Flight Performance & Aviation Analytics for AeroStat Airlines

MySQL | Power BI | DAX | Visualization

-by SANGEETA RAJPUT

Intoduction

AeroStat Airlines operates an extensive flight network, facing challenges like delays, cancellations, and operational inefficiencies. As a Data Analyst, your role is to leverage SQL for data extraction, Power BI/Tableau for visualization, and machine learning for predictive insights.
This project focuses on analyzing flight punctuality, cancellation patterns, and route efficiency to identify areas for improvement. By building an interactive dashboard with key performance indicators and trend analysis, you will provide executives with data-driven insights to optimize scheduling, reduce delays, enhance passenger experience, and improve financial efficiency, ensuring AeroStat Airlines remains competitive.

About MySQL

- **Powerful and Open-Source:** A widely used relational database management system (RDBMS).
- **Data Storage and Management:** Used for storing and managing structured data.
- **SQL Querying:** Provides robust tools for querying, analyzing, and manipulating data through SQL (Structured Query Language).
- **Scalability:** Can handle large datasets and grow with business needs.
- **Reliability:** Known for its stable performance and consistency.
- **Efficiency:** Optimized for fast data processing and complex business queries.

PART 1: SQL QUERIES

Flight Delay Analysis

1.Calculate the average departure and arrival delay for all flights in the last 6 months.

```sql
SELECT
    ROUND(AVG(DEPARTURE_DELAY), 2) AS avg_departure_delay,
    ROUND(AVG(ARRIVAL_DELAY), 2) AS avg_arrival_delay
FROM
    flights
WHERE
    DATE(CONCAT(YEAR, '-', MONTH, '-', DAY)) >= DATE_SUB(CURDATE(), INTERVAL 6 MONTH);
```

Most Frequent Routes

2.Identify the top 10 most popular flight routes based on the total number of flights.

```sql
SELECT
    ORIGIN_AIRPORT,
    DESTINATION_AIRPORT,
    COUNT(*) AS total_flights
FROM
    flights
GROUP BY ORIGIN_AIRPORT , DESTINATION_AIRPORT
ORDER BY total_flights DESC
LIMIT 10;
```

Airline Performance Ranking

3.Rank airlines by their on-time performance (lowest average delay).

```sql
SELECT
    f.AIRLINE,
    a.AIRLINE AS airline_name,
    ROUND(AVG(f.ARRIVAL_DELAY), 2) AS avg_arrival_delay,
    ROUND(AVG(f.DEPARTURE_DELAY), 2) AS avg_departure_delay,
    RANK() OVER (ORDER BY AVG(f.ARRIVAL_DELAY) ASC) AS performance_rank
FROM flights f
JOIN airlines a ON f.AIRLINE = a.IATA_CODE
WHERE f.CANCELLED = 0
GROUP BY f.AIRLINE, a.AIRLINE
ORDER BY avg_arrival_delay ASC;
```

Cancellation Trends

4.Find the top reasons for flight cancellations and their frequency

```sql
SELECT
    c.CANCELLATION_DESCRIPTION,
    COUNT(f.CANCELLATION_REASON) AS cancellation_count
FROM
    flights f
        JOIN
    cancellation_codes c ON f.CANCELLATION_REASON = c.CANCELLATION_REASON
WHERE
    f.CANCELLED = 1
GROUP BY c.CANCELLATION_DESCRIPTION
ORDER BY cancellation_count DESC;
```

Airport Congesion Analysis

5.Identify airports with the highest number of flight departures and arrivals.

```sql
SELECT
    ORIGIN_AIRPORT AS airport,
    COUNT(*) AS total_departures
FROM flights
GROUP BY ORIGIN_AIRPORT
ORDER BY total_departures DESC
LIMIT 10;

SELECT
    DESTINATION_AIRPORT AS airport, COUNT(*) AS total_arrivals
FROM
    flights
GROUP BY DESTINATION_AIRPORT
ORDER BY total_arrivals DESC
LIMIT 10;
```

