SQL + Power BI Project

Project Title: U.S. Airline Performance & Delay Analysis

Objective:

This project focuses on analysing large scale flight performance data to uncover patterns and insights that can help airlines and stakeholders improve operations.

Using tools like SQL for data cleaning and transformation, and Power BI for interactive visualizations, this project delivers a comprehensive dashboard that tracks key performance indicators (KPIs) such as total number of flights, average departure delay, average arrival delay, and cancellation rates. The data is explored across multiple dimensions including airline, airport, date, and delay type.

Problem Statements:

- Total cancellation flight, number of flights cancelled by weather, security, air system etc
- Total number of flights diverted.
- Departure delay and arrival delay statistics (min, max)
- Percentage of flights arriving within 15mins of the schedule.
- Flights cancellation rate
- Total number of flights, number of flights cancelled, number of flights diverted by each airline
- Total number of flights, number of flights cancelled, number of flights diverted by each airport
- Total number of flights, number of cancelled flights, number of diverted flights in each month.
- Average arrival delay by flight date
- Average departure delay by different airline
- Flight cancellation by different months

Dataset Summary:

Dataset contains 3 csv files, **airlines.csv**, **airports.csv** and **flights.csv**.

1. Airlines.csv has **15 rows** and **2 columns**, columns are:
    - **Iata_code:** Airline Identifier
    - **Airline:** Airline's Name

2. Airports.csv has **323 rows** and **7 columns**, columns are:
    - **Iata_code:** Airport Identifier
    - **Airport:** Airport's Name
    - **City:** City Name of the Airport
    - **State:** State of the Airport

- **Country:** Country Name of the Airport
- **Latitude:** Latitude of the Airport
- **Longitude:** Longitude of the Airport

3. Flights.csv has **1million+ rows** and **31 columns**, columns are:
    - **Year:** Year of the Flight Trip
    - **Month:** Month of the Flight Trip
    - **Day:** Day of the Flight Trip
    - **Day_of_week:** Day of week of the Flight Trip
    - **Airline:** Airline Identifier
    - **Flight_number:** Flight Identifier
    - **Tail_number:** Aircraft Identifier
    - **Origin_airport:** Starting Airport
    - **Destination_airport**: Destination Airport
    - **Scheduled_departure**: Planned Departure Time
    - **Departure_time**: WHEEL_OFF – TAXI_OUT
    - **Departure_delay**: Total delay on departure
    - **Taxi_out**: The time duration elapsed between departure from the origin airport gate and wheel off
    - **Wheels_off:** The time point that the aircraft's wheel lave the ground
    - **Scheduled_time:** Planned time amount needed for the flight trip
    - **Elapsed_time:** AIR_TIME + TAXI_IN + TAXI_OUT
    - **Air_time:** The time duration between wheel off and wheel_on time
    - **Distance:** Distance between two airports
    - **Wheels_on:** The time point that the aircraft's wheels touch on the ground
    - **Taxi_in:** The time duration elapsed between wheels-on and gate arrival at the destination airport
    - **Scheduled_arrival:** Planned arrival time
    - **Arrival_time:** WHEELS_ON+TAXI_IN
    - **Arrival_delay:** ARRIVAL_TIME-SCHEDULED_ARRIVAL
    - **Diverted:** Aircraft landed on airport that out of schedule
    - **Cancelled:** Flight Cancelled (1 = cancelled)
    - **Cancellation_reason:** Reason for Cancellation of flight: A - Airline/Carrier; B - Weather; C - National Air System; D - Security
    - **Air_system_delay:** Delay caused by air system
    - **Security_delay:** Delay caused by security
    - **Airline_delay:** Delay caused by the airline
    - **Late_aircraft_delay:** Delay caused by aircraft
    - **Weather_delay:** Delay caused by weather

Data Preparation (SQL work)

Steps Taken:

- Handle NULL values in Cancellation_reason, Air_system_delay, Security_delay, Airline_delay, Late_aircraft_delay, Weather_delay columns
- Joined 3 tables using INNER joins
- Aggregated flights cancelled, flights diverted by airline, by airport and month using GROUP BY
- Created calculated fields: scheduled_departure_ts, departure_time_ts, scheduled_arrival_ts, arrival_time_ts, departure_delay_time, cancellation_reason_desc, flight_date

```sql
UPDATE flights
SET air_system_delay = 0
WHERE air_system_delay IS NULL;
```

