Data Visualization(MCA573A)

Kalpana N
2347229

CIA-2 Assignment

Problem Statement:

In the airline industry, understanding the pricing trends, flight duration, and booking patterns is crucial for optimizing revenue and improving customer satisfaction. The dataset captures various aspects of flight operations such as departure and arrival times, airline classes, stops, and booking lead times. A key problem for the airline industry is to identify how booking time, airline, travel class, and the number of stops affect flight pricing.

Objective:

Data visualization is not a one-size-fits-all approach—each domain, whether retail, medical, financial, or education, presents unique challenges and opportunities for analysis. The task is to investigate and apply advanced data visualization techniques tailored to a specific domain, showcasing how each stage of the visualization process transforms raw data into actionable insights.

Dataset Overview

The dataset consists of flight-related information that can be used to analyze trends, delays, and performance.

Application of different Data visualization stages

1. Data Acquisition

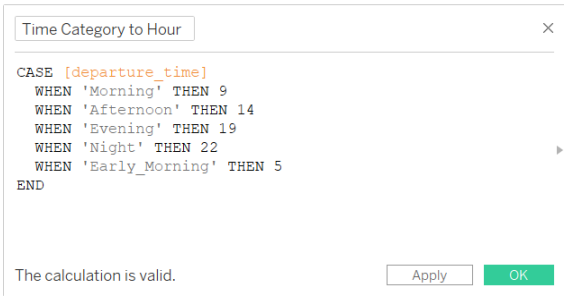
Field	Description	Usage
airline	The name of the airline operating the flight.	Compare performance metrics (e.g., price, duration) across different airlines.
flight	The flight number assigned by the airline.	Track specific flights for route analysis and performance metrics.

source_city	The city from which the flight originates.	Analyze flight patterns based on departure city, identify busiest source cities.
departure_time	The time of day the flight departs (e.g., Morning, Evening).	Study the relationship between departure times and delays, price, or duration.
stops	The number of stops (e.g., zero, one).	Analyze how the number of stops affects travel duration, pricing, and flight efficiency.
arrival_time	The time of day the flight arrives at the destination.	Analyze flight punctuality and arrival patterns based on time.
destination_city	The city to which the flight is traveling.	Track common destination routes, identify congested destinations.
class	The travel class (e.g., Economy, Business).	Compare pricing and duration between travel classes.
duration	The total flight duration (in hours).	Assess efficiency of routes and analyze factors influencing longer flight durations.
days_left	The number of days remaining until the flight's departure.	Examine booking trends based on lead time, and how price fluctuates as departure date nears.
price	The price of the flight ticket.	Study price trends across different airlines, travel classes, and booking lead times.

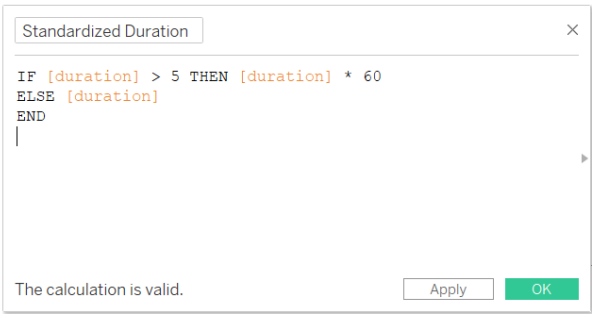
---

2. Parsing

- **Objective:** Convert and clean the raw data into a usable format for analysis. This involves standardizing categorical values and making the data easier to process.
- **Actions:**
    - **Convert Time Categories (e.g., "Morning," "Evening") into Numerical Ranges**