Weather related disruptions

6.Analyze how weather impacts delays and cancellations.

```sql
SELECT
    COUNT(*) AS total_flights,
    SUM(CASE
        WHEN
            CANCELLED = 1
                AND CANCELLATION_REASON = 'W'
        THEN
            1
        ELSE 0
    END) AS total_weather_cancellations,
    ROUND(SUM(WEATHER_DELAY) / COUNT(*), 2) AS avg_weather_delay,
    ROUND(SUM(WEATHER_DELAY), 2) AS total_weather_delay,
    ROUND((SUM(CASE
                WHEN
                    CANCELLED = 1
                        AND CANCELLATION_REASON = 'W'
                THEN
                    1
                ELSE 0
            END) / COUNT(*)) * 100,
            2) AS weather_cancellation_rate
FROM
    flights;
```

Flight distance duration Trends

7.Calculate the average flight duration for different distance ranges.

```sql
SELECT
    CASE
        WHEN DISTANCE <= 500 THEN '0-500 miles'
        WHEN DISTANCE <= 1000 THEN '501-1000 miles'
        WHEN DISTANCE <= 1500 THEN '1001-1500 miles'
        WHEN DISTANCE <= 2000 THEN '1501-2000 miles'
        WHEN DISTANCE <= 2500 THEN '2001-2500 miles'
        ELSE '2501+ miles'
    END AS distance_range,
    ROUND(AVG(ELAPSED_TIME), 2) AS avg_flight_duration
FROM
    flights
WHERE
    ELAPSED_TIME IS NOT NULL
GROUP BY distance_range
ORDER BY MIN(DISTANCE);
```

Day of Week Flight Performance

8.Find which day of the week has the highest on-time flight performance.

```sql
SELECT
    DAY_OF_WEEK,
    COUNT(*) AS total_flights,
    SUM(CASE
        WHEN ARRIVAL_DELAY <= 0 THEN 1
        ELSE 0
    END) AS on_time_flights,
    ROUND((SUM(CASE
                WHEN ARRIVAL_DELAY <= 0 THEN 1
                ELSE 0
            END) / COUNT(*)) * 100,
            2) AS on_time_percentage
FROM
    flights
WHERE
    CANCELLED = 0
GROUP BY DAY_OF_WEEK
ORDER BY on_time_percentage DESC;
```

Seasonality in Air Traffic

9.Identify which months have the highest number of flights and delays.

```sql
SELECT
    MONTH,
    COUNT(*) AS total_flights,
    SUM(DEPARTURE_DELAY + ARRIVAL_DELAY) AS total_delay_minutes,
    ROUND(AVG(DEPARTURE_DELAY + ARRIVAL_DELAY), 2) AS avg_delay_per_flight
FROM
    flights
WHERE
    CANCELLED = 0
GROUP BY MONTH
ORDER BY total_flights DESC;
```

Passenger Connectivity & Hub Efficency

10.Identify which airports act as the largest hubs based on the number of connecting flights.

```sql
SELECT
    f1.ORIGIN_AIRPORT AS hub_airport,
    a.AIRPORT AS airport_name,
    COUNT(*) AS total_connections
FROM
    flights f1
        JOIN
    flights f2 ON f1.TAIL_NUMBER = f2.TAIL_NUMBER
        AND f1.DESTINATION_AIRPORT = f2.ORIGIN_AIRPORT
        AND TIMESTAMPDIFF(MINUTE,
        f1.ARRIVAL_TIME,
        f2.DEPARTURE_TIME) BETWEEN 30 AND 180
        JOIN
    airports a ON f1.ORIGIN_AIRPORT = a.IATA_CODE
WHERE
    f1.CANCELLED = 0 AND f2.CANCELLED = 0
GROUP BY f1.ORIGIN_AIRPORT , a.AIRPORT
ORDER BY total_connections DESC
LIMIT 10;
```

Comprehensive Summary of SQL Analysis on Flight Operations

1.Flight Delay Analysis

- The average departure and arrival delays over the past six months were calculated.
- Only recent flight records were considered to support operational performance analysis.
- The results provide insights into delay patterns and help identify areas for improvement.

2.Top 10 Most Popular Flight Routes

- Flight routes were ranked based on the total number of flights between origin and destination airports.
- The most frequently traveled routes indicate major air corridors and key demand hubs.

3.**Airline Performance Based on Delays**

- Airlines were evaluated based on their average arrival and departure delays.
- Only non-cancelled flights were considered to ensure ranking accuracy.
- Airlines were ranked from best to worst, with the lowest average arrival delay being the best performer.

4.Flight Cancellations by Reason

- The total number of cancelled flights was analyzed.
- Cancellations were categorized using predefined cancellation codes (e.g., Weather, Security, Airline Issues).
- Results were sorted to highlight the most common reasons for flight cancellations.