Fill the NULL value with 0 in air_system_delay column

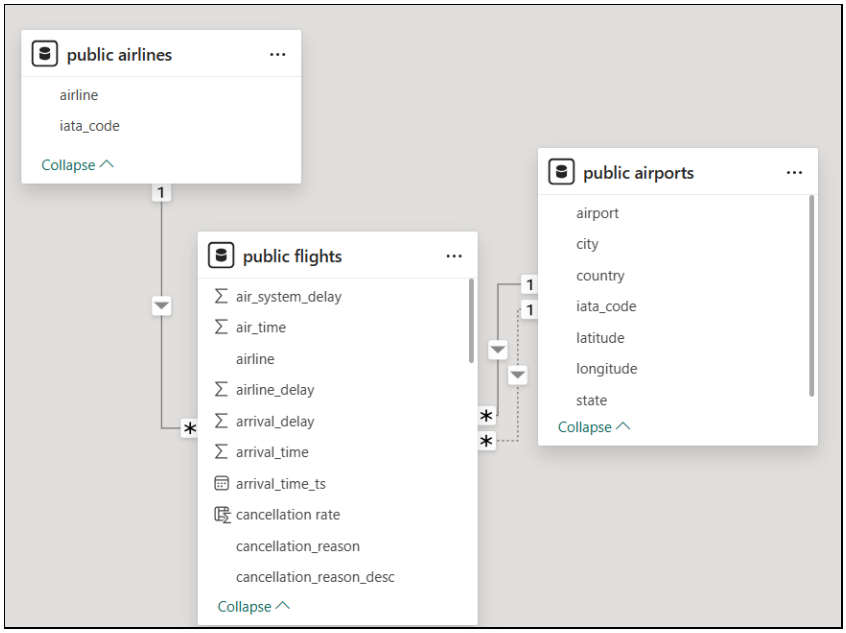
```sql
ALTER TABLE flights
ADD COLUMN cancellation_reason_desc TEXT;

UPDATE flights
SET cancellation_reason_desc =
    CASE cancellation_reason
        WHEN 'A' THEN 'Airline/Carrier'
        WHEN 'B' THEN 'Weather'
        WHEN 'C' THEN 'National Air System'
        WHEN 'D' THEN 'Security'
        ELSE 'Not Cancelled'
    END;
```

Create new column cancellation_reason_desc (based on cancellation_reason column)

```sql
SELECT  a.airline,
        COUNT(f.flight_number) AS number_of_flights,
        COUNT(f.cancelled) FILTER(WHERE f.cancelled = 1) AS flights_cancelled,
        COUNT(f.diverted) FILTER(WHERE f.diverted = 1) AS diverted_flights
FROM airlines AS a
JOIN flights AS f
ON a.iata_code = f.airline
GROUP BY a.airline
ORDER BY number_of_flights DESC
LIMIT 5;
```

Cancelled flights, diverted flights by airline

Power BI Data Model



Data Model Structure:

- **1 Fact Table**: flights (flight trip data)
- **2 Dimension Tables**: airlines, airports
- **Relationships:** Primary Key-Foreign key links (iata_code, flight_number, origin_airport, destination_airport)
- **DAX Measures:**
    - cancellation rate = DIVIDE(CALCULATE(COUNTROWS('public flights'), 'public flights'[cancelled] = 1), COUNTROWS('public flights'), 0)
    - On-Time Arrival Rate % = DIVIDE(CALCULATE(COUNTROWS('public flights'), 'public flights'[arrival_delay] <= 15), COUNTROWS('public flights'), 0)
    - Diverted Flight Rate % = DIVIDE(SUM('public flights'[diverted]), COUNTROWS('public flights'), 0)

Dashboard Overview

Section 1: Key Matrices (KPI Cards)

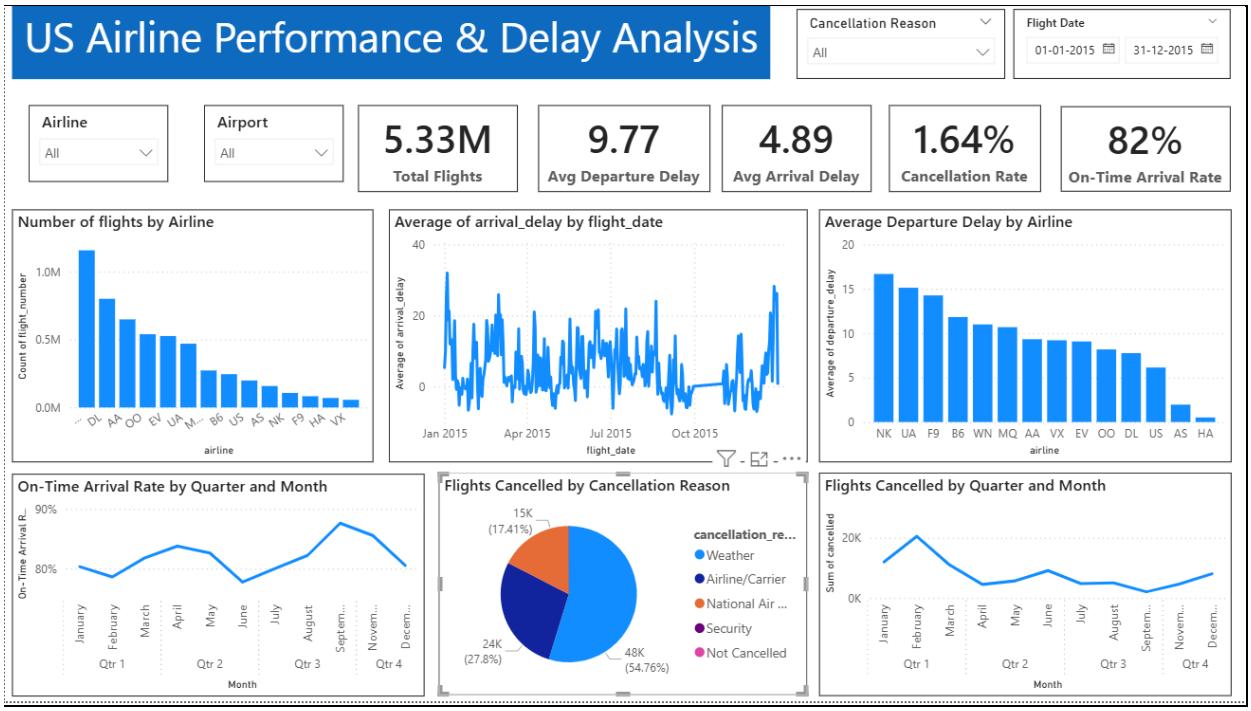
- Total Flights
- Average Departure Delay
- Average Arrival Delay
- Cancellation Rate
- On-time Arrival Rate

