Flight Data Analysis Project Report

Project Title:

U.S. Domestic Flights Data Analysis

Prepared by:

SHIVANSH DIXIT

Date:

2-07-2025

1. Project Overview

This project analyzes a comprehensive dataset of U.S. domestic flights to uncover operational insights related to delays, cancellations, and airline performance. The primary objective was to generate actionable metrics and summary reports that can help improve scheduling, resource allocation, and customer satisfaction.

2. Dataset Description

Table: Flights

Records: Approximately 5.8 million rows

Key Columns:

- AIRLINE Carrier code (e.g., UA, AA)
- FLIGHT_NUMBER Unique flight identifier
- ORIGIN_AIRPORT Departure airport code
- **DESTINATION_AIRPORT** Arrival airport code
- DEPARTURE_DELAY Departure delay in minutes
- ARRIVAL_DELAY Arrival delay in minutes
- **CANCELLED** 1 if the flight was cancelled, 0 otherwise
- CANCELLATION_REASON Reason code for cancellation
- TAXI_OUT Minutes spent taxiing before takeoff
- YEAR / MONTH Date fields

3. Analysis Queries and Results

Below are all queries performed, their purpose, and a description of what each uncovered. Full CSV exports of results are attached separately.

3.1 Total Records

SQL Query:

sql

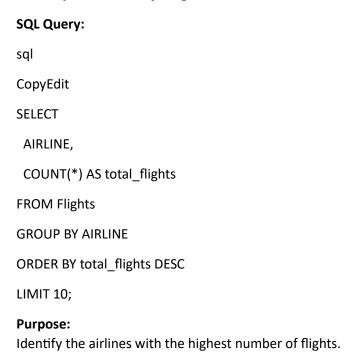
CopyEdit

SELECT COUNT(*) FROM Flights;

Purpose:

Confirm dataset size.

3.2 Top Airlines by Flight Volume



3.3 Average Arrival Delay per Airline

SQL Query:

sql

CopyEdit

SELECT

AIRLINE,

ROUND(AVG(ARRIVAL_DELAY), 2) AS avg_arrival_delay

FROM Flights

WHERE CANCELLED = 0

GROUP BY AIRLINE

ORDER BY avg_arrival_delay DESC;

Purpose:

Compare average arrival delays between airlines.

3.4 On-Time Performance

sql CopyEdit SELECT AIRLINE, COUNT(*) AS total_flights, SUM(CASE WHEN ARRIVAL_DELAY <= 0 THEN 1 ELSE 0 END) AS on_time_flights, ROUND(100.0 * SUM(CASE WHEN ARRIVAL_DELAY <= 0 THEN 1 ELSE 0 END) / COUNT(*), 2) AS on_time_percentage FROM Flights WHERE CANCELLED = 0 GROUP BY AIRLINE ORDER BY on_time_percentage DESC;

3.5 Top Departure Airports

Measure the percentage of flights arriving on time.

SQL Query:

Purpose:

sql

CopyEdit

SELECT

ORIGIN_AIRPORT,

COUNT(*) AS departures

FROM Flights

GROUP BY ORIGIN_AIRPORT

ORDER BY departures DESC

LIMIT 10;

Purpose:

Find the busiest departure airports.

3.6 Monthly Delay Trends

sql CopyEdit SELECT YEAR, MONTH, ROUND(AVG(DEPARTURE_DELAY),2) AS avg_departure_delay FROM Flights WHERE CANCELLED = 0 GROUP BY YEAR, MONTH ORDER BY YEAR, MONTH;

3.7 Cancellations by Reason

Observe seasonal patterns in departure delays.

SQL Query:

Purpose:

sql

CopyEdit

SELECT

CANCELLATION_REASON,

COUNT(*) AS cancellations

FROM Flights

WHERE CANCELLED = 1

GROUP BY CANCELLATION_REASON

ORDER BY cancellations DESC;

Purpose:

Analyze reasons for cancellations.

3.8 Longest Delays

sql CopyEdit SELECT AIRLINE, FLIGHT_NUMBER, ORIGIN_AIRPORT, DESTINATION_AIRPORT, DEPARTURE_DELAY, ARRIVAL_DELAY FROM Flights ORDER BY ARRIVAL_DELAY DESC LIMIT 10;

3.9 Longest Taxi-Out Times

Identify flights with the longest delays.

SQL Query:

Purpose:

sql

CopyEdit

SELECT

AIRLINE,

FLIGHT_NUMBER,

ORIGIN_AIRPORT,

DESTINATION_AIRPORT,

TAXI_OUT

FROM Flights

ORDER BY TAXI_OUT DESC

LIMIT 10;

Purpose:

Spot operational bottlenecks on the ground.

4. Key Insights

- Several airlines have significantly higher average arrival delays.
- Specific airports manage extremely high flight volumes and experience congestion.
- Cancellations are most commonly caused by weather and carrier-related issues.
- Seasonal peaks in delays align with major holidays and weather disruptions.

5. Recommendations

- Address taxi-out delays at high-traffic airports by reviewing gate scheduling and runway procedures.
- Investigate frequent carrier-related cancellations to identify systemic operational issues.
- Consider adjusting flight schedules and buffer times during peak months.

6. Attachments

- CSV exports of all queries
- Screenshots of query executions
- Visual charts in Excel or Power BI