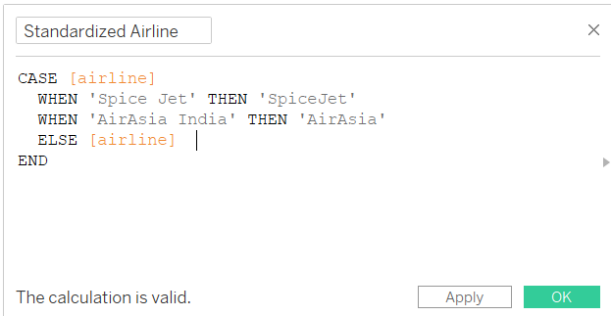
- **Standardize Flight Duration**



- **Transform "Stops" into Numerical Data**

- **Parse Categorical Data (e.g., Airline, Class) for Consistency**

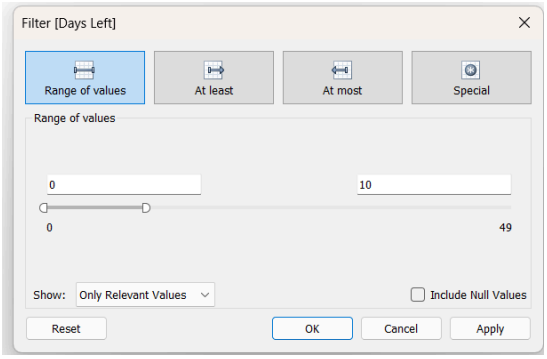
This is important because inconsistencies in categorical data can lead to incorrect or confusing visualizations. For example, you might have different spellings or variations for the same category (e.g., "SpiceJet" vs. "Spice Jet").



---

3. Filtering

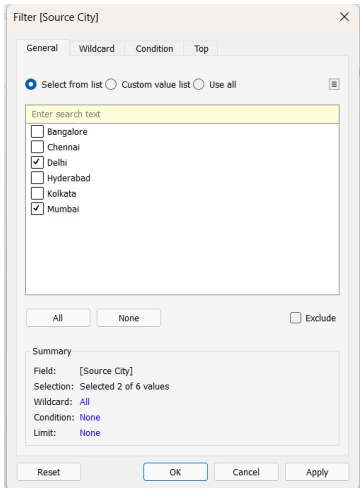
- **Objective:** Remove irrelevant or redundant data, focusing on the key variables that help solve the problem.
- **Actions:**
    - **Filter Flights Based on Days Left (e.g., Last 10 Days)**



    - **Remove Unnecessary Columns (e.g., Flight Number)**

      Eg: If you don't want to see the flight column in your Data Source view, right-click the field and select Hide. This removes it from the list of available fields without deleting it from the dataset.

    - Focus on analyzing specific cities (e.g., major hubs like Delhi and Mumbai) or airlines that are critical to the problem statement.

