

Airline Flight Price Analysis Case Study

Botond Koroknai

August 10, 2025

1 Ask: Define the Business Problem/Question

1.1 Business Problem

A travel agency seeks to enhance its service by understanding the underlying factors that dictate airline ticket prices. The goal is to provide more informed recommendations to customers and to refine internal marketing and sales strategies.

1.2 Key Questions to Answer

To address the business problem, the following key questions will guide our analysis:

1. How do flight prices vary across different airlines? Are some airlines consistently more expensive or affordable than others?
2. Does the number of stops (direct, one stop, two or more stops) significantly affect the flight price? If so, what is the typical price difference?
3. Is there a direct relationship or correlation between the duration of a flight and its price? Do longer flights necessarily mean higher costs?
4. How does the "days left" until departure influence the ticket price? Is there an optimal booking window for the lowest prices?
5. Are there specific departure or arrival times (e.g., Early Morning, Evening, Night) that consistently correlate with higher or lower flight prices?

2 Prepare

The data set I am using can be accessed and used under a Open Database Contents License.

Source: <https://www.kaggle.com/datasets/rohitgrewal/airlines-flights-data>.

2.1 Data Integrity

The dataset achieved a perfect 10.00 usability score on Kaggle, with over 10,000 downloads. This could be a strong indicator of its technical quality, and this score suggests the data is well-structured, easy to use, and likely free of common technical issues like missing values or incompatible file formats.

However, we must understand that high usability does not guarantee data integrity. The score reflects how convenient the data is to work with, not its factual accuracy or freedom from bias.

2.2 Data Security and Privacy

The dataset did not contain any personally identifiable information (PII), which means that the data was already anonymized. Therefore we do not need to make additional steps for privacy protection.

2.3 Plan for data cleaning

Despite the high usability score, I wanted to double check the cleanliness of the dataset.

- As per usual the first step is to look for NA values.
- Secondly we need a method to identify and remove duplicate rows, to ensure each observations stays unique.
- Once these steps are done I intent to look for human errors such as inconsistent spellings or formats.

3 Process

In this phase we have completed the steps mentioned previously in the data cleaning plan. Additionally, I have double checked if each column has the correct data types assigned to them.

4 Analyze

4.1 Q1 - How do flight prices vary across different airlines?

My analysis began by calculating the average flight price for each airline. The results, visualized in the first figure, showed a significant inconsistency in the average flight prices, as Vistara and Air India appeared to be substantially more expensive than the other four carriers.

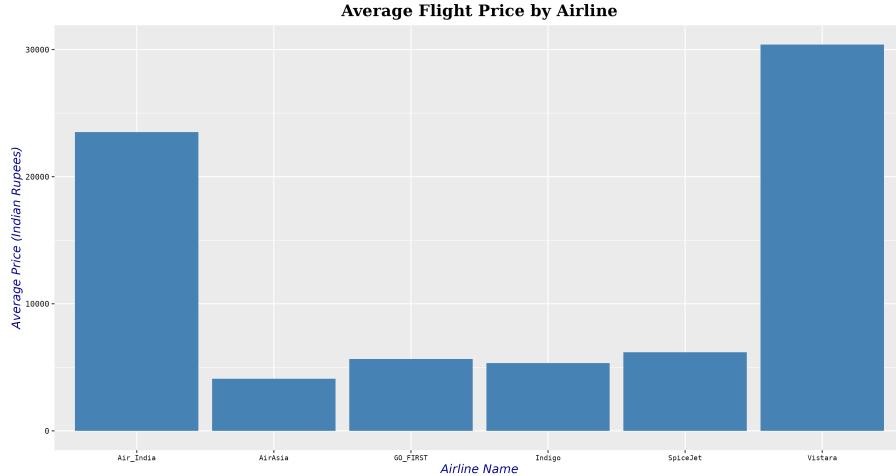


Figure 1:

However, a closer look at this preliminary finding suggested it might be misleading. These averages could be heavily influenced by factors like the class of travel and the number of stops. For example, we must consider that Vistara and Air India offer expensive Business Class tickets, which naturally drives up their overall average price compared to airlines that offer only Economy class.

To create a more accurate comparison, I have created a new sub-set to perform a controlled analysis. My idea was to filter for only direct flights in Economy class. This approach allowed me to compare airlines under the exact same conditions, providing a more reliable insight into their true pricing differences.