5.Busiest Airports by Departures & Arrivals

- The top 10 busiest airports were identified based on the number of departing and arriving flights.
- The first query ranked airports by total departures, while the second ranked them by total arrivals.
- These findings help pinpoint key hub airports handling high flight traffic.

6.Impact of Weather on Flight Operations

- The effect of weather disruptions was assessed by calculating:
  - Total number of flights
  - Total weather-related cancellations
  - Average and total weather delays
  - Weather cancellation rate
- These insights help evaluate how significantly weather impacts flight schedules.

7.Duration by Distance Range

- Flights were categorized into distance bands (e.g., 0-500 miles, 500-1000 miles).
- The average flight duration was calculated for each range.
- This helps understand how trip length affects flight time.

8.Flight Punctuality by Day of the Week

- The total flights, on-time flights, and on-time performance rate were calculated for each day of the week.
- Results were sorted to highlight the best-performing days in terms of punctuality.

9.Monthly Flight Activity and Delays

- Monthly flight trends were analyzed by calculating:
  - Total flights per month
  - Total delay minutes
  - Average delay per flight
- The busiest months were ranked, revealing seasonal trends in air travel demand.

10. Top 10 Busiest Hub Airports (Flight Connections)

- The busiest hub airports were identified based on flight connections within 30 to 180 minutes.
- The total number of flight connections at each hub was counted.
- Results ranked hubs based on total connections, highlighting key transit points in air travel.

About Power BI :

- **DAX Queries:** Used for creating complex calculations and data manipulations.
- **Visuals:** Allows users to create a wide range of charts and visuals (e.g., bar charts, line graphs, maps) to represent data insights.
- **Filters:** Enable users to segment data and focus on specific insights for better analysis.
- **Data Modeling:** Helps structure data relationships to ensure consistency and accuracy in reports.
- **Calendar Table:** Used for time-based analysis, supporting time intelligence functions like year-over-year comparisons and trend analysis.

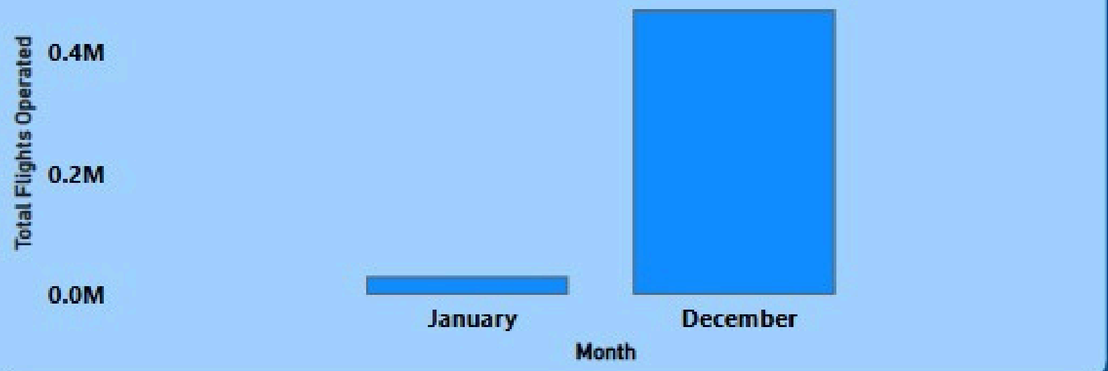
PART 2: DASHBOARD VISUALIZATION

Aerostat Airlines Dashboard

500K	61.16%	136.58	6.23	10.12
Total Flights Operated	On-Time Performance Rate	Average Flight Duration	Average Arrival Delay	Average Departure Delay

On-Time Performance Rate by AIRLINE

Total Flights Operated by Month



DAY

31 December 1899	02 January 1900	04 January 1900	06 January 1900	08 January 1900	10 January 1900	12 January 1900
01 January 1900	03 January 1900	05 January 1900	07 January 1900	09 January 1900	11 January 1900	13 January 1900

