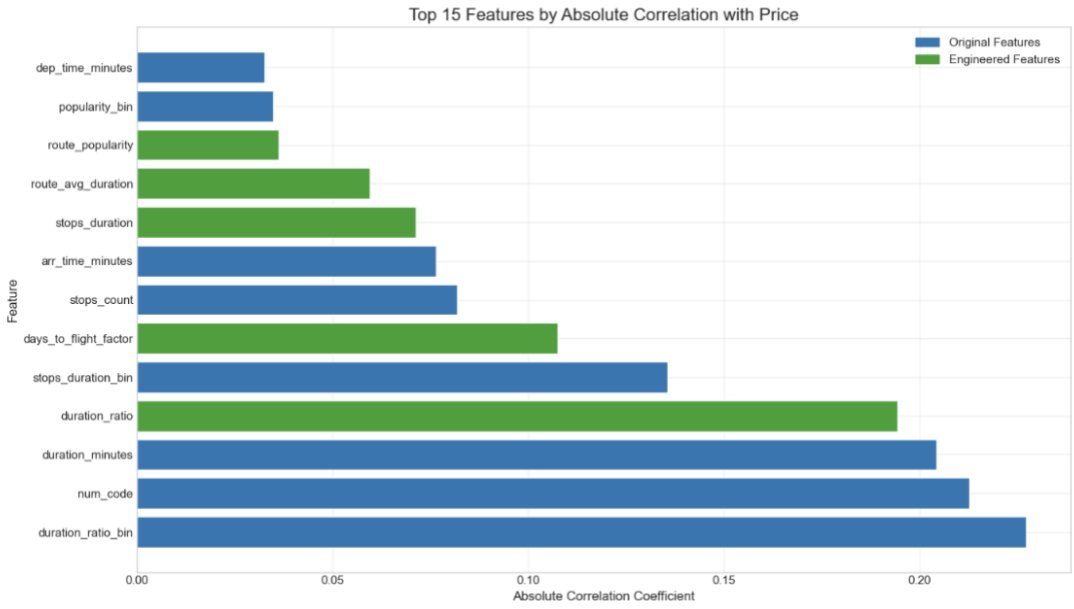
By comparing an individual flight's duration to the average duration for its route, you can identify anomalies (e.g., unusually long or short flights), which might influence pricing decisions or reflect operational efficiency

— **How calculated**

```
engineered_df['duration_ratio'] = engineered_df['duration_minutes'] /
engineered_df['route_avg_duration']
```

Exhibit 3.7

Features Correlation with Price



Model Development & Evaluation

We have previously build a model with both business and economy class flight tickets and the 'class' variable have profound inflencue in prediction, which covers all other variables's influence in predicting flight prices. Build a single model that includes "class" as a feature allows the model to learn differences in pricing within one framework, which is simpler to maintain but may not capture subtle differences as well if the patterns are very distinct. Therefore, we decide to Build distinct models for Business and Economy to capture the unique pricing patterns of each class more accurately. Since we also have enough sample size for both classes, we don't need to worry about overfitting.

The 19 features we used to build the models are:

— **6 categorical features**

airline, from, to, day_of_week, dep_time_category, arr_time_category

— **13 numerical features**

duration_minutes, stops_count, days_before_departure, dep_time_minutes, arr_time_minutes, is_weekend, is_peak_departure, is_peak_arrival, days_to_flight_factor, route_popularity, stops_duration, route_avg_duration, duration_ratio

— **Be aware of potential feature problem**

- 'route_popularity' - is a fixed number counting the number of tickets available in the website
- 'route_avg_duration', - calculated from the historical data that the average of the duration of the route
- These features are derived from the full dataset with historical data

Business

Transformation

Log-transform not used due to normal shape.

No log transformation applied.

