Introduction:

This project analyzes flight delay data from January 2024 to January 2025 across major U.S. airports. Flight delays directly impact passenger satisfaction, airline costs, and overall operational efficiency. Understanding when and where delays are most likely to occur can help improve decision-making and resource allocation within the aviation industry.

The primary goal of this analysis is to uncover patterns in delay frequency, identify airports with the highest delay percentages, explore common causes of delays, and detect any notable trends over time. By using SQL for data exploration and Tableau for visualization, this project delivers data-driven insights that can support improvements in airline performance and customer experience.

Data Preparation:

The dataset was sourced from the U.S. Bureau of Transportation Statistics and includes detailed flight delay data from January 2024 to January 2025 across major U.S. airports. I imported the CSV file into an SQLite database using DB Browser for SQLite to perform structured queries.

As part of the setup, I renamed the columns for clarity, ensuring they accurately reflected the data they contained (e.g., total flights, delayed flights, delay causes). I also verified the data types and checked for any formatting or missing value issues to ensure clean and accurate analysis.

SQL Analysis:

To better understand delay trends and operational performance, I used SQL to explore key questions within the flight delay dataset. The queries below were designed to uncover insights into delay frequency, contributing factors, and performance across time and locations, which are crucial for improving airline operations.

**Q1: What percentage of total flights were delayed?**

To get a baseline understanding of operational performance, I calculated the percentage of flights that experienced any delays. The total number of flights to these major airports was 819,010, and the total number of delayed flights was 197,376. Using a quick proportion, we can discover that between January 2024 and January 2025, 24.1% of all flights experienced delays.

SQL Query + Result:

SELECT SUM(arr\_flights) AS total\_flights,

SUM(arr\_del15) AS total\_delayed\_flights

FROM flight\_delays

A screenshot of a computer

AI-generated content may be incorrect.

ANSWER- 24.1% of total flights delayed

**Q2: What are the top 5 airports with the highest percentage of flights delayed?**

Seasonality often plays a significant role in airline delays, so I decided to analyze delay percentages by month. By calculating the ratio of delayed flights to total flights each month, I was able to identify trends over time. The data indicated that the summer months, particularly July and August, experienced the most delays. This was likely due to increased travel numbers and weather-related issues. In contrast, January and February experienced the lowest number of delays.

SQL Query + Result:

SELECT

airport,

airport\_name,

SUM(arr\_flights) AS total\_flights,

SUM(arr\_del15) AS delayed\_flights,

ROUND(SUM(arr\_del15) \* 100.0 / SUM(arr\_flights), 2) AS delay\_percentage

FROM flight\_delays

GROUP BY airport, airport\_name

ORDER BY delay\_percentage DESC

LIMIT 5;

A screenshot of a computer

AI-generated content may be incorrect.

Answer:

1. EWR

2. SFO

3. IAD

4. BWI

5. BNA

**Q3: How did delay rates vary by month?**

To understand where delays were most common, I calculated the number of delayed arrival flights by destination airport. This allowed me to identify which airports had the highest absolute number of delays. Major hubs like Chicago O’Hare (ORD) and Dallas/Fort Worth (DFW) appeared near the top of the list, likely due to their high traffic volumes.

**SQL Query + Result:**

SELECT

year,

month,

SUM(arr\_flights) AS total\_flights,

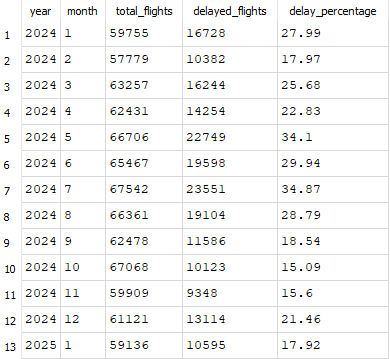
SUM(arr\_del15) AS delayed\_flights,

ROUND(SUM(arr\_del15) \* 100.0 / SUM(arr\_flights), 2) AS delay\_percentage

FROM flight\_delays

GROUP BY year, month

ORDER BY year, month;



**Tableau Visualization:**A graph on a white background

AI-generated content may be incorrect.

**Q4: What was the most common cause of delays for American Airlines?**

Rather than looking at the total minutes delayed, I focused on the number of flights delayed due to each specific cause. The dataset provides columns indicating whether a delay was attributed to carrier, weather, NAS (National Aviation System), security, or late aircraft. By counting how many flights had a delay greater than zero in each category, I was able to determine which causes were most frequent. The results showed that late aircraft and NAS delays were the most common, while security delays were the least frequent.

SQL Query + Result:

SELECT

ROUND(SUM(carrier\_ct), 0) AS carrier\_delays,

ROUND(SUM(weather\_ct), 0) AS weather\_delays,

ROUND(SUM(nas\_ct), 0) AS nas\_delays,

ROUND(SUM(security\_ct), 0) AS security\_delays,

ROUND(SUM(late\_aircraft\_ct), 0) AS late\_aircraft\_delays

FROM flight\_delays;



Tableau Visualizations:A graph with different colored squares

AI-generated content may be incorrect.

Conclusion:

This analysis provided a deeper understanding of flight delays within the airline industry, focusing specifically on arrival delays between January 2024 and January 2025. Through SQL queries, we uncovered that nearly a quarter of all flights in the dataset experienced delays, with certain months and airports standing out as particularly problematic. Late aircraft and NAS-related issues emerged as the most frequent causes of these delays, highlighting systemic challenges that go beyond weather conditions or security.

These insights not only help identify when and where delays are most likely to occur but also offer a data-driven foundation for potential operational improvements. With this kind of analysis, airline stakeholders can better prioritize interventions, allocate resources, and monitor the impact of changes over time.