**TECHNICAL REPORT**

*FOR*

**CASE STUDY - AIRPORT AUTHORITY DATA ANALYSIS**

**Authors: James Benjamin Cole, Long Ho Mak, Mary Grace Lunar, Sirjan Bhalla**

**Roll No: 041178602, 041172682, 041169815, 041155238**

**Course Info: 25W_CST2102_010**

**Instructor Name: Yasser Jafer**

**Submission Date: March 07, 2025**

# TABLE OF CONTENTS

# ABSTRACT

To ensure that flights depart and arrive at their scheduled time, it is very important to make sure that all airport operations are efficient and in order. Any disruption in the daily operations can cause delays in flights or other day-to-day activities at the airport. This negatively impacts the airline companies as well as the passenger experience.

Using the Airline Delays and Cancellations Dataset (Kaggle, 2018), this study analyses flight schedules and patterns across significant airports for the year 2018. The key insights in this report include flight punctuality, customer satisfaction and ground processing times. Using PowerBI as a visualization tool, interactive reports are targeted for end users such as airport administration and airline companies.

These dashboards aim to help airport authorities in optimizing their operations and scheduling to prevent significant flight delays.

Findings show that passenger dissatisfaction peaked around August, especially at SFO airport. Although airline-specific trends vary, general observation showed aircraft delays and ground processing times peaking from June to August and from November to January, which corresponds with peak travel seasons.

# INTRODUCTION

## Background:

Airports are a hub for domestic and global connectivity. Operational efficiency plays a key role in ensuring smooth flight schedules. Disruption to these can have significant consequences for both airlines and passengers. Delays can lead to increased operational costs, logistical inefficiencies, and reduced passenger satisfaction, which in turn affect an airline's reputation and profitability.

Understanding flight patterns and identifying bottlenecks in airport operations can help airport authorities optimize scheduling, streamline processes, and improve overall efficiency. By leveraging data-driven insights, decision-makers can address these issues that contribute to flight delays and passenger dissatisfaction.

## Objective:

The primary goal is to develop an interactive Power BI dashboard that provides real-time insights into flight punctuality, passenger feedback, and operational bottlenecks.

The dashboards will allow users to:

- Track the total number of flights (arrivals and departures).
- Look at flight delays and ground processing times.
- Monitor passenger feedback trends.
- Filter data based on date, flight source, and destination.

A study by Yimga (2020), suggests that delays not only disrupt airport operations but also incur monetary losses for airlines. These financial setbacks, coupled with reduced customer trust, can negatively impact an airline's public image. By providing a clear visualization of key airport metrics, this report aims to support airport authorities in optimizing operations, improving scheduling, and minimizing delays, thereby reducing financial and reputational risks for airlines.

## Overview of the process:

The methodology involves four key steps:

1. Data Storage & Preparation:

    - The dataset was imported to Microsoft SQL Server.
    - Data cleaning and transformation ensured consistency and accuracy.
    - New columns were created to categorize flight delays and customer satisfaction.

2. Database Integration with Power BI:

    - Microsoft SQL Server was connected to Power BI to create the dashboards.
    - Data relationships were defined to support drill-through functionality.

3. Dashboard Development:

    - Airport-Level Metrics: Total flights, delays, and flight trends.
    - Flight-Level Metrics: Ground processing times, average delays, and passenger satisfaction.
    - Interactive Features: Users can filter by airport, date, and flight details for deeper analysis.

4. Map Visualization:

    - A geospatial map was created using latitude and longitude data to display airport locations.
    - The map allows users to explore flight routes between source and destination airports.

# METHODOLOGY

To extract meaningful insights from the 2018 airport dataset, a structured approach was implemented using Microsoft SQL Server for database creation and Power BI for visual analysis. The objective was to examine flight punctuality, customer satisfaction, and operational bottlenecks across major airports.

## Step 1: Database Creation in Microsoft SQL Server

The dataset was loaded into Microsoft SQL Server to facilitate structured storage and efficient querying. The SQL Server Import and Export Wizard was used for this purpose. The steps included:

- Creating a new database in SQL Server.
- Selecting "Flat File Data" as the data source.
- Importing the 2018 airport dataset as a CSV file.
- Choosing the Microsoft OLE DB Provider for SQL Server as the destination to store the data.