Data split:

- Training set: 74,789 samples
- Testing set: 18,698 samples

Model Performance

Exhibit 4.1

Model Comparison

Model	RMSE	MAE	R^2
Random Forest	3649.753592	1677.436567	0.921232
XGBoost	5096.803493	3354.626522	0.846391
Gradient Boosting	7064.949250	5078.194044	0.704852
Ridge Regression	8896.079022	6813.028836	0.532029
Linear Regression	8896.091273	6812.948302	0.532028
Lasso Regression	8896.101238	6813.014479	0.532027

Exhibit 4.2

Best model based on RMSE: Random Forest

Economy

Transformation

Using log transformation for target variable based on distribution

Original price range: $1,105.00 - $42,349.00

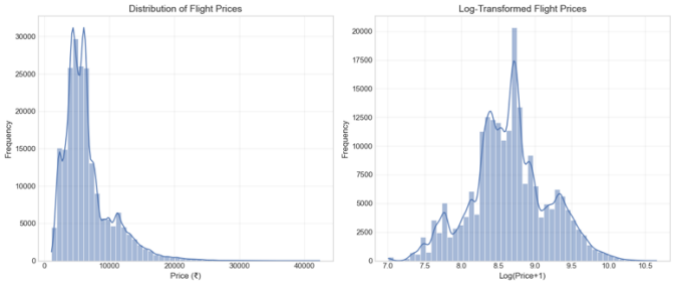
Log-transformed range: 7.01 - 10.65

Data split:

- Training set: 165,419 samples
- Testing set: 41,355 samples

Exhibit 4.3

Variable distribution



Model Performance

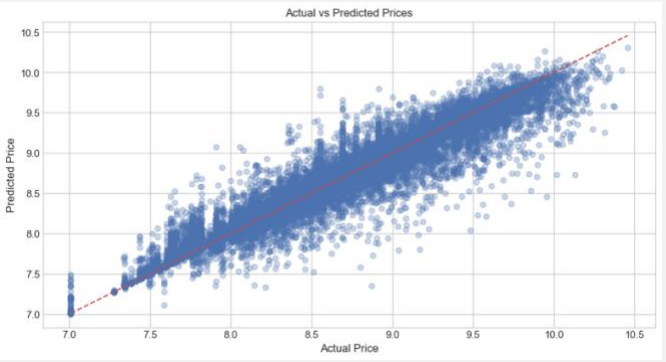
Exhibit 4.4

Model Comparison

Model	RMSE	MAE	R^2
Random Forest	0.132093	0.064871	0.937052
XGBoost	0.210470	0.151481	0.840188
Gradient Boosting	0.255485	0.193129	0.764518
Ridge Regression	0.316463	0.246611	0.638694
Linear Regression	0.316464	0.246613	0.638692
Lasso Regression	0.413585	0.322784	0.382898

