Business Intelligence Project

Flight Delay Analysis
Wings of Data Project
IT300 - Business Intelligence
December 14, 2025

1 1 Industry and Organization Description

Industry: Aviation & Air Transportation

The aviation industry is a highly complex, time-sensitive, and data-intensive sector where operational efficiency and punctuality are critical. Flight delays directly impact customer satisfaction, airline costs, airport congestion, and brand reputation. Even small inefficiencies can cascade across networks, affecting thousands of passengers and multiple stakeholders.

Organization Context

This use case represents a commercial airline operating a large domestic flight network. The airline collaborates with multiple airports and operates hundreds of daily flights. Management seeks to leverage historical flight data to better understand delay patterns and improve operational planning.

The dataset used (Wings of Data) simulates real operational data that an airline or aviation authority would analyze to enhance performance.

---

2 2 Business Problem

Flight delays remain a persistent operational challenge. Despite existing scheduling and buffer strategies, the organization experiences:

- **Inconsistent on-time performance**
- Higher operational costs due to delays
- Passenger dissatisfaction and reputational risk

The core problem is the lack of clear, data-driven understanding of when, where, and why delays occur. Without this insight, management decisions rely on assumptions rather than evidence.

3 3 Analytical Questions (10)

1. For each airline, how many flights are operated and how do their average departure and arrival delays compare across carriers?
2. How does the proportion of flights in each delay band (on-time, short delay, moderate delay, severe delay) differ by time of day, and are certain periods concentrated in higher delay segments?
3. How concentrated is flight activity among a small number of airlines or airports (for example, what percentage of all flights comes from the top 5 airlines or top 5 airports), and does this indicate operational dependency on specific players?
4. Which airports have the highest share of severely delayed flights, and which are mostly concentrated in on-time or low-delay segments?
5. What is the distribution of flight distances (short-haul, medium-haul, long-haul) by daily band and airline?
6. Are there noticeable temporal gaps in performance, such as specific days of the week or months with significantly higher delays compared to others at similar traffic volumes?
7. How do departure delays translate into arrival delays across different airlines, and are some carriers better at absorbing or recovering from initial delays?
8. Does flight frequency on a given route correlate with higher delay rates, indicating congestion or scheduling pressure?
9. How do delay patterns differ between origin airports and destination airports, and are certain airports consistently acting as delay propagation hubs?
10. Are there structural inefficiencies in the network, such as routes or airports with relatively few flights but disproportionately high average delays?

---

4 4 Key Performance Indicators (KPIs)

KPI 1: Total Flights

Meaning: Overall size of the flight dataset.

$$\text{TotalFlights} = |\{\text{Flight ID}\}|$$

KPI 2: Flights per Airline

Meaning: Operational scale of each airline a.

$$\text{FlightsPerAirline}(a) = |\{\text{FlightID} \mid \text{Airline} = a\}|$$

KPI 3: Average Departure Delay by Airline

Meaning: Typical departure punctuality for airline a.

$$\text{AvgDepDelay}(a) = \frac{1}{n_a} \sum_{i=1}^{n_a} \text{DepartureDelay}_i$$

where n_a is the number of flights for airline a.

KPI 4: Average Arrival Delay by Airline

Meaning: End-to-end delay experienced by passengers for airline a.

$$\text{AvgArrDelay}(a) = \frac{1}{n_a} \sum_{i=1}^{n_a} \text{ArrivalDelay}_i$$

KPI 5: Delay Rate

Meaning: Proportion of flights delayed beyond the acceptable threshold.

$$\text{DelayRate} = \frac{|\{\text{FlightID} \mid \text{DepartureDelay} > 0\}|}{\text{TotalFlights}}$$

KPI 6: Delay Band Distribution

Meaning: Share of flights in each delay category (on-time, short, moderate, severe).

For each delay band b:

$$\%\text{FlightsInBand}(b) = \frac{|\{\text{FlightID} \mid \text{DelayBand} = b\}|}{\text{TotalFlights}} \times 100$$

KPI 7: Airport Delay Intensity

Meaning: Measures average delay at specific airports.

For airport p:

$$\text{AvgDelayAtAirport}(p) = \frac{1}{n_p} \sum_{i=1}^{n_p} \text{DepartureDelay}_i$$

where n_p is the number of flights at airport p.

KPI 8: Peak-Time Delay Index

Meaning: Measures how much delays increase during peak hours.

$$\text{PeakDelayIndex} = \frac{\text{AvgDelay}_{\text{peak}}}{\text{AvgDelay}_{\text{off-peak}}}$$

KPI 9: Delay Propagation Ratio

Meaning: Degree to which departure delays convert into arrival delays.

$$\text{PropagationRatio} = \frac{\text{AvgArrivalDelay}}{\text{AvgDepartureDelay}}$$

KPI 10: Concentration Ratio (Top 5 Airlines / Airports)

Meaning: Dependency on a small number of operational entities.

For top 5 airlines:

$$\text{CR}_{\text{airlines}} = \frac{\text{Flights from Top 5 Airlines}}{\text{TotalFlights}}$$

For top 5 airports:

$$\text{CR}_{\text{airports}} = \frac{\text{Flights from Top 5 Airports}}{\text{TotalFlights}}$$