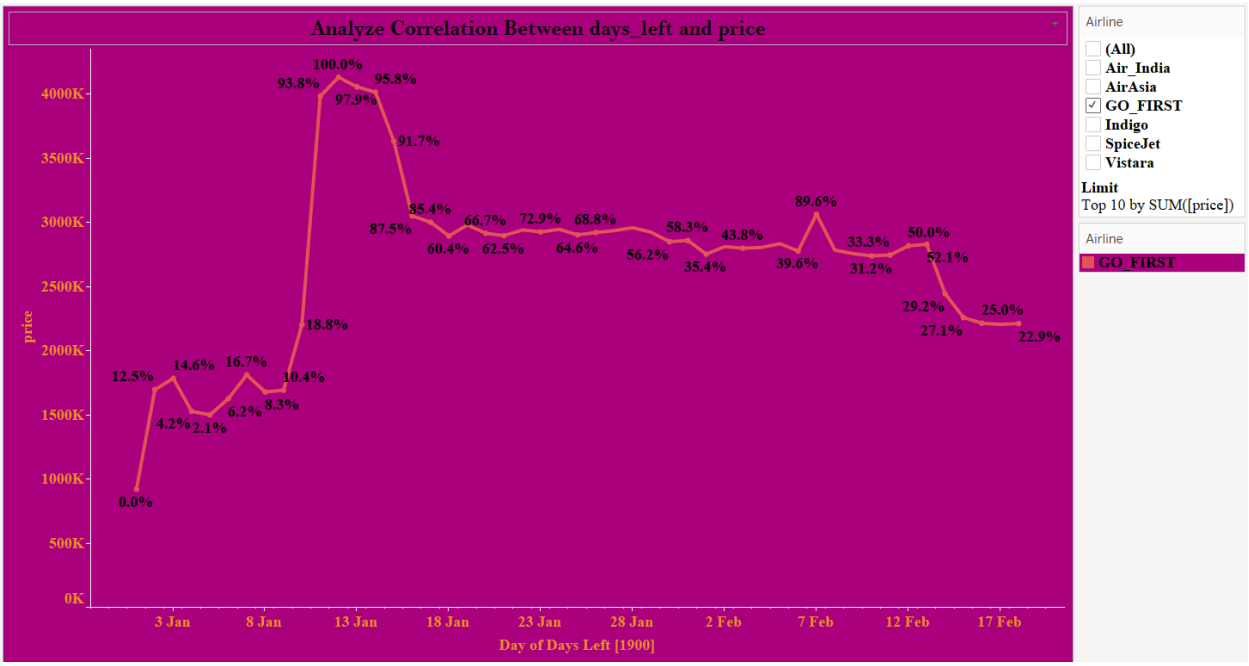


- Filter out classes (e.g., only analyze economy-class tickets if the problem focuses on budget travel trends).

---

4. Data Mining

- **Objective:** Discover hidden patterns and trends in the dataset to gain insights.
- **Actions:**
  - Use clustering techniques to group flights based on similar characteristics such as price, duration, or stops.
  - **Analyze the Correlation between days_left and price**



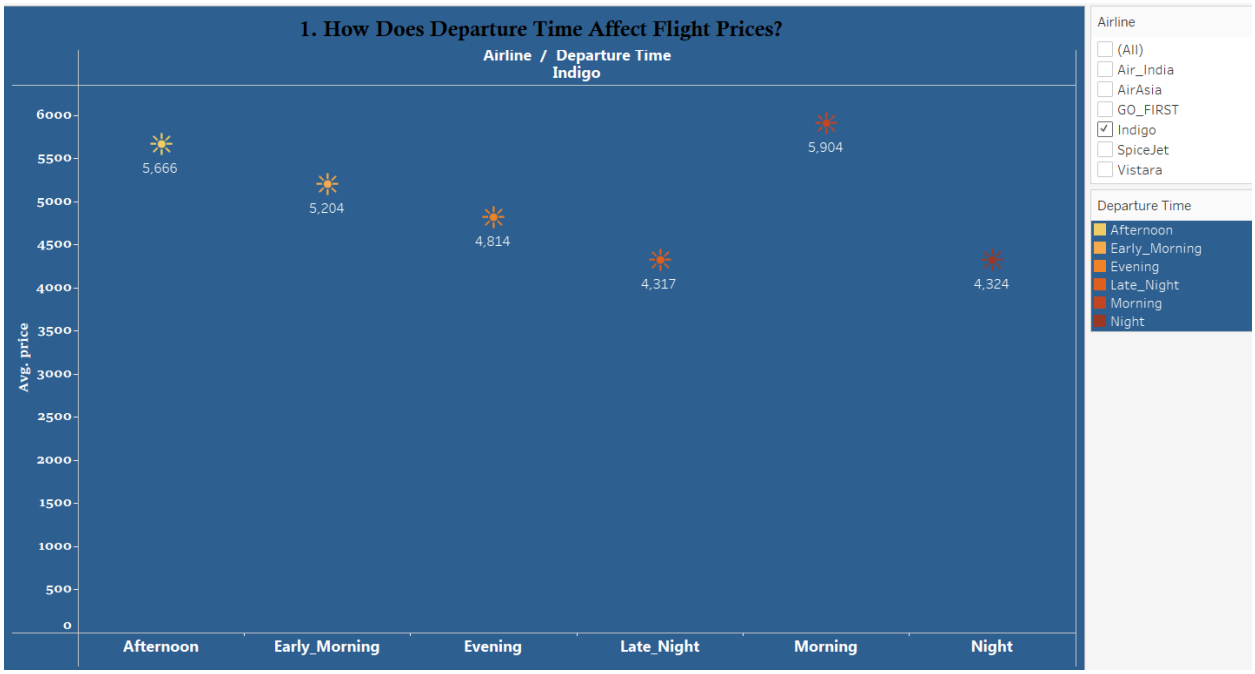
- Investigate if certain times of day (e.g., Early Morning, Afternoon) are more likely to have lower prices or shorter durations.
- Identify which airlines offer the most affordable flights while maintaining shorter durations.

