Infosys Springboard

Internship Report

AirFly Insights – Flight Delay Analysis and Visualization

SUBMITTED BY:

NAME : PUJASHREE E

DURATION : SEP-NOV 2025

Abstract

This internship project, *AirFly Insights*, presents a comprehensive analysis of flight delays using the publicly available **Flight Delay and Causes** dataset from Kaggle. The dataset encompasses extensive flight operation records across various airlines and airports, with detailed information about departure and arrival times, delay causes, and distances. The main objective of the project was to identify key patterns, correlations, and contributing factors behind delays while developing meaningful visualizations to support data-driven decision-making in the aviation sector. The study involved establishing a structured workflow encompassing data acquisition, preprocessing, cleaning, and exploratory data analysis to ensure reliability and insight accuracy.

The project applied a combination of **Python (pandas, NumPy, matplotlib)** for data preprocessing and analytics, along with **Power BI** for creating an interactive dashboard. Data cleaning involved handling missing values, optimizing memory usage, and engineering new features such as flight route, day of the week, and total delay time. Exploratory visualizations revealed patterns in delay frequencies based on routes, airlines, and seasons. The insights derived from this analysis assist in understanding operational inefficiencies, identifying high-risk time periods, and improving flight scheduling strategies. The project concludes by highlighting potential for **predictive modeling and automation** to further enhance reliability and performance monitoring in airline operations.

1. Introduction & Dataset Overview

The AirFly Insights project focuses on understanding flight delays and their causes using a large dataset sourced from Kaggle. The dataset contains flight details, airport operations, and delay reasons for over 480,000 flight records. After preprocessing, the dataset retained approximately 484,310 rows and 33 columns.

Dataset Source: https://www.kaggle.com/datasets/undersc0re/flight-delay-and-causes

2. Data Foundation and Cleaning

Week 1: Project Initialization and Setup

• Defined goals, KPIs, and workflow.
• Loaded CSV files using pandas.
• Explored schema, data types, and handled missing values.
• Performed sampling and memory optimization.

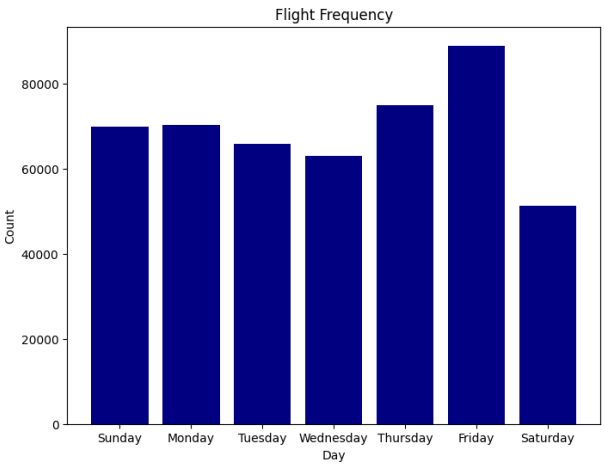
Week 2: Preprocessing and Feature Engineering

Handled missing values in delay and cancellation columns, created derived features such as Month, Day of Week, Hour, and Route. Implemented duplicate removal and data type conversions to standardize the dataset.

Key Performance Metrics

• Minimum Flight Distance: 31 miles
• Maximum Flight Distance: 4,502 miles
• Average: 752.38 miles

Figure 1: Flights by Day



3. Exploratory Data Analysis (EDA)

Exploratory analysis was conducted to identify trends in flight frequency, delay types, and operational patterns. Various charts were created using Power BI and matplotlib to visualize these distributions.