After the import, the dataset was successfully stored in the database, ensuring optimized access for further analysis.

## Step 2: Connecting SQL Server to Power BI

With the database ready, establishing a connection between Power BI Desktop and SQL Server was next to enable visualization. The steps included:

- Selecting SQL Server as the data source in Power BI.
- Entering server and database name for authentication.
- Choosing relevant tables to load into Power BI.

Data transformation and relationship modeling were conducted within Power BI to prepare the dataset for meaningful analysis.

## Step 3: Data Cleaning and Modeling

1. **Fact Table**
    - The 2018 Airport Data Table serves as the main fact table, containing detailed flight records and all relevant foreign keys.
    - Flight ID column added to the 2018 table as a surrogate key.
    - Both DepartureTimeDisplay and ArrivalTimeDisplay columns in the 2018 table were edited to display time in the appropriate format (1:00) rather than as a number (100).
    - CustomerSatisfaction column was created using a formula that indicates a “satisfied customer” for flights that arrive 11 minutes earlier than expected and are not cancelled, “unsatisfied customer” for flights that are cancelled or delayed by 16 minutes or more, and “neutral” for flights that arrive anywhere between 10 minutes earlier than expected and 15 minutes later but are not cancelled.
    - DelayedOrNot column was created using a formula that indicates “delayed” if flight is not cancelled and arrival delay time greater than 0, or “Notdelayed” if flight is not cancelled, and arrival delay time is less than or equal to 0.

2. **Dimension Tables**
    - AirportList Table: Created by merging distinct Origin (ORIGIN) and Destination (DEST) airport codes to create Airport column.
    - AirportsInformation Table: External data from OpenFlights (2018) was used to supplement airport details, such as Name, Country and Coordinates. 1 to 1 relationship established between AirportInformation and AirportList Table using airport codes (IATA).
    - DestTable: Created by duplicating the main 2018 table and removing all other columns except Flight ID and destination.
    - OriginTable: Created by duplicating the main 2018 table and removing all other columns except Flight ID and Origin.
    - AirportAppend Table: Created by combining both Dest and Origin tables. Many to 1 relationship established between AirportAppend and 2018 table using flight ID.

- AirlineTable: Derived from OP_CARRIER column to store airline details. Airline names were retrieved from IATA's Airline Code Search (2018) to standardize carrier details. Many to 1 relationship established between 2018 table and AirlineTable.

## Step 4: Creating Power BI Dashboards

The Power BI dashboard was designed to visualize key airport-level and flight-level metrics:

1. **Airport-Level Metrics**

- Total Incoming Flights: Includes on-time and delayed arrivals (excluding cancellations).
- Total Departing Flights: Includes on-time and delayed departures (excluding cancellations).
- Delayed Incoming Flights: Arrivals delayed by ≥1 minute (excluding cancellations).
- Delayed Departing Flights: Departures delayed by ≥1 minute (excluding cancellations).
- Map: Generated from longitude and latitude columns in AirportsInformation Table. The map is filtered by location using the Airport column in AirportList Table.
- Drill-through Functionality: Enables filtering by airport and date for in-depth analysis.

2. **Flight-Level Metrics**

- Total Number of Flights: Count of flights based on surrogate key created in Flight ID in 2018 table
- Average Ground Processing Time: Average (TAXI_IN + TAXI_OUT) from 2018 table
- Average Arrival Delay: Average ARR_DELAY from 2018 table

- Average Departure Delay: Average DEP_DELAY from 2018 table
- Customer Satisfaction: Makes Use of the created CustomerSatisfaction column in 2018 table.

# RESULTS

## Delays and Ground Processing Times:

- Flights experienced longer ground processing times and higher average arrival and departure delays during June to August and November to January compared to other months. Ground processing times increased by approximately 30% during these months.