Count of FLIGHT_NUMBER by ARRIVAL_DELAY

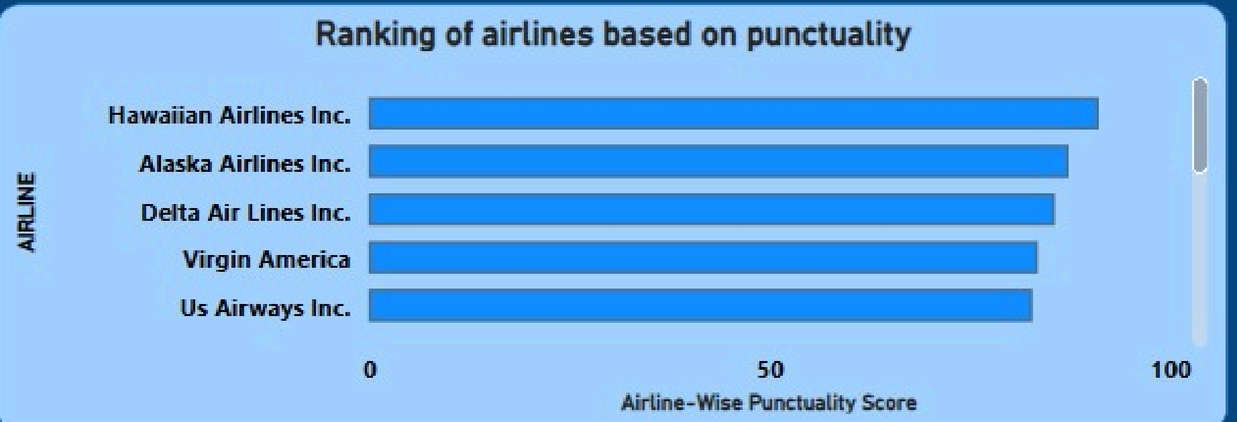
Airline Performance Comparison

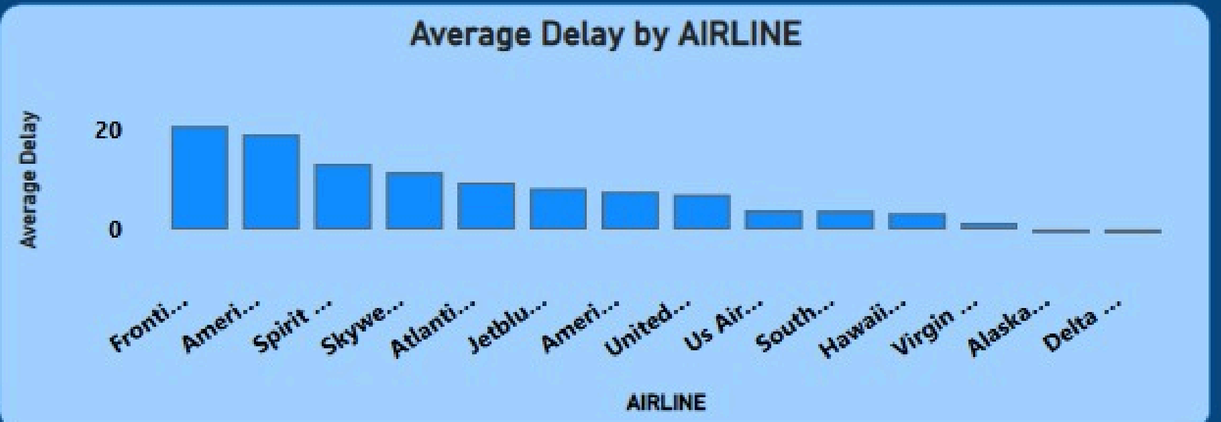
77.59
Airline-Wise Punctuality Score

3.51%
CancellationRate

6.47
Average Delay

DAY
31-12-1899
28-01-1900





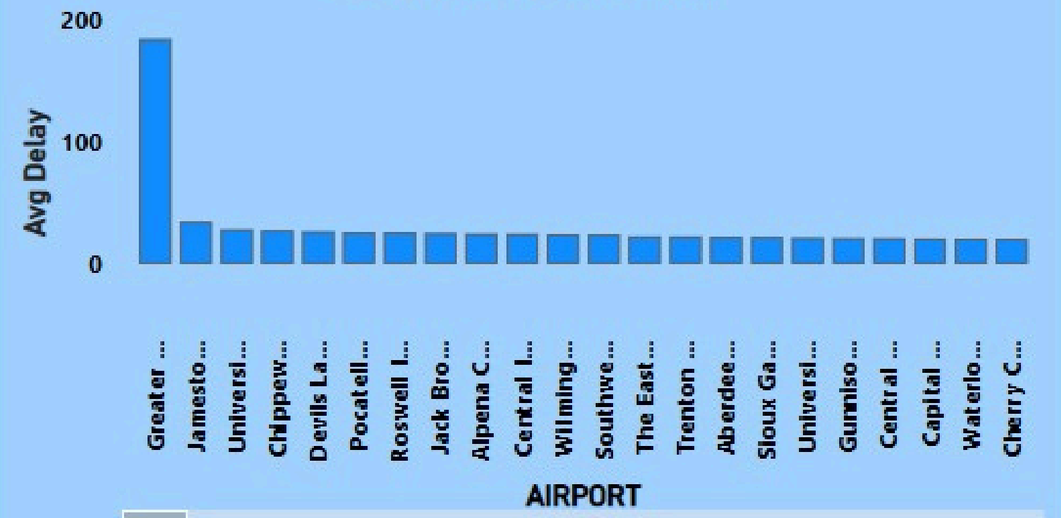
AIRLINE
- Alaska Airlines Inc.
- American Airlines Inc.
- American Eagle Airlines Inc.
- Atlantic Southeast Airlines
- Delta Air Lines Inc.
- Frontier Airlines Inc.
- Hawaiian Airlines Inc.

ORIGIN_AIRPORT