Figure 2: Flights by Month

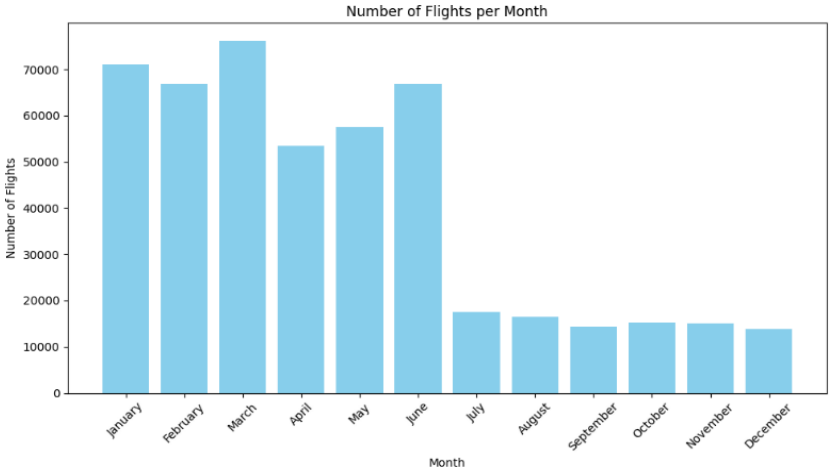


Figure 3: Average Delay by Delay Type

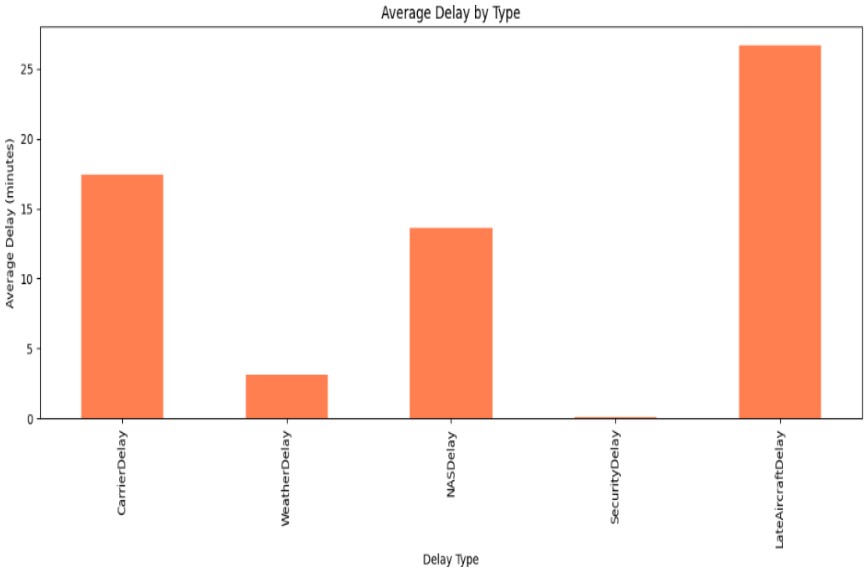


Figure 4: Market Share by Airline (insert chart here)