Exhibit 4.5

Best model based on RMSE: Random Forest

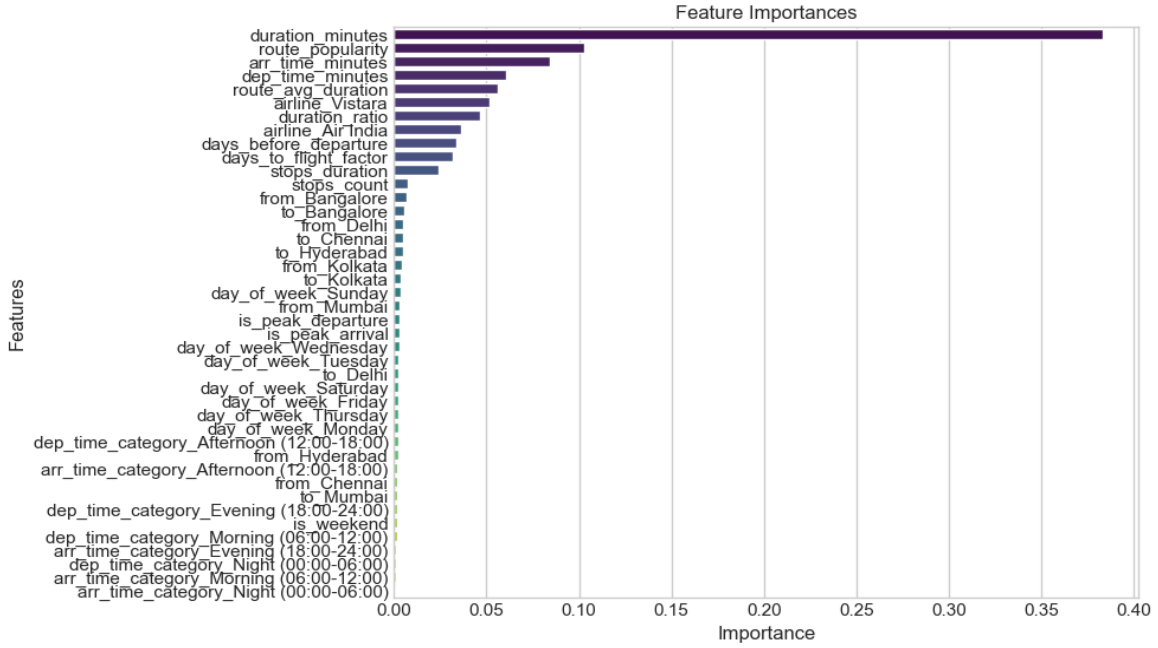


Insights & Recommendation

Insights from model

Exhibit 5.1

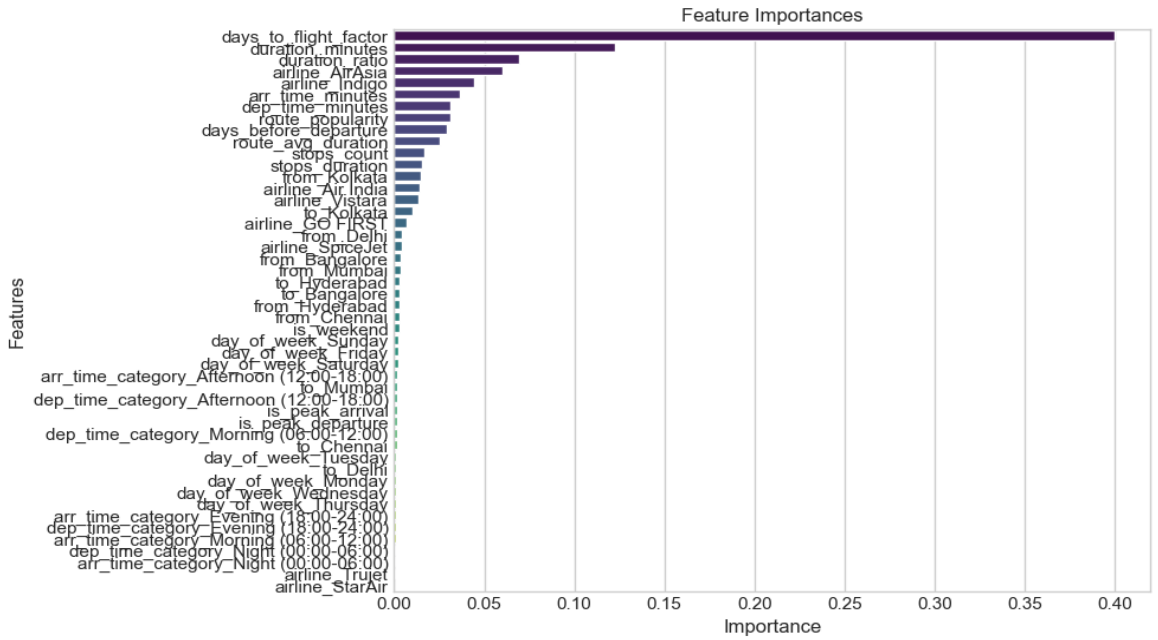
Key Price Drivers for Business tickets



- Duration (minutes) stands out as the most critical predictor for business class prices, suggesting that longer flights drive higher fares.
- Route popularity also ranks high, implying that well-traveled or in-demand routes command different price levels.
- Arrival time (minutes) and route_avg_duration further emphasize how timing and typical flight performance on a route influence pricing.
- Specific route and airline factors appear in the middle tier of importance, indicating that which route is flown and which carrier is operating still have a meaningful effect on the final price.
- Airline company is also important factor to consider.

Exhibit 5.2

Key Price Drivers for Economy tickets



- Days to flight factor is the most influential predictor for economy class pricing, highlighting the impact of how far in advance a ticket is booked.
- Duration minutes and duration ratio are also key factors, suggesting that longer flights are more expensive, but their impact is lower compared to business class pricing.
- Airline type (e.g., AirAsia, Indigo) plays a more significant role in economy pricing compared to business class, where specific airlines were less influential.
- Arrival and departure time in minutes still hold importance but rank lower than in business class predictions.
- Route popularity and days before departure are moderately important but not as dominant as in business class pricing.
- Stops count and stops duration contribute to pricing, though their impact is not as pronounced as other factors.