- ABE
- ABI
- ABQ
- ABR
- ABY
- ACT
- ACV

Route & Airport Analytics

ORIGIN_AIRPORT	TotalFlights
ATL	31332
DEN	18142
DFW	24651
IAH	14255
LAS	12361
LAX	18457
MCO	10717
ORD	24967
PHX	14010
SFO	13712
Total	**182604**

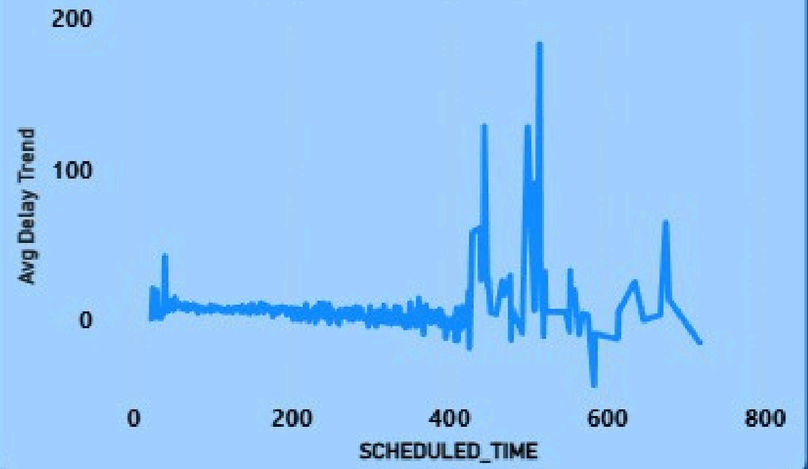
Airport Congestion Index



Heatmap of flight movements across airports

ORIGIN_AIRPORT	DESTINATION_AIRPORT	Flight Routes
ATL	ABE	80
DTW	ABE	72
ORD	ABE	20
DFW	ABI	254
ATL	ABQ	56
BWI	ABQ	33
DAL	ABQ	153
DEN	ABQ	121
DFW	ABQ	132
HOU	ABQ	95
Total		**499998**