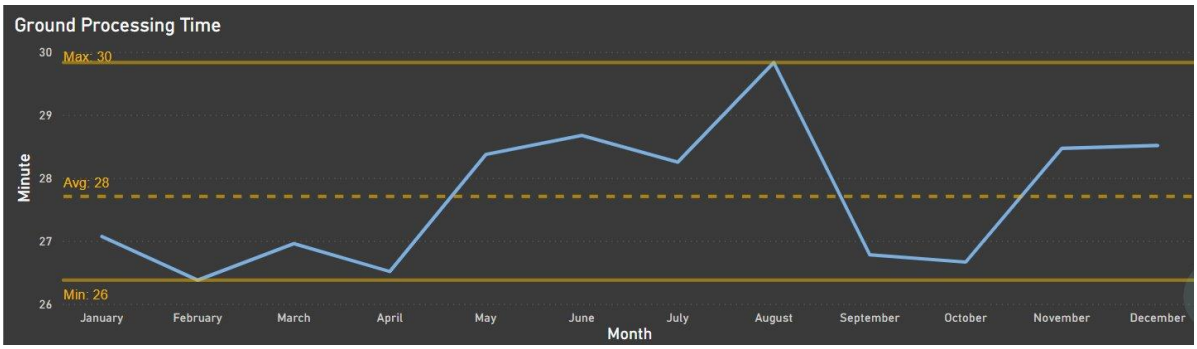


Fig 1.

- These periods coincide with peak travel seasons such as summer vacations and the winter holiday season. Arrival and departure delays increased by approximately 45% compared to other months.

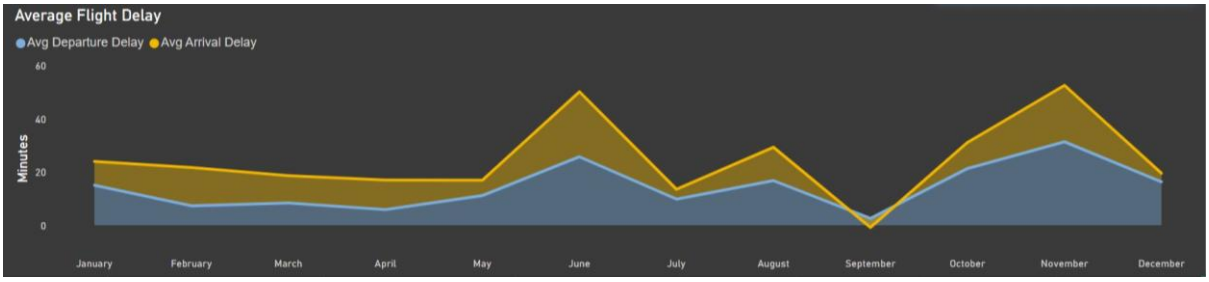


Fig 2.

## Customer Satisfaction:

- In 2018, Neutral (46%) and Satisfied (35%) customers were more common than Unsatisfied (19%) customers.

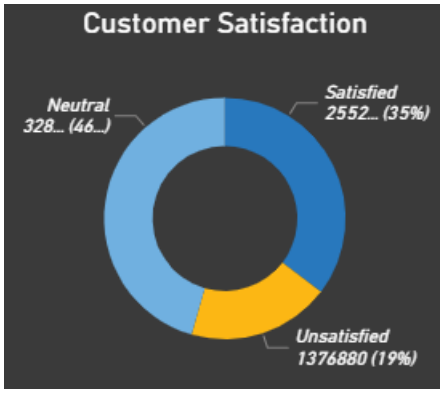


Fig 3.

- Unsatisfactory ratings started increasing in April 2018, reaching their peak in August, before declining toward the end of the year.