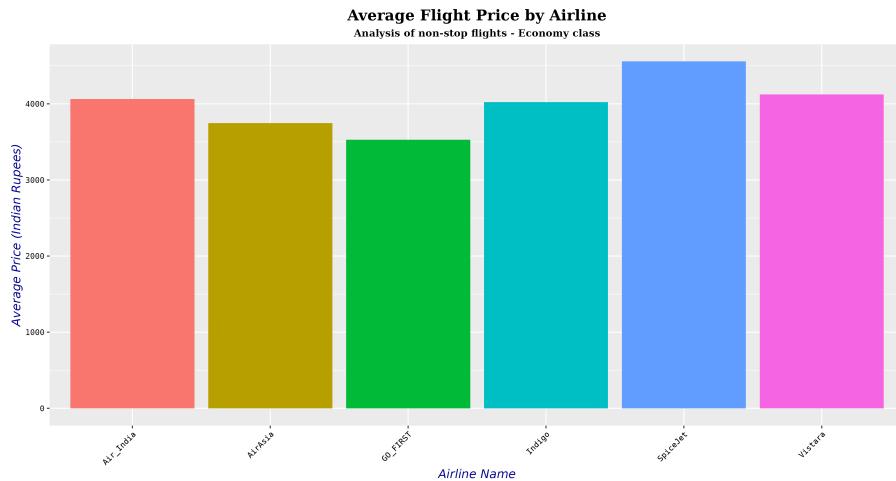


Figure 2:

After analyzing this specific subset of flights, it's clear that on average, SpiceJet offers the highest prices while GO FIRST offers the lowest. The price difference between them is approximately 1000 Rupees.

4.2 Q2 - Does the number of stops (direct, one stop, two or more stops) significantly affect the flight price? If so, what is the typical price difference?

To investigate the impact of stops on flight prices, I once again focused exclusively on economy class flights. This approach allows for a consistent comparison, as it removes the price inflation associated with business class and provides a larger, more reliable sample size for our analysis.



Figure 3: Enter Caption

This finding suggests that the added complexity and longer duration of multi-stop flights often correlate with higher prices. Conversely, direct flights offer a simple, efficient option at the lowest average cost. The price difference between these categories is significant, it means around 2 000 rupees additional cost with each stop.

4.3 Q3 - Is there a direct relationship or correlation between the duration of a flight and its price? Do longer flights necessarily mean higher costs?

My first attempt was to create a scatter plot visualising the relationship between duration and price.

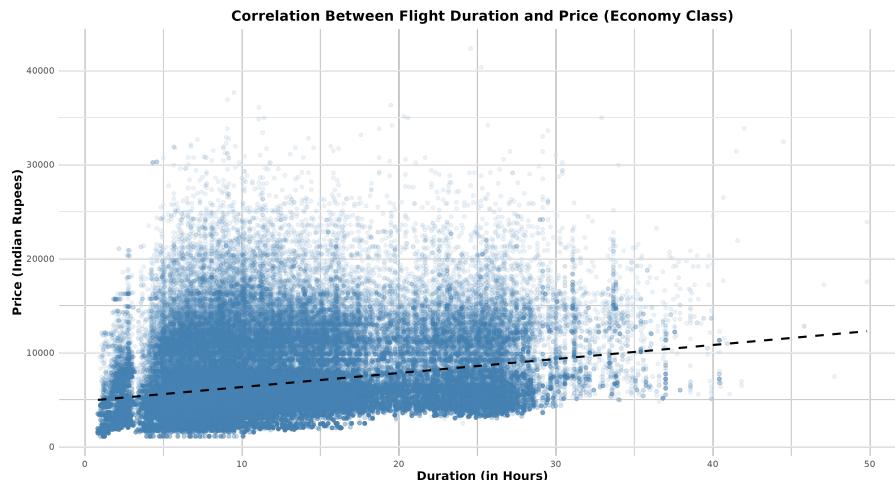


Figure 4:

As the prices vary over a wide range, it is difficult to see the relationship between the two variables. To address this, I rounded the flight durations and grouped the data by this new metric. Having done so, I calculated an average price for each category.