---

5. Representation

- **Objective:** Display the mined data using appropriate visualizations to communicate insights clearly.

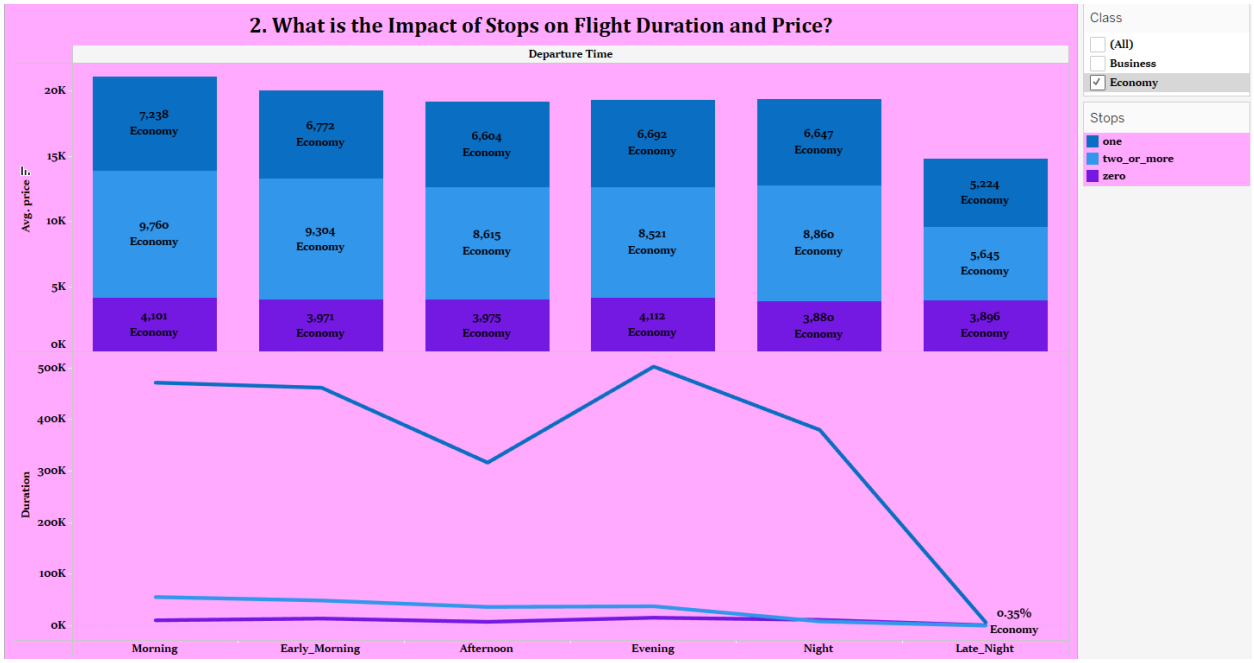
1. How Does Departure Time Affect Flight Prices?