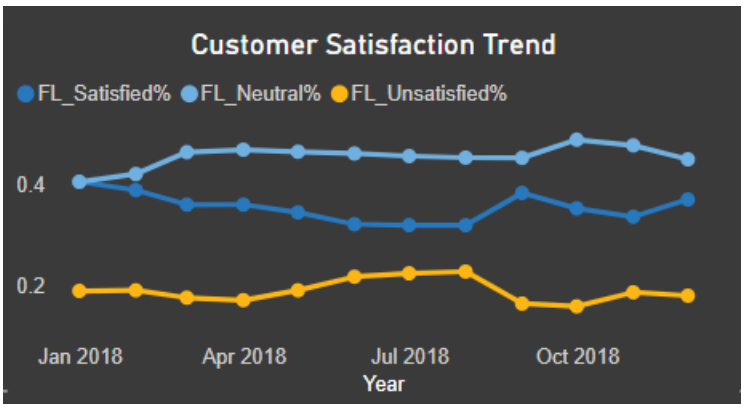


Fig. 4

- Customer dissatisfaction trends vary across airlines and airports. For example, JetBlue Airways at SFO, the highest unsatisfactory rate was in March (38%), compared to August (34%).

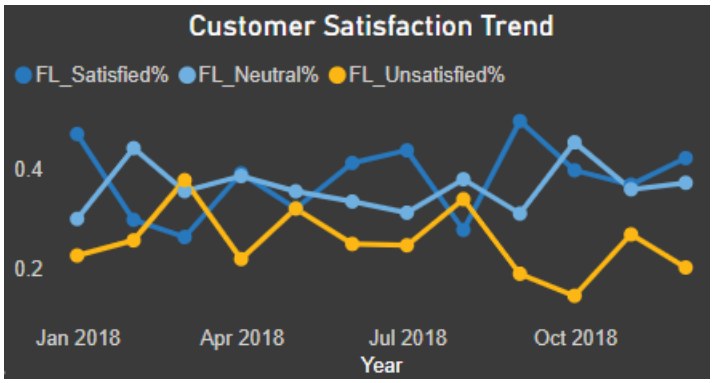


Fig. 5

- At an airport specific level, for example, at the SFO airport customer dissatisfaction showed an upward trend with fluctuations from January to

August, followed by a sharp drop in September before rising again in November, though it did not reach August levels. Total flight volume was highest in August, aligning with the peak in dissatisfaction.