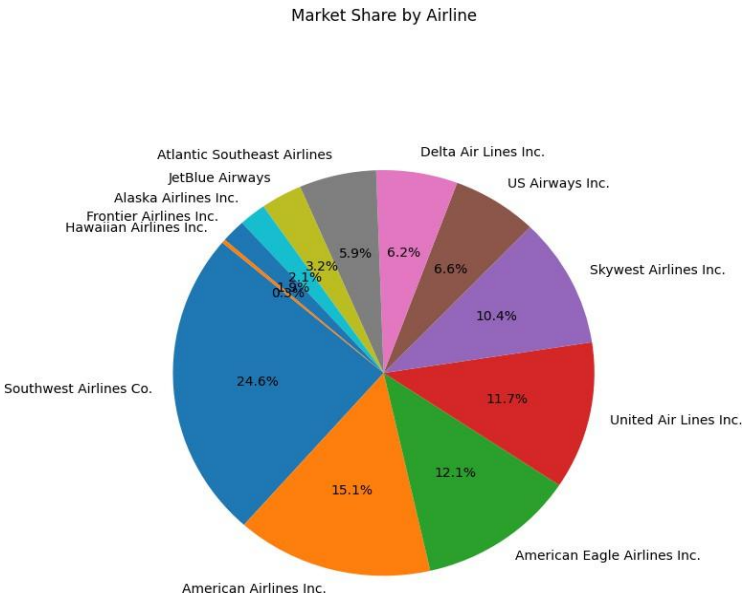


Figure 5: Flight Departures by Time of Day

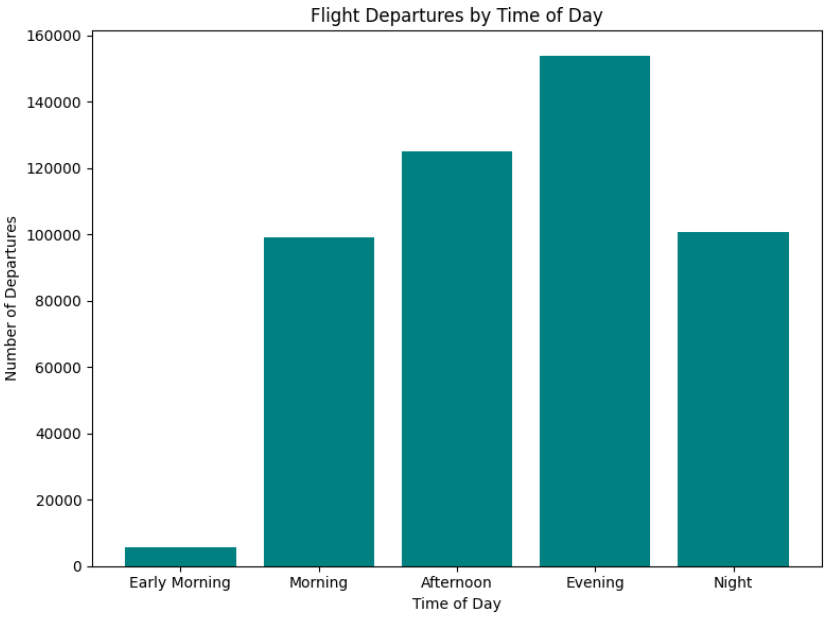


Figure 6: Top 5 Routes by Frequency

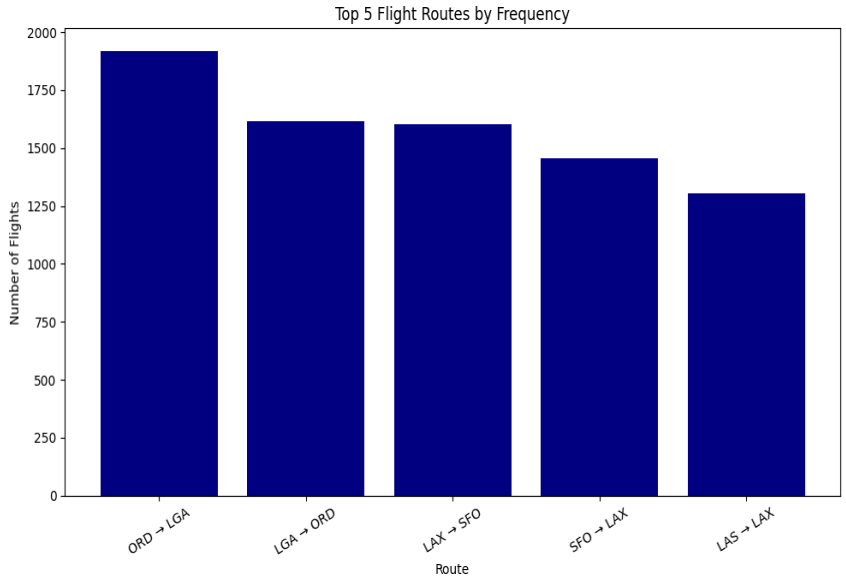
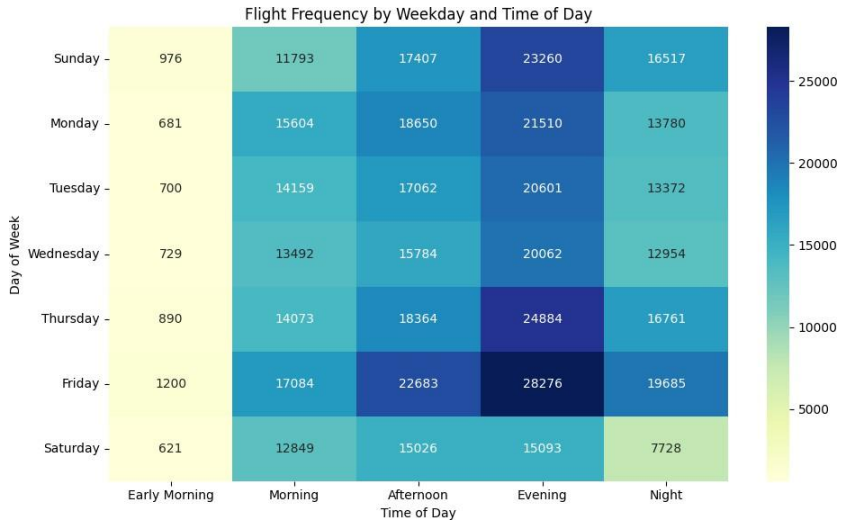


Figure 7: Flight frequency by time and day of week



4. Delay Trends and Insights

Friday was identified as the busiest travel day with 88,972 flights, while Saturday recorded the lowest traffic. ORD–LGA route emerged as the most delay-prone, while ALB–CVG route experienced the least delays. Tuesday exhibited higher delays compared to Wednesday, which had smoother operations.