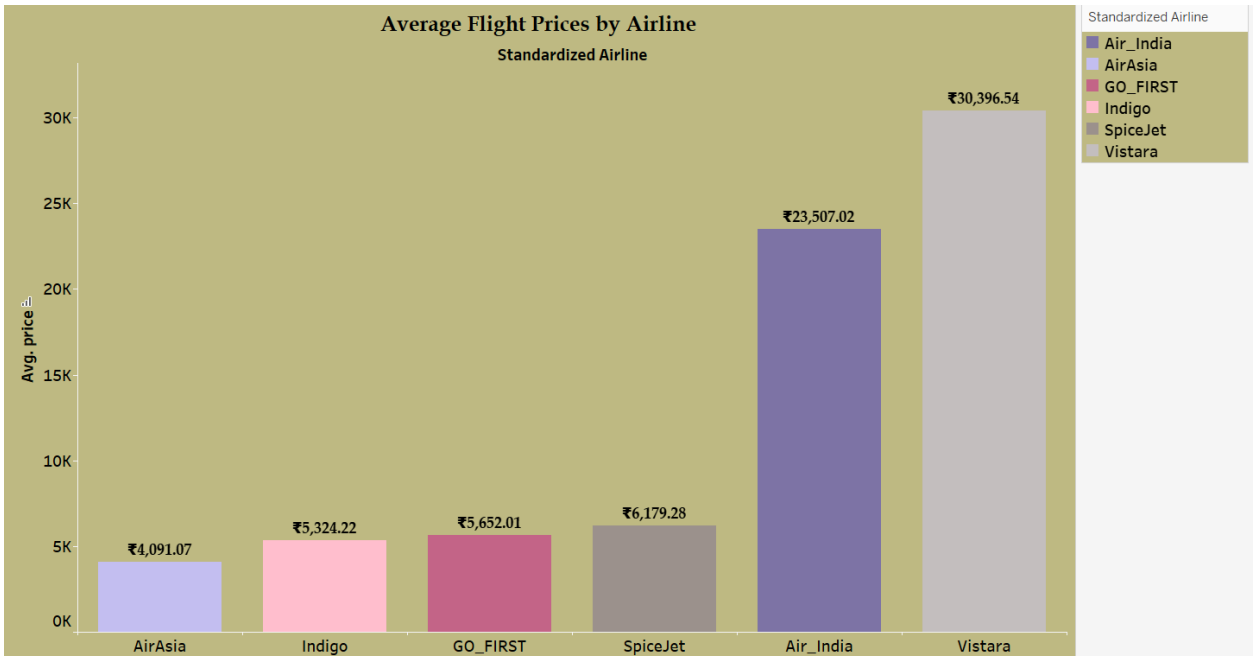
Interpretation:

Morning flights with Indigo are the most expensive, averaging ₹5,901. Late-night flights with Indigo are the cheapest at ₹4,237. Flight prices with Indigo vary significantly depending on the time of day.

2. What is the Impact of Stops on Flight Duration and Price?



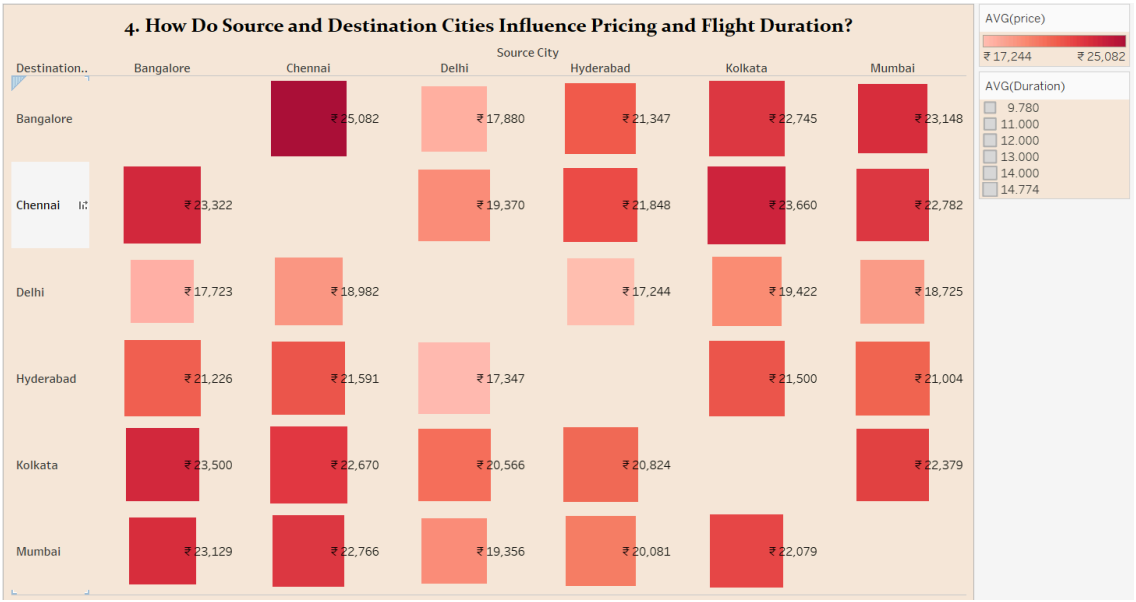
3. Which Airlines Offer the Most Competitive Prices?