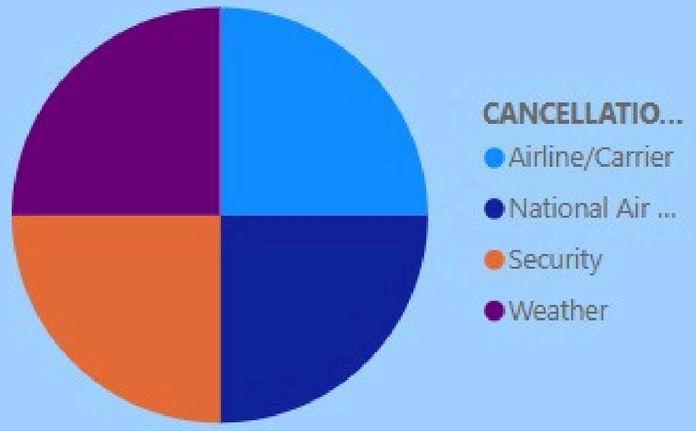
Delay trends by airport



Traffic flow between major hubs

Flight Cancellation & Delay Patterns

Cancellation Rate by Reason



340K

Weather Delay impact

39.65

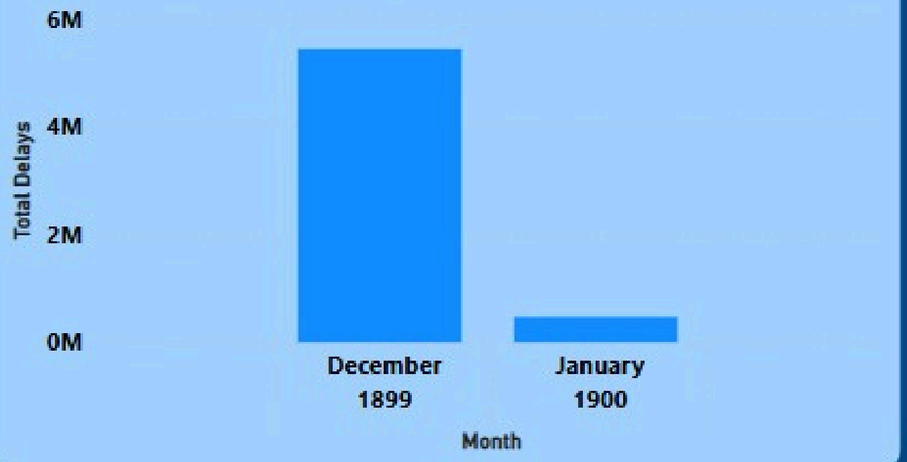
Late Aircraft Delay Contribution

Cancellation Rate by Month

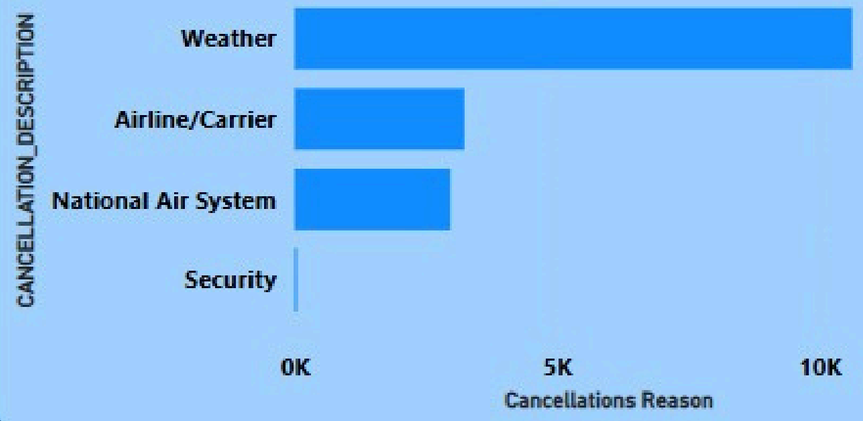
MONTH

31-12-1899 01-01-1900

Flight delay reasons analysis



Cancellation breakdown by external factors



CANCELLATION_DESCRIPTION

- [ ] (Blank)
- [ ] Airline/Carrier
- [ ] National Air System
- [ ] Security
- [ ] Weather