Figure 8: Total Delay by Route

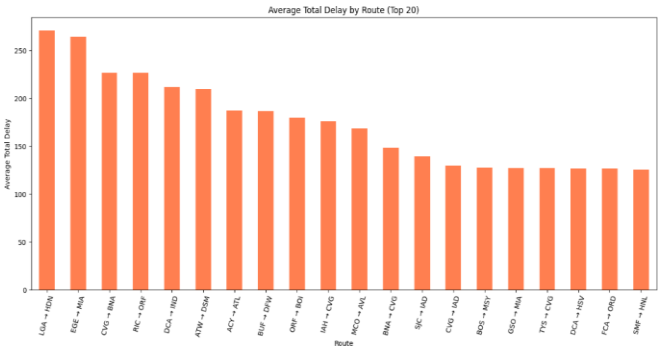


Figure 9: Average Total Delay by Airline

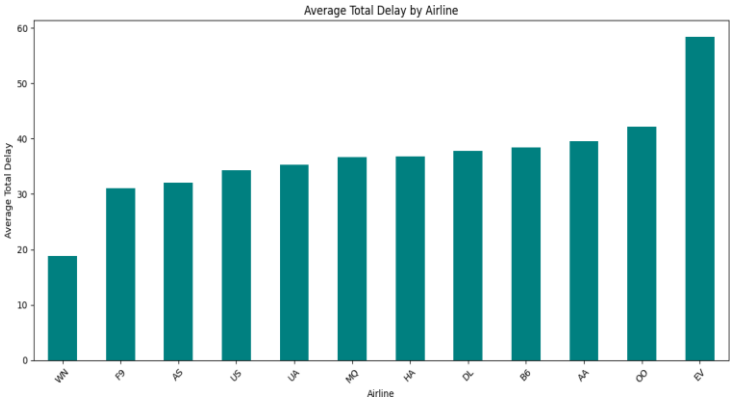


Figure 10: Average Delay types by Airline

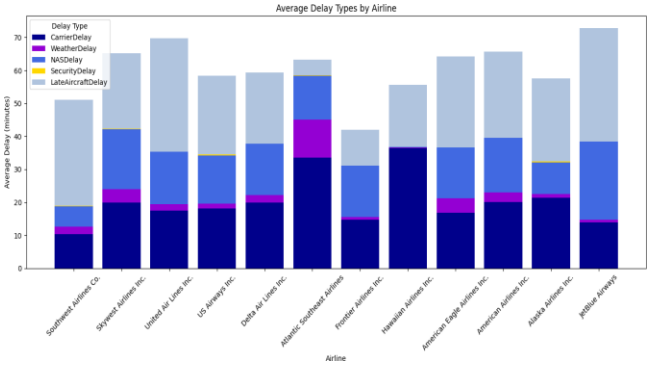


Figure 11: Flight Delay Distribution by month

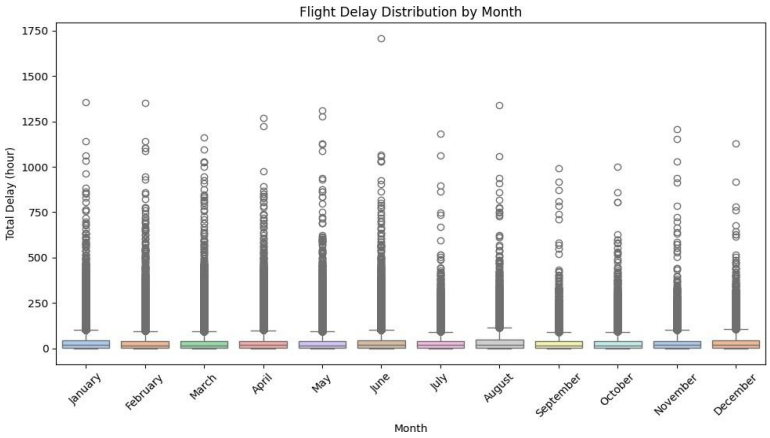


Figure 12: Top 10 Airports by Security Delay