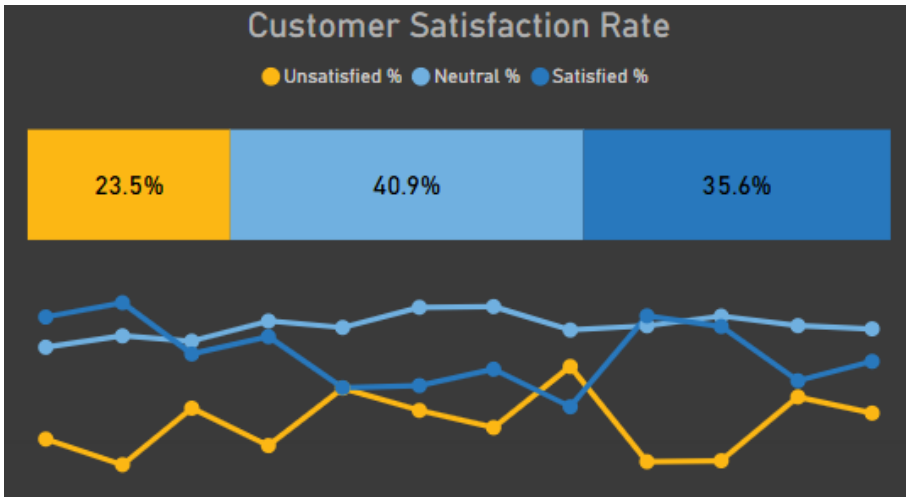


Fig. 6

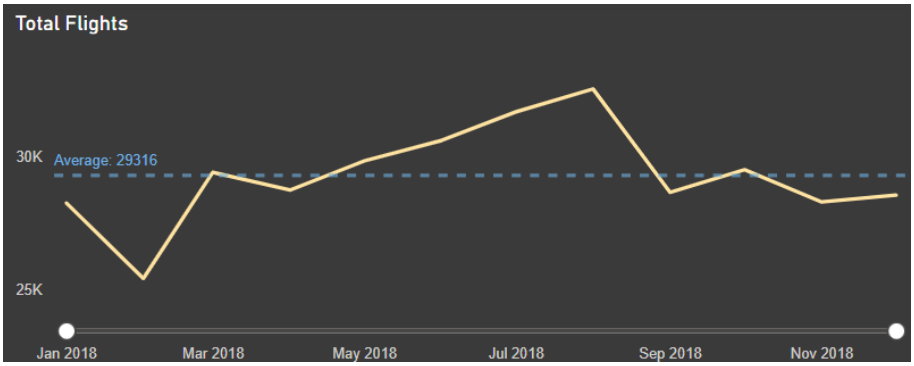


Fig. 7

# DISCUSSION

## Impact of Seasonal Weather conditions on Ground Processing Times and Delays:

1. **Late Summer and Monsoon (June–August)**
   - Thunderstorms and hurricanes in the Southeast and Midwest cause ground stops and rerouted flights, increasing arrival delays.
   - Heat waves affect aircraft performance, requiring longer take-off distances and flight adjustments.
   - Based on a study by the National Center for Atmospheric Research (2018), thunderstorms account for 70% of weather-related delays during this time of the year.

2. **Winter Delays (November–January)**
   - Snowstorms and ice in Denver, New York, and Chicago necessitate de-icing, adding 15–30 minutes to ground processing times.
   - Runway closures for snow removal further delay take-offs and landings.
   - Fog-related visibility issues cause holding patterns and departure delays.
   - A report by the Bureau of Transportation Statistics (2018) highlights that winter weather contributes to 40% of flight cancellations.

# CONCLUSION

An analysis of flight delays, ground processing times, and customer satisfaction trends during 2018 shows operational difficulties encountered by airports, particularly during peak travel seasons.

- Flight delays and ground processing times were significantly higher from June–August and November–January, aligning with summer vacations and the winter holiday season.
- Ground processing times increased by 20–40%, while arrival and departure delays rose by 30–60% during these periods.
- Customer dissatisfaction peaked in August, particularly at SFO, where the highest flight volume was recorded. However, dissatisfaction levels varied across airlines and airports, suggesting multiple contributing factors beyond congestion.

The primary cause of these delays were seasonal weather disruptions, including thunderstorms, hurricanes, snowstorms, and fog, which might have resulted in flight rerouting, de-icing procedures, and runway closures. These conditions significantly impact airport operations, particularly in high-traffic hubs like Atlanta, Chicago, New York, and Denver.

This study highlights the importance of proactive operational strategies, such as enhanced weather forecasting, optimized scheduling, and improved ground processing efficiency, to minimize delays and maintain service quality during high-demand periods. Leveraging data-driven insights through interactive dashboards using PowerBI can help airport authorities to anticipate disruptions, allocate resources and staff effectively, and enhance overall passenger experience based on historical data.

# REFERENCES


Bureau of Transportation Statistics (BTS). (2018). *Winter weather and flight cancellations: A statistical analysis*. Retrieved from https://www.bts.gov

International Air Transport Association (IATA). (2018). *IATA Airline Code Search*. Retrieved from https://www.iata.org/en/publications/directories/code-search/

Kaggle. (2018). *Airline Delays and Cancellations Dataset*. Retrieved from https://www.kaggle.com/

National Center for Atmospheric Research (NCAR). (2018). *The impact of thunderstorms on summer flight delays*. Retrieved from https://ncar.ucar.edu

OpenFlights. (2018). *Airport Database*. Retrieved from https://openflights.org/data.php

Yimga, J. (2020). *Price and marginal cost effects of on-time performance: Evidence from the US airline industry.Journal of Air Transport Management, 84,* 101769. https://doi.org/10.1016/j.jairtraman.2020.101769