With the help of `geom_smooth()` I have used a linear ($y \sim x$) model to describe the relation between

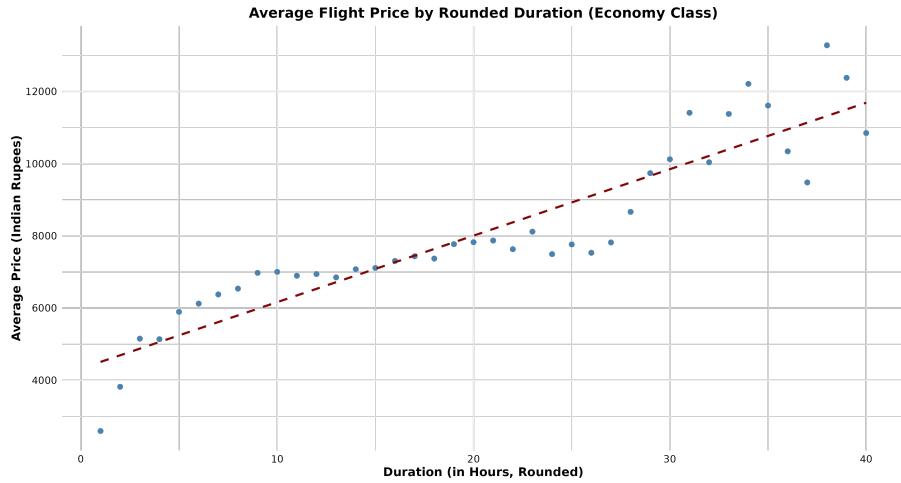


Figure 5:

flight prices and flight duration. By using this trick I have significantly reduced the number of data points while kept the information lying behind them.

While the linear model seems a good choice it is important to mention that we can clearly separate 4 sections of the price change:

- 0 - 10 hours: The price goes from under 500 up to 7 000 rupees
- 11 - 27 hours: The price always almost between 7 000 and 8 000 rupees
- 28 - 31 hours: We can see a drastic increase again up to around 11 500 rupees.
- 32 - hours: The price fluctuates in the $11\,000 \pm 2\,000$ price range.

4.4 Q4 - How does the "days left" until departure influence the ticket price? Is there an optimal booking window for the lowest prices?

To answer this question, I grouped the dataset by 'days left' and calculated the average price for each group.

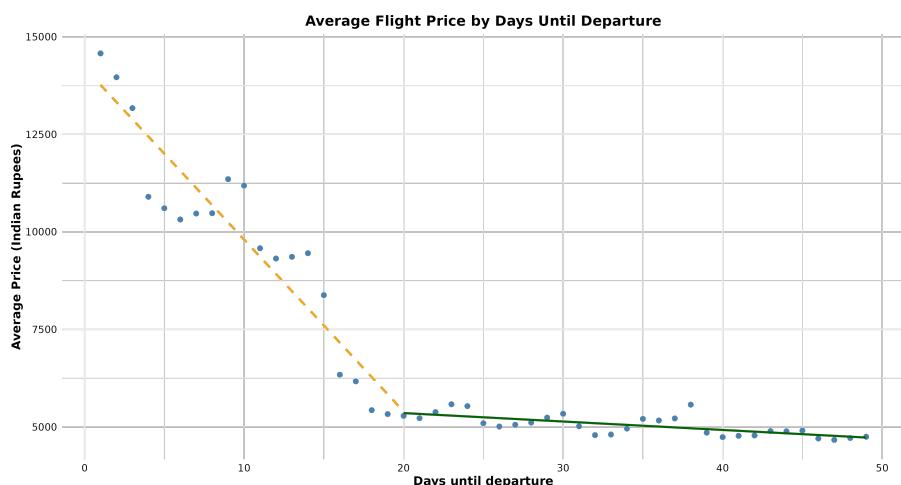


Figure 6:

While creating the visualisation for this task, I realised that it could not be described by a single linear model. The reason behind this is that flight ticket prices increase drastically once we are within 23–20 days of the flight's departure.

It is also worth mentioning that we can clearly see the weekly fluctuation in ticket prices by examining the miniature sinus-like waves.

4.5 Q5 - Are there specific departure times (e.g., Early Morning, Evening, Night) that correlate with higher or lower flight prices?

I took a simple approach to examining the problem. I grouped the flight data by departure time and calculated an average price for each category.