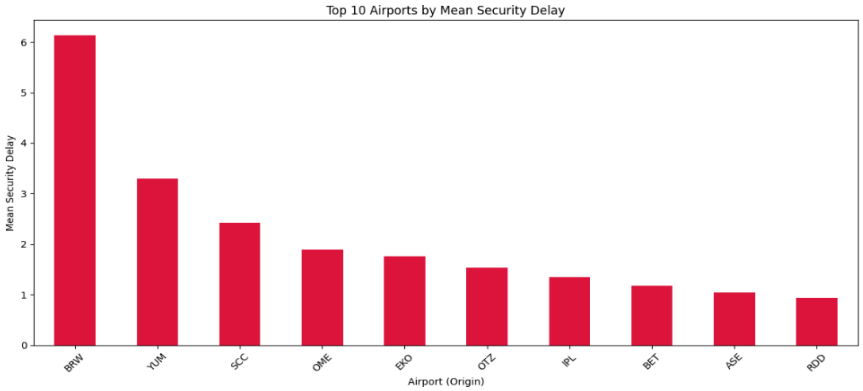


Figure 13: Count of Departure delays by Time of Day

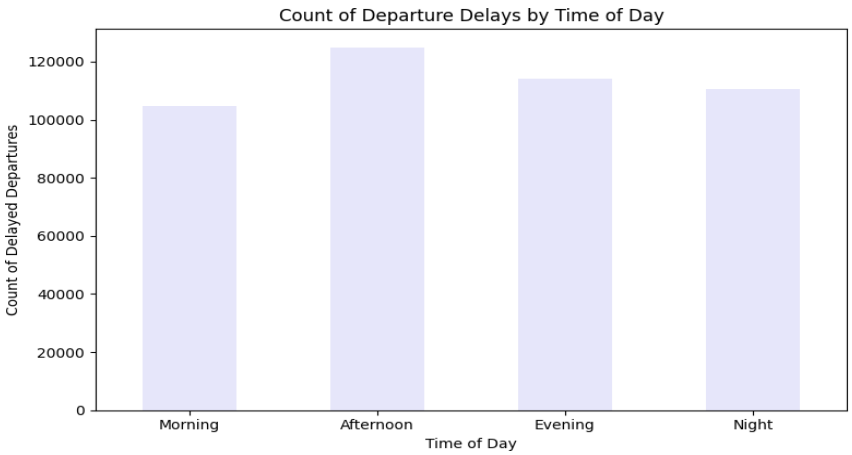


Figure 14: Carrier Delays vs Airlines by month

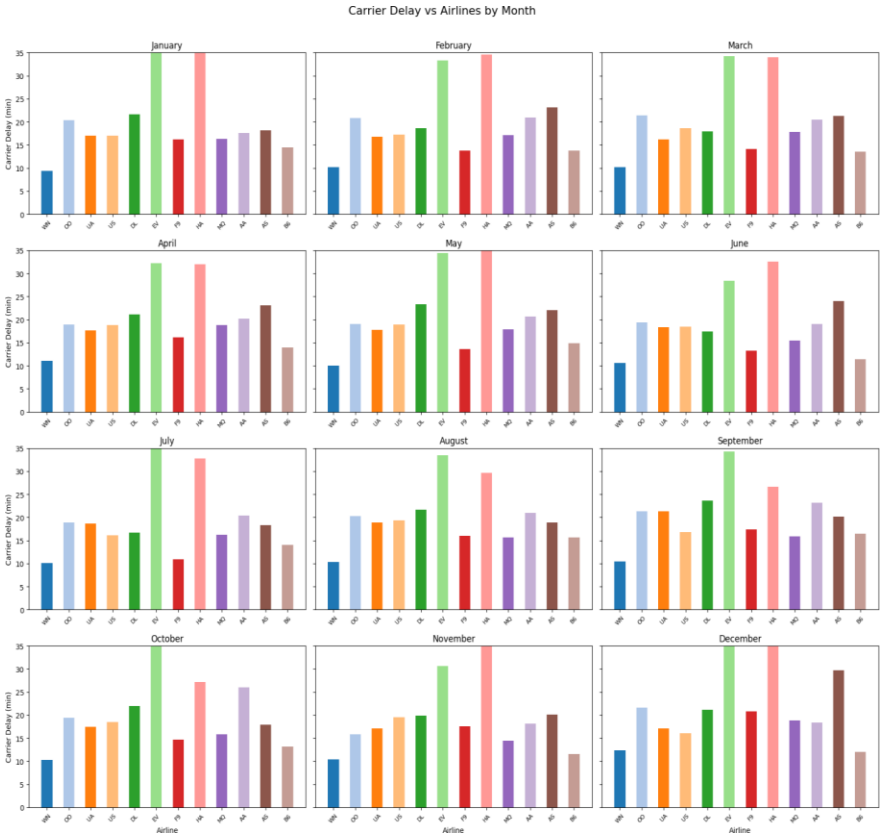


Figure 15: Heatmap of Top 5 Airlines by Mean Delay Types

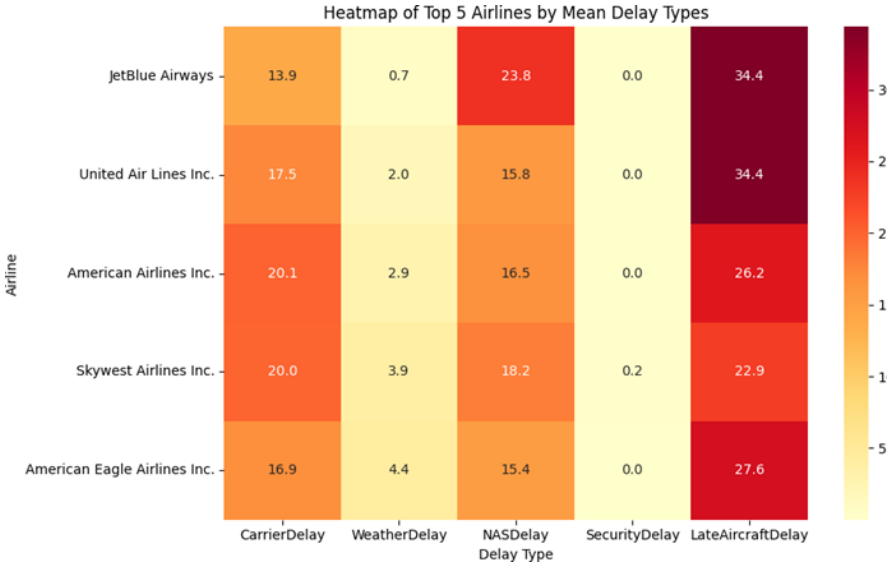


Figure 16: Right Skewed Delay

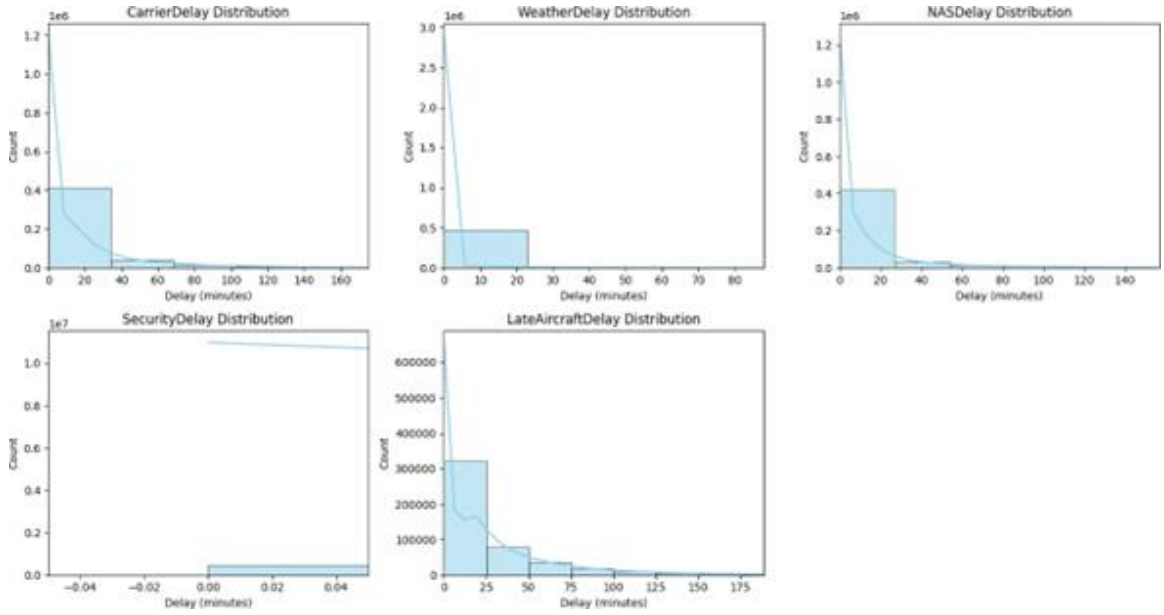