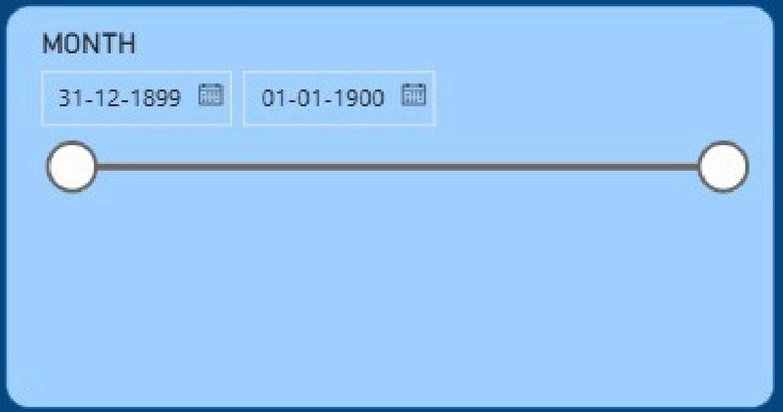
Passenger Experience & Service Quality

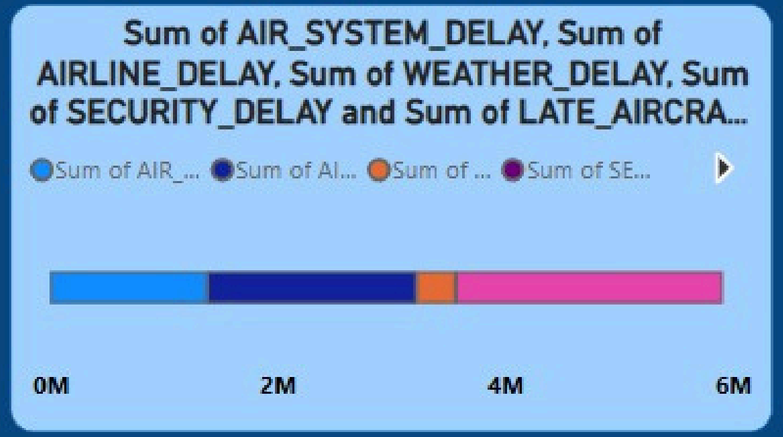
60.12
Customer_Satisfaction proxy score

12.20
Missed_Connections_Rate

38.46%
Delay Recovery Effectiveness

6.23
Avg_Arrival_Delay





Financial & Operational Efficiency

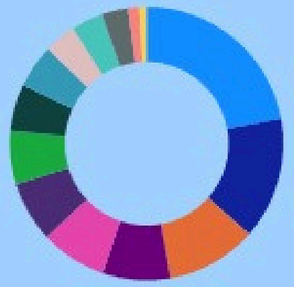
$12.2958...
Flight Cost per Mile

$175.6bn
Total Revenue Loss

1.22
Fuel Efficiency

AIRLINE
- AA
- AS
- B6
- DL
- EV
- F9
- HA
- MQ
- NK
- OO
- UA

Financial impact of cancellations and delays



AIRLINE
- MQ
- B6
- EV
- US
- UA

MONTH

31-12-1899 01-01-1900

Cost efficiency comparison across different routes

Airline fuel and operational efficiency insights

Comprehensive Summary of visualizations on Flight Operations

1. Flight Operations Overview

- Total Flights Operated: 500K
- On-Time Performance Rate: 61.16%
  - Key Delay Metrics:Average Flight Duration: 136.58 minutes
  - Average Arrival Delay: 6.23 minutes
  - Average Departure Delay: 10.12 minutes
  - Operational Insights:Flight performance trends are monitored by airline and monthly
    flight volumes to enhance punctuality.

2. Airline Performance Comparison

- Punctuality Score: 77.59%
- Cancellation Rate: 3.51%
- Average Delay: 6.47 minutes
    - Insights:Top Punctual Airlines: Hawaiian, Alaska, Delta
    - High Delay Airlines: Frontier, American
    - Cancellation Concerns: Some airlines exhibit elevated cancellation rates that need attention.

3. Route & Airport Analytics

- Busiest Airports: ATL, DEN, and DFW handle the highest flight volumes.
- Top Routes: ATL to ABE and DFW to ABI experience high traffic flow.
- Congestion Hotspots: Certain airports have significantly higher average delays.
- Delay Trends: Delays peak during rush hours.

4. Flight Cancellation & Delay Patterns

- Top Cancellation Cause: Weather contributes the most to flight cancellations.
- Delay Contributors: Late aircraft delays have a major impact on on-time performance.
- Monthly Trend: Cancellation rates decline towards the end of the year.

5. Passenger Experience & Service Quality

- Satisfaction Proxy Score: 60.12, indicating room for improvement.
- Missed Connections Rate: 12.2%, highlighting passenger inconvenience.
- Top Delay Contributors: Airlines like F9, MQ, and NK have the highest average arrival delays.

6. Financial & Operational Efficiency

- Flight Cost per Mile: $12.30, indicating relatively high operational expenses.
- Total Revenue Loss: $175.6 billion due to cancellations, delays, and inefficiencies.
- Fuel Efficiency: Average of 1.22, with VX and UA leading in operational efficiency.

Let’s Connect !!

I hope you found these insights valuable.

Feel free to connect with me on LinkedIn to explore more about my projects, share feedback, or discuss new opportunities:

Linkedin Profile

Github Profile

Or drop me a message directly at:
Email: **ranputsangeeta70@gmail.com**

Together, let’s turn data into meaningful stories!