Project Scoping Document

Name: Youssef

Analyzing Global Airline Operations

Business Problem :

The airline industry is a critical component of the global transportation network, enabling billions of passengers and goods to move across countries each year. However, operational efficiency in aviation is often hampered by unpredictable delays, cancellations, and other disruptions. These issues not only lead to financial losses for airlines but also cause dissatisfaction among passengers.

This project seeks to answer the following question:

Which factors are most correlated with flight duration, pricing, and delays, and how can we identify patterns that help optimize airline operations and customer satisfaction?

Understanding these dynamics is crucial for airlines aiming to improve their reliability, pricing strategies, and operational planning.

Business Impact :

- Provide actionable insights into which flight variables most affect pricing and travel duration.
- Support airlines in optimizing routes and schedules to minimize delays and maximize efficiency.
- Offer passengers a better understanding of how factors like ticket class, stops, or timing impact cost and duration.
- Help build a predictive model for pricing or delay classification based on user-defined inputs (airline, date, class, etc.).

Data :

Primary Dataset:

Airline Dataset – Kaggle

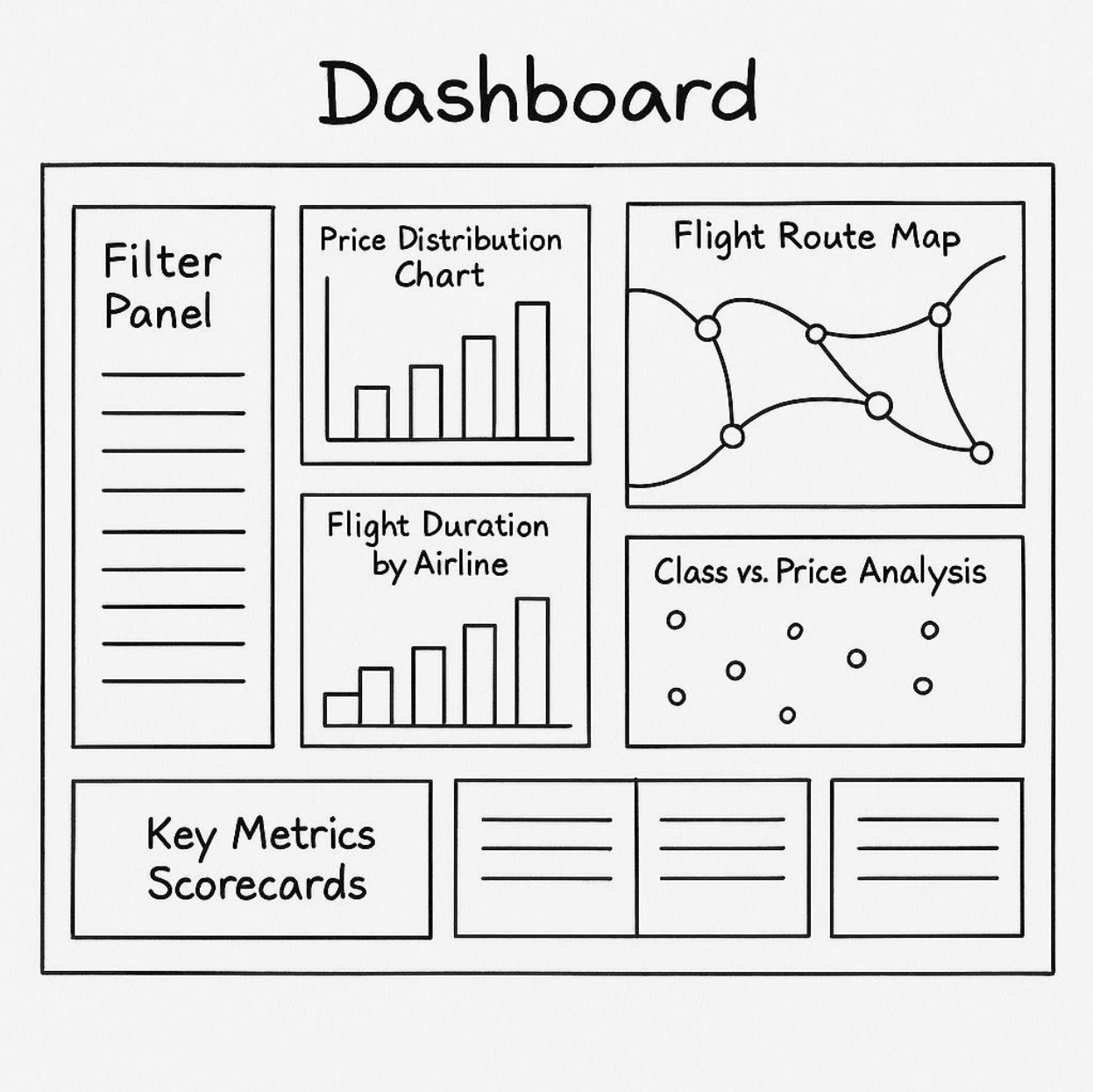
Key Columns:

- Airline: Carrier name
- Date_of_Journey: Flight date
- Source & Destination: Departure/arrival cities
- Dep_Time, Arrival_Time: Time details
- Duration: Total travel time
- Total_Stops: Number of stops
- Additional_Info: Comments or special attributes
- Price: Ticket cost
- Class: Business/Economy
- *(Optional: Add delay data if available from a secondary source)*

Potential Weaknesses:

- No delay or cancellation field in this dataset (can be added from external APIs).
- No weather data included (optional enhancement if time permits).

Dashboard



Methods

Variables:

- **Independent Variables:** Airline, Class, Total Stops, Source, Destination, Time of Day, Date, Route
- **Target Variables:** Duration, Price
- *(Optional: Add Delay as a target if using an additional dataset)*

Analytical Techniques:

- Data Cleaning & Feature Engineering (date parsing, time blocks, flight categories)
- Exploratory Data Analysis (correlations, distributions, group comparisons)
- Predictive Modeling:
    - **Regression Model** for predicting Price
    - **Classification Model** for predicting long vs. short duration flights
- Data Visualization (interactive dashboards for stakeholders)

Milestones

Milestone	Description
1. Define Scope & Objectives	Finalize business problem, understand dataset structure
2. EDA & Data Cleaning	Analyze key variables, clean date/time fields, transform categorical features
3. Feature Engineering	Create new columns (e.g., day part, flight type) for better modeling
4. Modeling	Train regression/classification models on pricing and duration
5. Dashboard Design	Build visual tools showing airline performance, pricing behavior, etc.
6. Final Report	Write summary of insights, limitations, and recommendations

Timeline

Week	Tasks
Week 1	Explore dataset, define objectives, clean data, initial EDA
Week 2	Feature engineering, test basic visualizations, start modeling
Week 3	Improve model accuracy, finalize visualizations and dashboard
Week 4 *(if needed)*	Polish results, draft final report, QA review