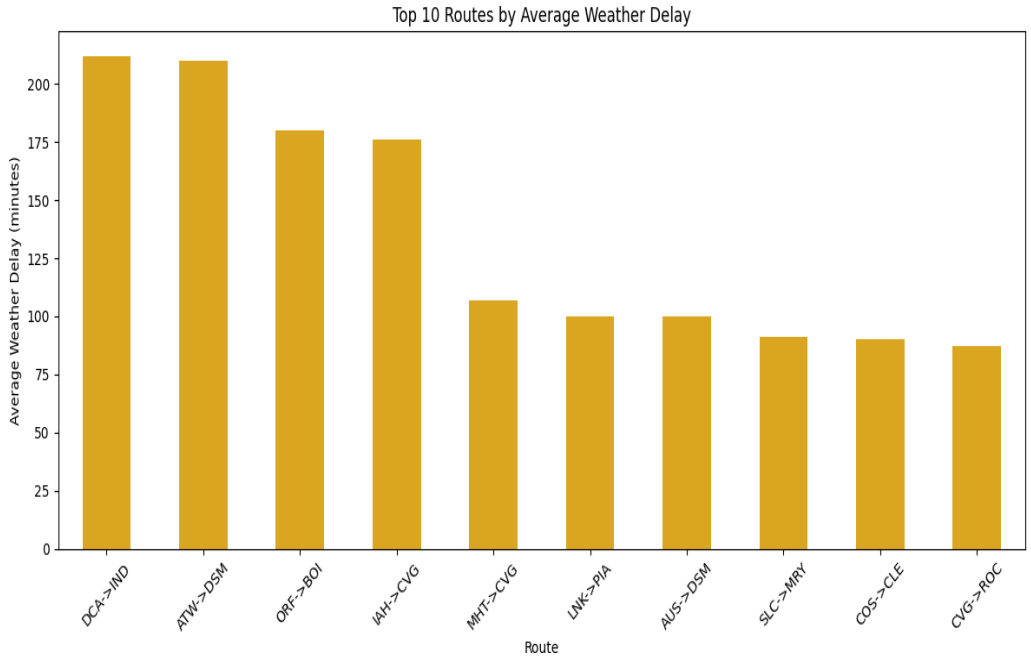


Figure 17: Top 10 Routes by Average Weather Delay



5. Seasonal and Cancellation Analysis

Seasonal variations in delays and cancellations were studied. Winter months showed a higher cancellation rate compared to other seasons. Analysis by delay type indicated that carrier and weather-related issues were predominant during peak travel periods.