Section 2: Visual Analytics (Charts)

- Trend Analysis
    - Line Chart: Flights cancellation rate over months/quarters
    - Line Chart: Average delay by year/month/quarter/day
    - Line Chart: On-time arrival rate by month/quarts
- Segmentation
    - Bar Chart: Departure delay by airline
    - Pie Chart: Flight cancellation distribution by cancellation reason description
    - Bar Chart: Flight trips by different airline

Section 3: Slicer for Interactivity

- Filtering by Flight Date
- Filtering by Airline
- Filtering by Airport
- Filtering by Cancellation Reason



Key Insights & Recommendations:

1. Flight Volume

- **5.33M** total flights, with **Southwest Airlines Co. (WN)** operating the most flights – showing their dominant market share. **July** was the busiest month – indicating seasonal travel peak.
- **Recommendation:**
    - Focus operational analysis on Southwest Airlines Co. (WN) to identify best practices for high volume efficiency
    - Increase staffing, aircraft availability and maintenance schedules in summer months

2. Delays

- Average departure delay is **9.77mins and** arrival delay is **4.89mins,** delays are more common at departure. **Spirit Airlines (NK)** has the longest average departure delay**, January** has the highest average arrival delay due to **winter weather.**
- **Recommendation:**
    - **Review boarding process and pre-flight checks to reduce departure delay**
    - Investigate NK's operational bottlenecks and provide punctuality improvement strategies.
    - Implement winter weather contingency plans, including standby crews and flexible scheduling.

3. Cancellation

- cancellation rate is **1.64%,** with **ORD airport** responsible for **9% of all cancellation,** mainly in **February** due to **winter weather**
- **Recommendation:**
    - Deploy more ground staff, enhance snow/ice clearance, and communicate proactively with passengers.

4. Diversions

- Overall diversion rate is **0.27%**, with **ORD** accounting for **5%** of diversions. **Southwest Airlines Co. (WN)** has the highest number of diverted flights.
- **Recommendation:**
    - Review air traffic flow, improve runway availability, and enhance weather monitoring at ORD.
    - Evaluate route planning and in-flight decision protocols for WN to minimize diversions.

5. In-Time Performance

- On-time arrival rate is **82%**, with **Southwest Airlines Co. (WN)** leading in punctuality. **October** is the best month for on-time arrivals — likely due to favourable weather and balanced demand.
- **Recommendation:**
    - Benchmark WN's scheduling and crew management practices for other airlines.
    - Promote October travel in marketing campaigns to enhance customer experience.

Summary:

- Airlines face operational challenges from delays, cancellations, and diversions, which impact passenger satisfaction and operational costs. The objective was to identify performance bottlenecks and seasonal/airport-specific problem areas to support decision-making.
- **SQL** – Extracted, cleaned, and aggregated flight records to calculate KPIs (e.g., on-time rates, delay averages, cancellation and diversion rates).
- **Power BI** – Built interactive dashboards with slicers, charts, and KPI cards to visualize performance across airlines, airports, and months.
- **Main Insights:**
    - Total flights: 5.33M, with Southwest Airlines Co. (WN) having the highest volume.
    - Cancellations: Overall rate of 1.64%, with ORD accounting for 9% of cancellations, primarily in February due to winter weather.
    - Delays: Average departure delay was 9.77 mins; Spirit Airlines (NK) had the longest average departure delays.
    - On-time performance: Overall rate of 82%, peaking in October.
    - Diversions: Overall rate 0.27%, highest from ORD airport.