# DASHBOARD OVERVIEW

## Airport Level View:

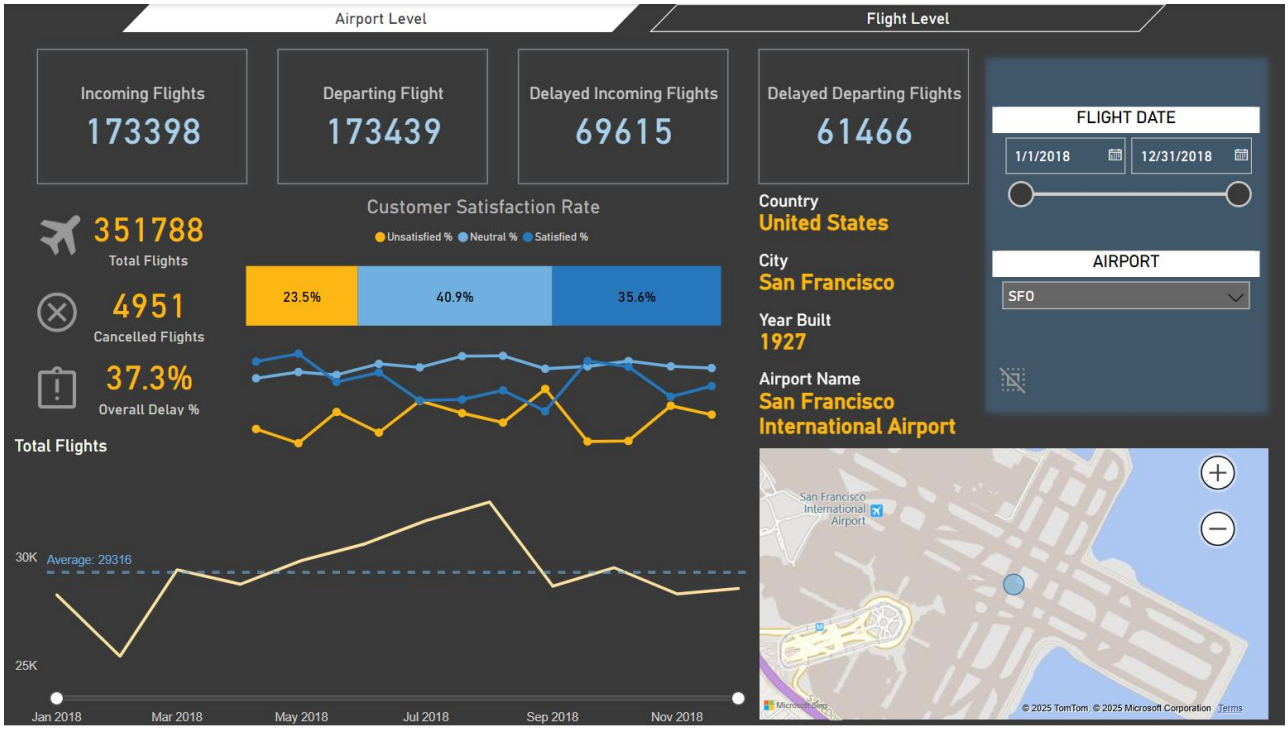


Fig. 8

## Departing Flight