Figure 18: Cancellation by Month

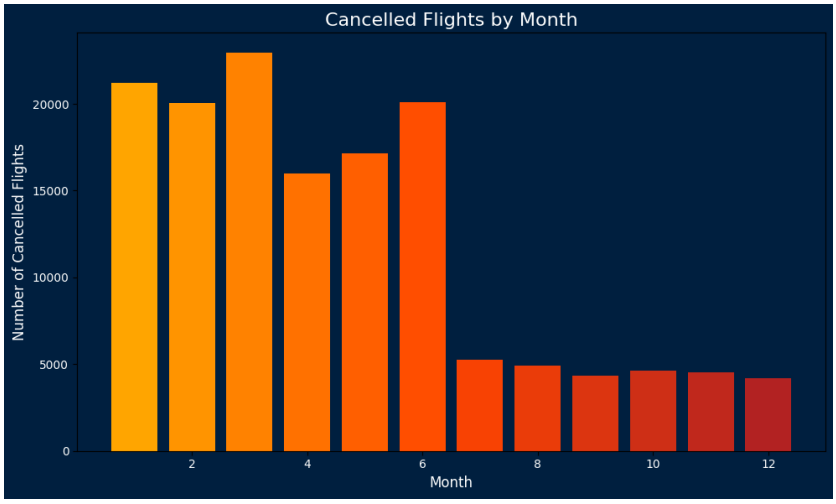


Figure 19: Winter Month Cancellations

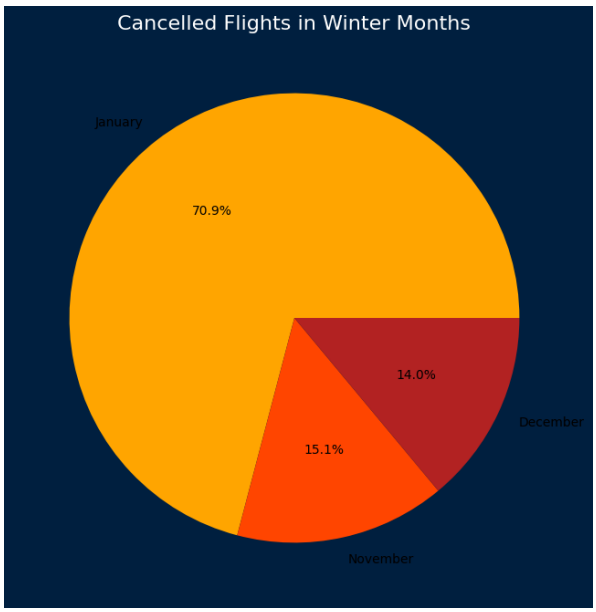


Figure 20: Winter Month vs Non Winter month Cancellations

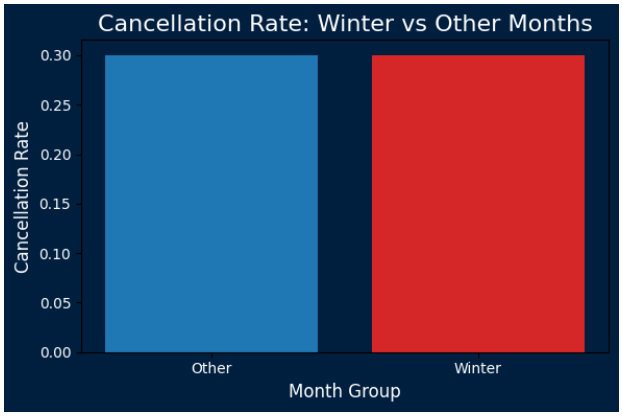


Figure 21: Cancellation by Delay Type

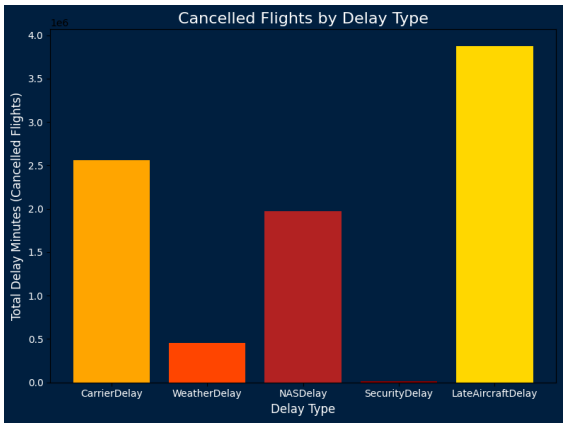


Figure 22: Cancellation Flight by month and delay types

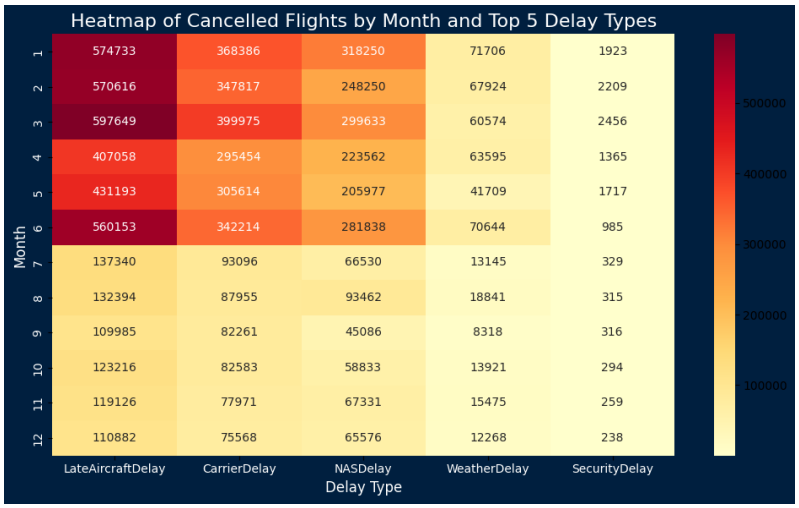
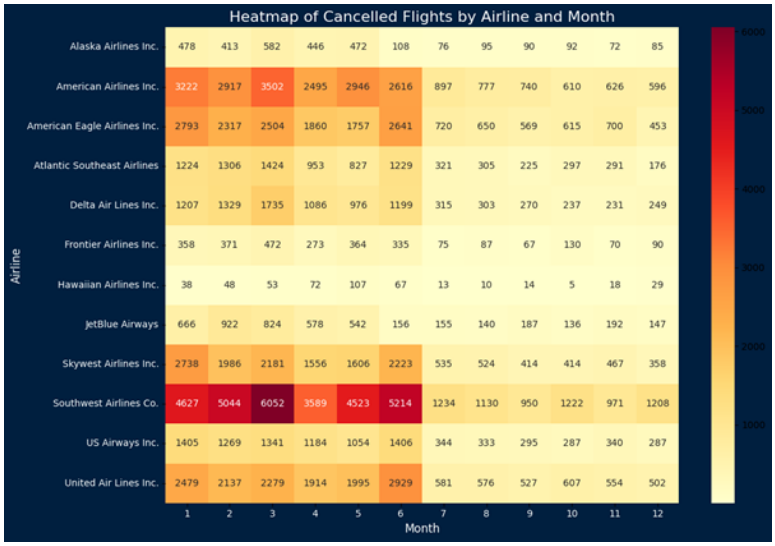


Figure 23: Cancellation Flight by month and airlines



6. Dashboard and Visualization Summary

A Power BI dashboard was created to summarize insights visually. It included charts showing delay trends, route-level performance, and airline comparisons. Interactive filters allowed users to explore flight patterns by month, airline, and delay cause.