Difference between Business and Economy

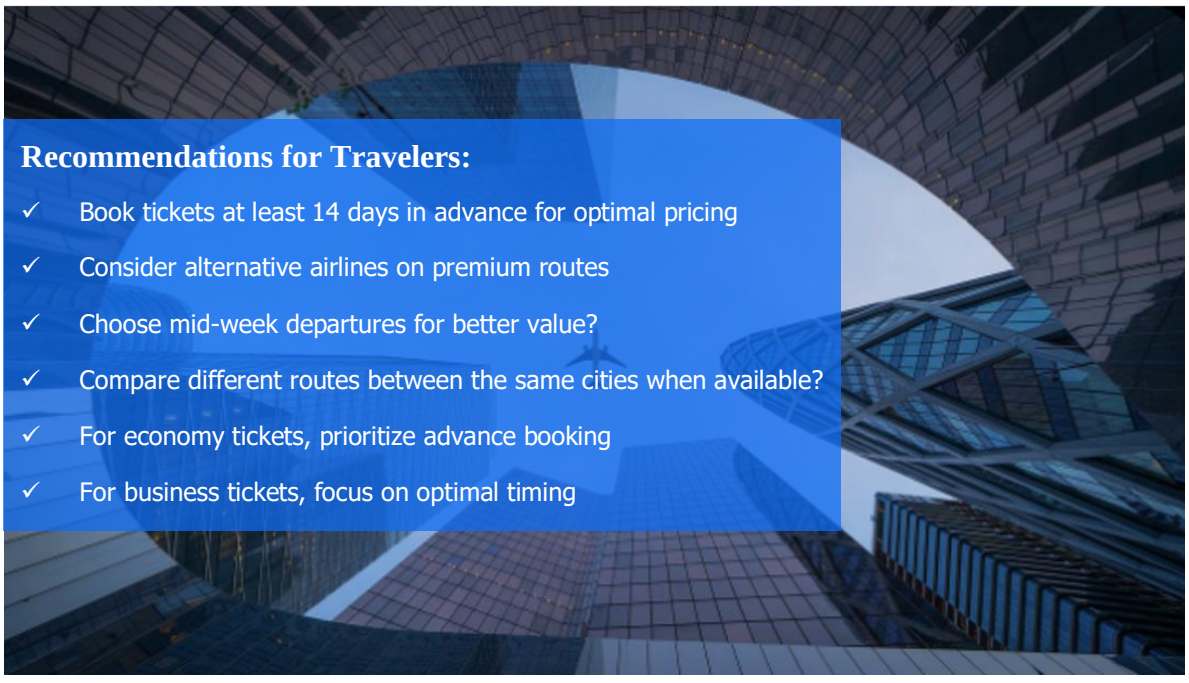
Booking time (Days to Flight Factor) matters more in economy class, meaning economy fares fluctuate more significantly based on how early the ticket is booked, whereas business class pricing is more stable.

Flight duration remains crucial in both, but **business class places more emphasis on total flight time**, while **economy class also considers duration efficiency (duration ratio).**

Airline choice plays a larger role in economy class pricing, likely due to budget airlines offering lower fares compared to premium carriers.

Route popularity is more influential in business class, possibly because business travelers prefer established routes with premium services.

Departure and arrival times have a stronger effect in business class, suggesting that business travelers may be more sensitive to flight timing for productivity reasons.



Recommendations for Travelers:

- ✓ Book tickets at least 14 days in advance for optimal pricing
- ✓ Consider alternative airlines on premium routes
- ✓ Choose mid-week departures for better value?
- ✓ Compare different routes between the same cities when available?
- ✓ For economy tickets, prioritize advance booking
- ✓ For business tickets, focus on optimal timing

Get in touch

Yukai Huang
Machine Learning Engineer
huang8@uchicago.edu

Zoey Hu
Data Scientist
ziyihu@uchicago.edu

Yueyang Liu
Data Analyst
yliu514@uchicago.edu

Wayne Wu
Business Intelligence Consultant
wenhanwu@uchicago.edu