Interpretation:

The bar chart compares the average ticket prices across different airlines, clearly highlighting pricing strategies. This visualization allows consumers to easily identify which airlines offer more competitive fares, enabling informed decision-making when booking flights.

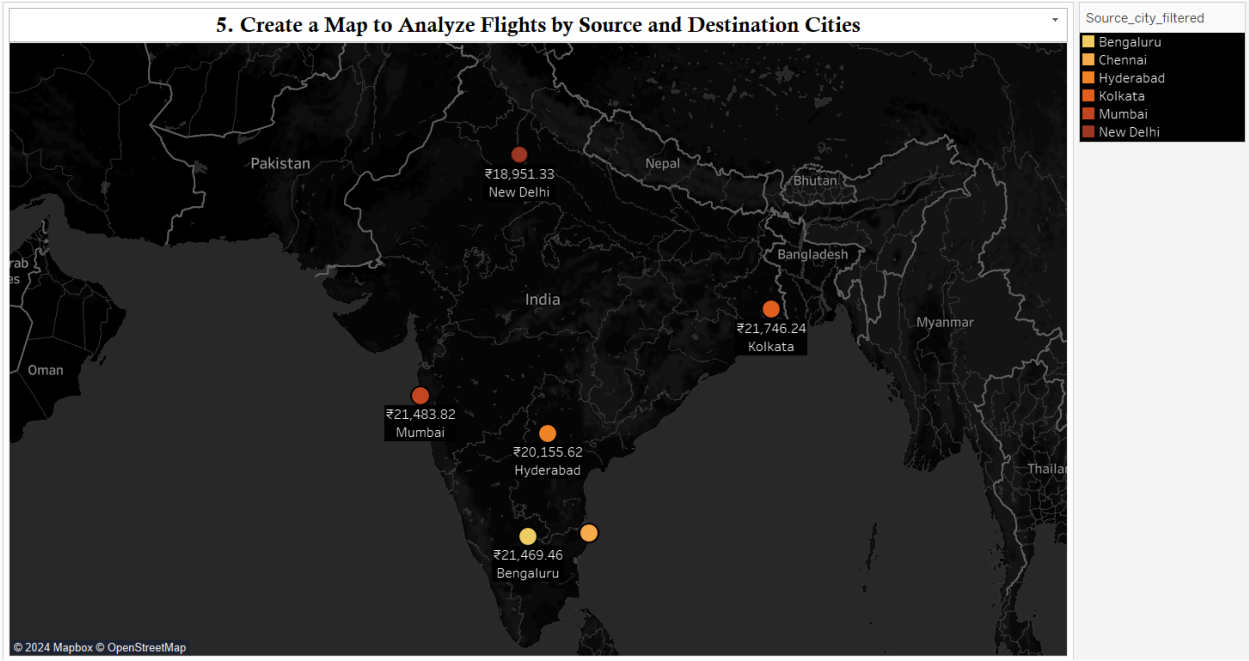
4. How Do Source and Destination Cities Influence Pricing and Flight Duration?



Interpretation :

The heat map illustrates the average flight prices and durations across various source and destination city pairs. This visualization helps identify lucrative routes and assess market competitiveness, providing insights into how geographic factors influence pricing and travel times.