Figure 24: Power BI Dashboard Overview

Figure 25: Power BI Dashboard Overview

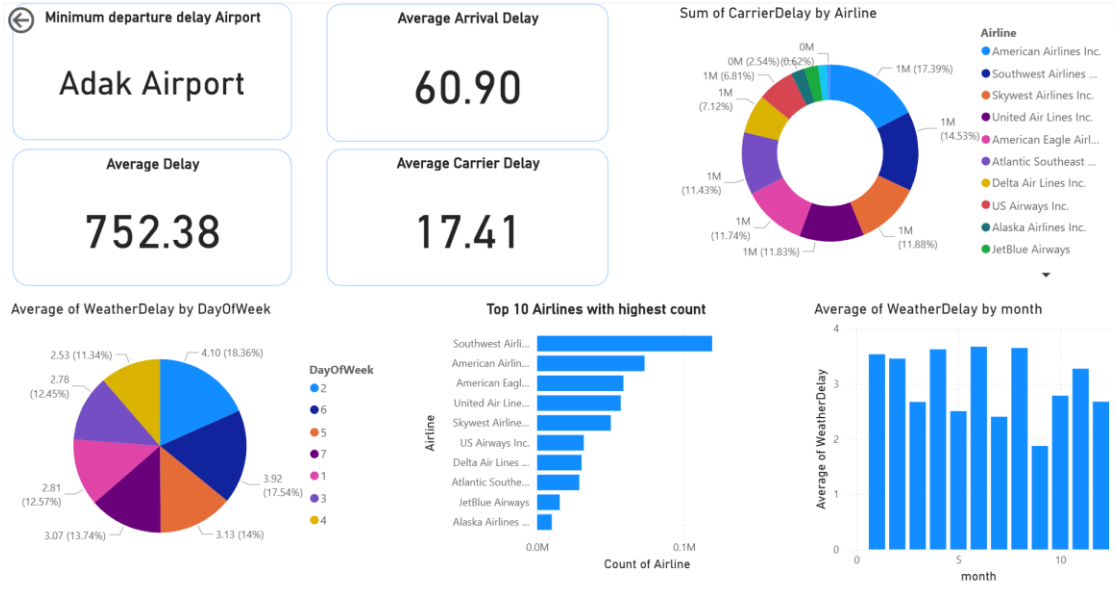
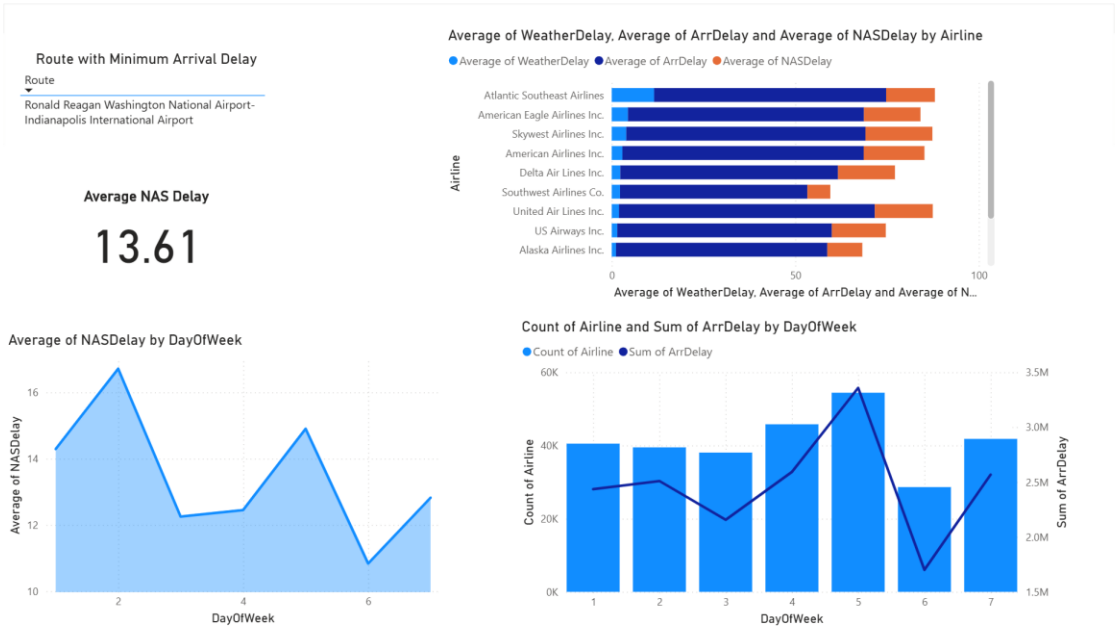


Figure 26: Power BI Dashboard Overview



7. Conclusion and Future Scope

The AirFly Insights project successfully identified patterns in flight delays, emphasizing the influence of specific routes and seasons on performance. Future extensions of this work may involve predictive modeling using machine learning to forecast delays and improve operational efficiency.