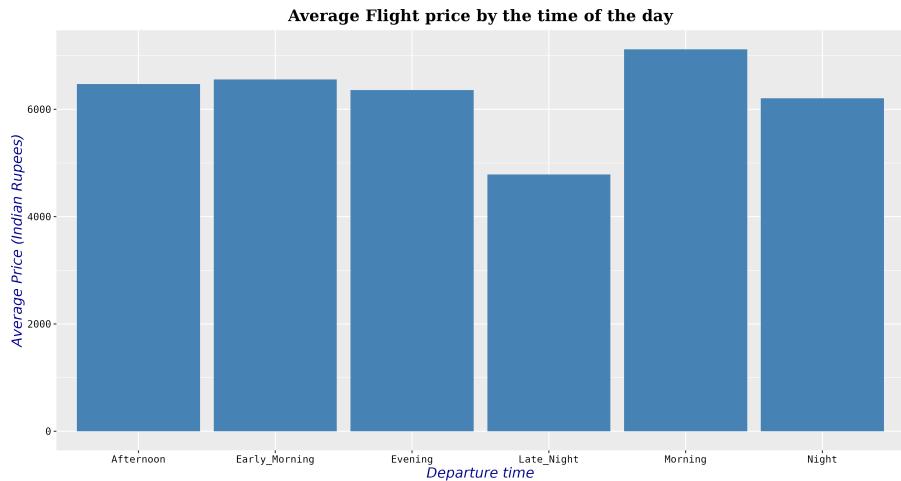


Figure 7:

As the figure clearly shows, late-night departures are the cheapest, which seems logical to me, given that nobody wants to arrive at their destination in the middle of the night.

5 Share

In this section I have summarized the key insights from my analysis into a clear and concise summary for the travel agency's stakeholders. My findings highlight the most influential factors driving airline ticket prices and offer valuable intelligence for enhancing customer recommendations and business strategy.

5.1 Pricing Drivers and Discrepancies

The initial analysis revealed a significant price difference among airlines, making Vistara and Air India appeared substantially more expensive on average. This initial finding, however, was misleading, as it was heavily influenced by the inclusion of expensive Business Class tickets offered by these carriers.

A controlled analysis focusing on direct economy class flights provided a more accurate picture. In this standardised scenario, SpiceJet offered the highest average prices, while GO FIRST was the most affordable, with an average price difference of 1,000 rupees between the two airlines.

5.2 The Impact of Stops and Duration

I have found out that the number of stops has a direct correlation with flight prices. Flights with two or more stops were the most expensive, followed by one-stop flights. Direct flights were the most affordable. There was a significant price difference between these categories, with an additional cost of around 2,000 rupees for each stop.

Analysis of flight duration revealed a positive linear relationship with price. This suggests that, in general, longer flights tend to be more expensive, although this relationship is not consistent for all flight durations.

5.3 Booking Window and Departure Time

It was also shown that the timing of a booking is a crucial factor. Flight prices increase drastically within 20–23 days of the flight's departure. The data also showed weekly fluctuations in prices that appeared as small, sinusoidal waves.

By analysing the departure times, I found that late-night departures are consistently the cheapest option. This is probably because late-night arrivals are less convenient for most travellers, resulting in lower demand and more affordable prices.

6 Act

Based on the analysis of the airline flight data, here are actionable recommendations for the travel agency to enhance its services and refine its business strategies.

6.1 Recommendations for Customers

- **Educate on Booking Window:** Advise customers to book flights as far in advance as possible to secure the lowest prices, as costs increase sharply within 20-23 days of departure.
- **Offer Price-Based Alternatives:** For budget-conscious travelers, recommend airlines like GO FIRST as they were found to have the lowest average prices.
- **Explain Price Factors:** When presenting flight options, highlight the impact of travel class, the number of stops, and the booking window to help customers understand price differences.
- **Suggest Non-Peak Departures:** Recommend late-night departures to customers seeking the lowest prices, as this time slot was consistently found to be the cheapest.

6.2 Recommendations for Internal Strategy

- **Optimize Marketing:** Develop targeted marketing campaigns based on these findings. For example, create "book early" promotions for holiday seasons or highlight savings on late-night flights for flexible travelers.
- **Develop a Pricing Tool:** Build an internal tool that uses these insights to predict flight prices and suggest the best booking times for specific routes. This can be a valuable asset for travel agents, allowing them to provide more informed advice.