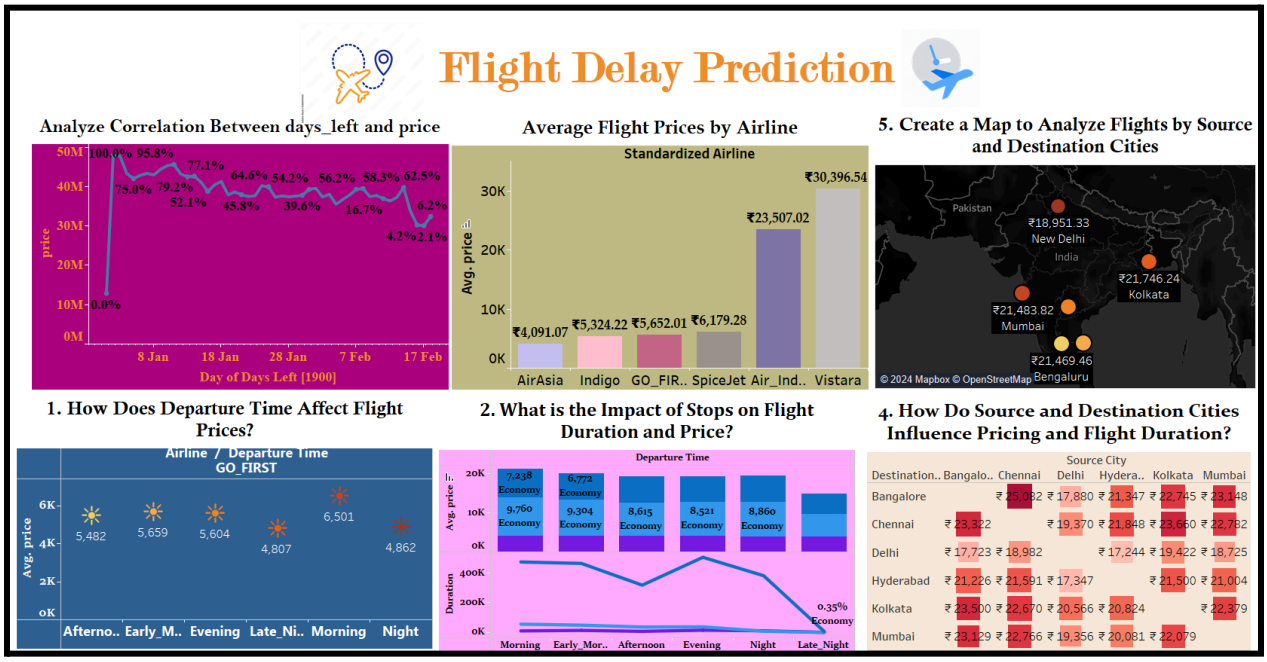
5. Create a Map to Analyze Flights by Source and Destination Cities



Interpretation :

The map visualizes average flight prices originating from various cities, highlighting pricing trends geographically. Darker colors indicate higher average prices, allowing stakeholders to identify lucrative markets and understand pricing dynamics based on geography.

Dashboard



The interactive dashboard provides a comprehensive overview of flight data, allowing users to explore pricing trends, flight durations, and other key metrics effectively. By incorporating filters and visualizations, stakeholders can make informed decisions based on their specific interests and quickly identify patterns in the airline industry.

Challenges in Flight Data Analysis Dashboard

1. Data Volume
2. Data Integration
3. Dynamic Pricing Models
4. User Engagement
5. Geographic Variability

6. Refinement

Refinement is vital in flight data analysis to present insights clearly and effectively for quick decision-making in the competitive airline industry. Simplifying designs and removing clutter help highlight critical metrics, while clear labels and consistent colors enhance readability. Interactive elements enable users to explore data based on their interests, and gathering feedback ensures visualizations remain relevant and user-friendly, ultimately supporting better decisions.

---

7. Interaction

Incorporating interaction into the flight data analysis dashboard enhances user engagement and allows stakeholders to explore data dynamically. Features such as filters for airlines and cities enable focused analysis, while drill-down capabilities provide detailed insights into specific data points. Hover tooltips offer immediate access to relevant metrics without cluttering visualizations, and interactive legends allow users to customize their view. Overall, effective interaction empowers users to identify trends and make informed decisions based on their specific interests.

Conclusion

This flight data analysis project delivers insights into pricing trends and flight durations in the airline industry through visualizations and interactive dashboards. By examining factors like departure time, stops, and city influence, stakeholders can make informed decisions on pricing strategies. The incorporation of interactivity enhances user engagement, allowing for tailored exploration of the data. Ultimately, this project empowers airlines and travelers to optimize booking strategies and improve operational efficiency.