Details (Drill Through from Airport Level):

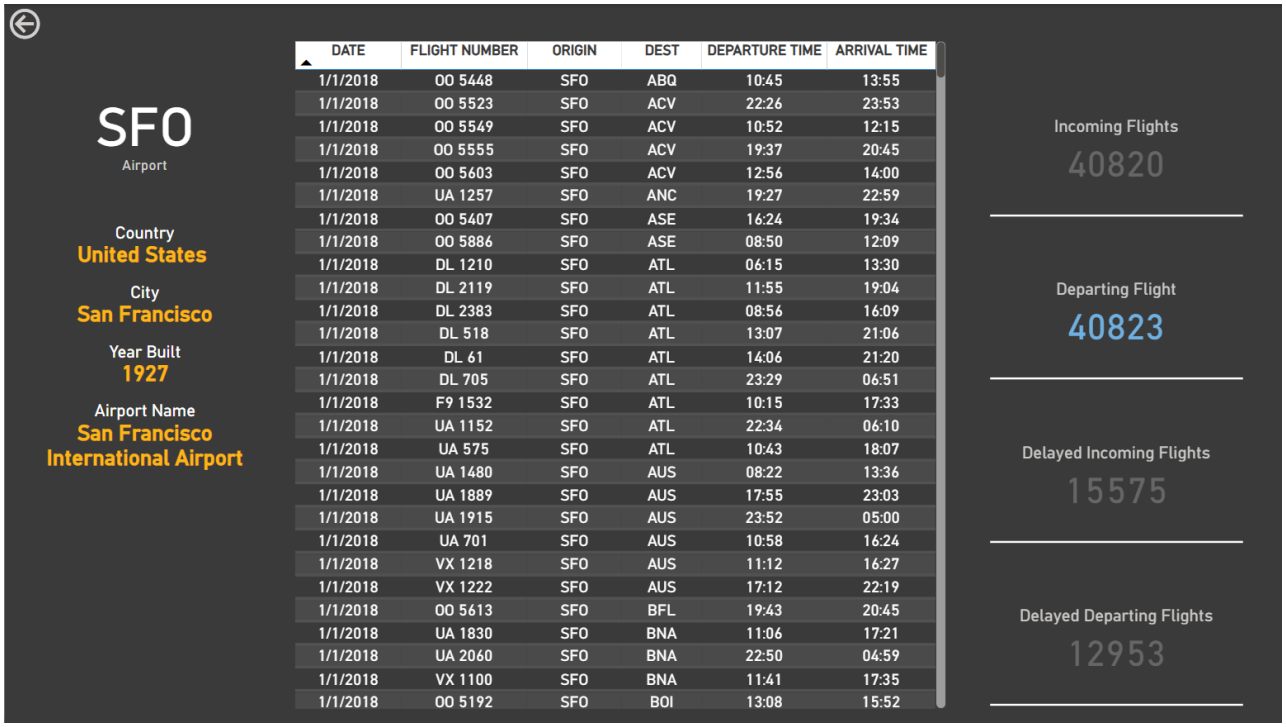


Fig. 9

## Flight Level View:

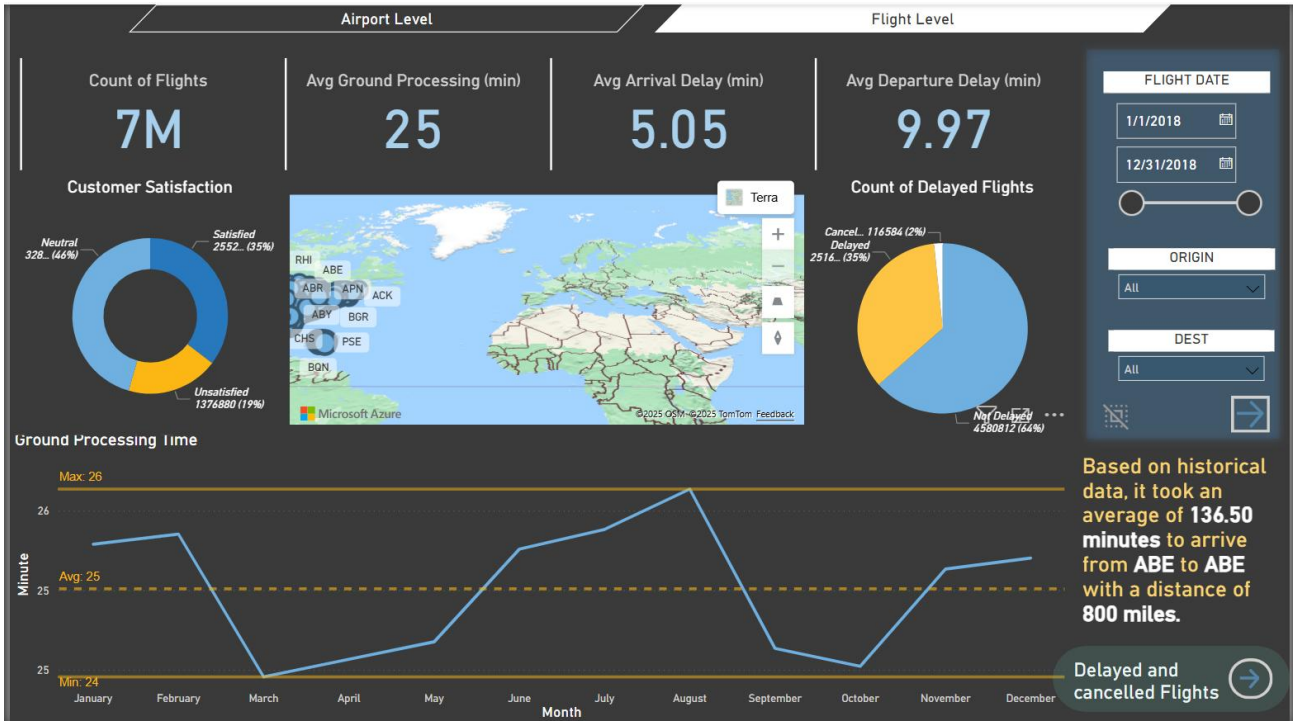


Fig. 10

## Airline View:

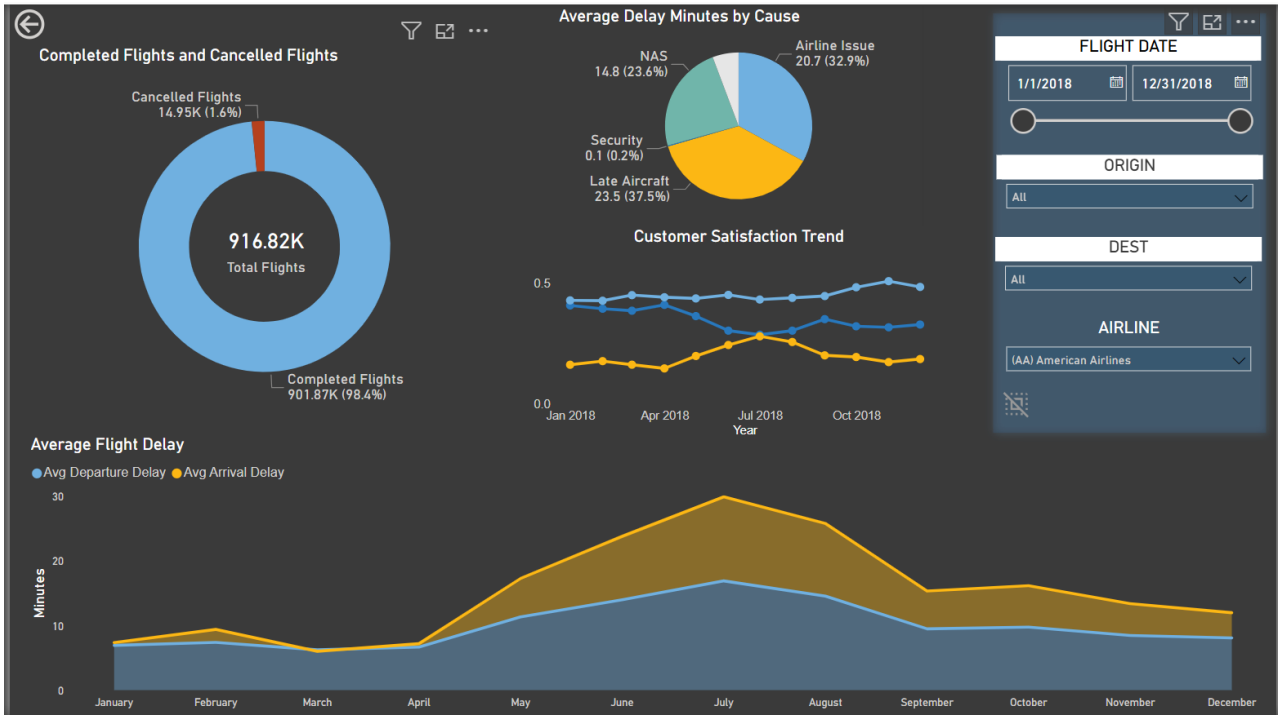


Fig. 11