Power BI Project Presentation.

Topic: Flight Delay Analytics

Submitter by

Nandhakishor S

Table of Content.

Motivation

- Flight delays are a major operational and customer-experience challenge
- Delays are influenced by time, location, airline operations and external factors
- Time based data provides opportunities for pattern discovery
- It involves multiple stakeholders (airlines, airports, regulators)

2. Problem Statement

The objective of this project is to analyze historical airline flight data to identify when delays occur, what causes them, and how they vary across airlines, airports, and time periods, in order to support data-driven operational decisions.

3. Data Set and Sources

Source:

Maven Analytics – Airline Delay Dataset (US Flights, 2015)

Dataset overview:

The dataset contains detailed flight-level data for US airlines across an entire year, with over five million records. It includes information on departure and arrival times, delays, cancellations, diversions, taxi-in and taxi-out times, flight duration, distance, airlines, and airports.

Key characteristics:

- Large-scale (millions of rows)
- Time-based (year, month, weekday, hour)
- Mix of operational metrics and categorical attributes
- Suitable for performance, operational, and reliability analysis

4. Data Model Overview & Rationale

Modeling approach: Star Schema

The data was already provided in a partially structured format with one large transactional table and multiple smaller reference tables. I identified common fields and built a star schema to separate facts from dimensions.

Fact Table:

Flight-level operational metrics (delays, taxi times, duration, distance, cancellation/diversion flags)

Dimension Tables:

- Airline Dimension (IATA code as unique identifier)
- Airport Dimension (origin & destination airports with geographic data)
- Date/Time Dimension (year, month, weekday, hour)
- Additional airport table added for destination airport context

Data Preparation:

- Cleaned and standardized fields using Pandas
- Addressed missing and inconsistent values
- Performed transformations in Power Query
- Built relationships in Model View

Star Schema over Snowflake Schema.

- Chose Star Schema over Snowflake Schema because the dimension tables were already in normalized format.
- Simple and Faster

5. Key Insights.

Four Important Insights:

Insight 1: Flight delays follow predictable time patterns

- Delays peak during late afternoon and evening hours
- Certain weekdays consistently experience higher delays

Implication:

Delay risk can be anticipated, not treated as random.

Insight 2: Longer scheduled flights show higher unpredictability

- As scheduled flight duration increases, arrival delays become more scattered
- Both extreme delays and early arrivals increase for longer duration flights

Implication:

Fixed buffers are insufficient for long-duration flights.

Insight 3: Ground operations significantly impact delays

- Certain airports consistently show higher taxi-out times
- Delays are not purely airline-driven; airport congestion matters

Implication:

Operational inefficiencies at airports contribute directly to delays.

Insight 4: Weather is the dominant cause of cancellations

- Weather-related reasons account for the majority of cancellations
- Strong seasonal patterns are visible

Implication:

Many cancellations are predictable, not unexpected.

6. Recommendations for Stakeholders

Recommendation 1: Time-aware scheduling

- Airlines should adjust schedules and buffer times during historically high-delay hours instead of using uniform scheduling across the day.
- Based on: Hour × weekday delay patterns

Recommendation 2: Dynamic buffers for long flights

- Longer scheduled flights should have flexible turnaround buffers to reduce cascading delays (delay after delay).
- Based on: Scheduled time vs arrival delay analysis

Recommendation 3: Improve airport ground operations

- Airports with consistently high taxi-out times should be prioritized for operational improvements such as better gate management and traffic flow optimization.
- Based on: Airport-level taxi-out performance

Recommendation 4: Proactive weather-based planning

- Airlines should adopt proactive cancellation strategies during high-risk weather periods to minimize passenger disruption.
- Based on: Cancellation reason analysis

7. Conclusion

This project demonstrates that airline delays and cancellations are largely influenced by predictable factors such as time of day, flight duration, airport operations, and weather. By leveraging historical data and time-based analysis, airlines and airports can move from reactive problem-solving to proactive operational planning.

Improvements & Future work

- Add azure maps to display flight routes enabling to understand which geographical regions are prone to weather related disruptions and compare with existing data
- Filter through routes to understand which routes have highest number of flights, top airlines using the routes etc.
- Understand distance vs delay relation visually.
- Improve design