# AirFly Insights – Data Visualization and Analysis of Airline Operations

## Internship Report

Submitted for: Infosys Springboard 6.0 Internship

This report presents the detailed documentation of the AirFly Insights internship project, executed over seven weeks. The project aimed to analyze large-scale airline data, uncover operational patterns, and visualize flight delays, cancellations, and trends. All work was conducted in Google Colab using Python-based analytical tools, emphasizing real-world data handling, visualization, and storytelling through data.

## Executive Summary

The AirFly Insights project focused on understanding airline operational efficiency using advanced data visualization and analytics techniques. Google Colab served as the primary environment for data exploration, cleaning, and visualization. The internship journey involved structured weekly milestones: starting with raw data preparation and culminating in the creation of a comprehensive visual dashboard. The project combined technical and analytical thinking to extract meaningful insights that could assist in operational decision-making for airlines.

This report outlines week-by-week progress, technologies used, and outcomes achieved, demonstrating practical exposure to real-world data analysis using Python, Power BI, and visualization libraries.

## Week 1 – Data Acquisition and Understanding

The first week was dedicated to setting up the project environment and exploring the dataset. The Kaggle Airlines dataset, containing over 480,000 flight records, was imported into Google Colab and explored using pandas. The dataset included essential columns such as flight dates, airline codes, origin and destination airports, delay reasons, and cancellation codes. Data profiling was done to check null values, duplicates, and data types. Memory optimization techniques were explored to make the dataset efficient for analysis.

**What I Did:**

• Imported the Kaggle Airlines dataset into Google Colab. • Explored data structure, schema, and column types using pandas. • Identified missing values and duplicate records. • Performed initial memory optimization and profiling for large datasets.

## Week 2 – Data Cleaning and Feature Engineering

Week 2 focused on preparing the dataset for detailed analysis. Data cleaning included handling missing values, removing duplicates, and filling categorical fields logically. Feature engineering was introduced to make the data more analytical. New columns like 'Route' (Origin–Destination pair), 'Month', 'Day of Week', and 'Hour' were created. Time fields stored in HHMM format were converted to proper datetime objects for more accurate analysis. These transformations formed the foundation for future visual insights.

**What I Did:**

• Cleaned delay and cancellation columns, filled missing values with zeros or placeholders. • Converted time columns into datetime format using pandas. • Created derived features: Route, Month, Day of Week, and Hour. • Verified dataset accuracy and exported cleaned version for future analysis.

## Week 3 – Univariate and Bivariate Visual Analysis

In Week 3, visualization was introduced. Univariate analysis explored single-variable trends such as the most frequent airlines, routes, and flight volumes. Bivariate analysis studied relationships between two variables — for example, delay duration vs. airline and flight volume vs. month. Visualizations helped reveal which airlines had the most consistent operations and which routes were more prone to delays.

**What I Did:**

• Conducted univariate and bivariate analysis using matplotlib and seaborn. • Visualized top airlines, busiest months, and flight distributions. • Compared airline performance and delay trends through plots. • Documented observations and visual insights in Colab notebooks.

## Week 4 – Delay Cause Analysis and Time-Based Trends

Week 4 emphasized understanding the reasons behind delays. Each delay type (Carrier, Weather, NAS, Security) was analyzed separately and compared across airlines. Time-based visualizations helped discover when delays were more frequent — for example, peak delay hours and days with maximum congestion. These visual insights provided a clear understanding of the operational inefficiencies within airline schedules.

**What I Did:**

• Grouped and analyzed different delay categories using pandas. • Created comparative bar and line charts for carrier and weather delays. • Identified top delay-prone airlines and peak delay hours. • Summarized key findings through plots in seaborn and matplotlib.

## Week 5 – Route and Airport-Level Analysis

Week 5 expanded the analysis to airports and routes. Heatmaps and route-level visualizations were developed to show which airports experienced the highest average delays. The busiest routes and congested airport pairs were identified. Folium was explored for interactive map creation, providing a spatial understanding of operational delays.

**What I Did:**

• Visualized top 10 busiest routes and most delayed airports. • Created heatmaps for average delays across airports. • Used plotly and folium for interactive route-level visualization. • Analyzed delay trends regionally and across flight paths.

## Week 6 – Seasonal and Cancellation Analysis

The sixth week focused on seasonal behavior and cancellation trends. Seasonal visualizations showed how operations varied across months. Cancellation reasons such as Carrier, Weather, NAS, and Security were analyzed to find the dominant cause. Patterns revealed that winter months had increased weather-related cancellations.

**What I Did:**

• Analyzed monthly and seasonal flight data. • Visualized cancellation trends and reasons using bar and pie charts. • Identified peak cancellation months and leading causes. • Summarized findings in structured markdown and plots.

## Week 7 – Final Visual Report and Dashboard Preparation

The final week involved compiling the complete analysis into a structured visual report. All plots were refined with proper titles, labels, legends, and color consistency. Insights from previous weeks

were summarized into a coherent storyline, and Power BI was used to create a professional dashboard highlighting the key operational metrics of the airline dataset.

**What I Did:**

• Reviewed and refined all previous week's visualizations. • Organized plots into a cohesive analytical narrative. • Created a Power BI dashboard summarizing key metrics. • Prepared and exported final visual report for submission.

## Tools and Technologies Used

- **Development Platform:** Google Colab - **Programming Language:** Python - **Libraries:** pandas, numpy, matplotlib, seaborn, plotly, folium - **Dashboard Tool:** Power BI - **Dataset:** Kaggle Airlines Flight Data (484,000+ records) - **Documentation:** Jupyter notebooks and final PDF reports

## Conclusion and Learnings

This seven-week internship provided hands-on experience with real-world data analysis workflows. Working with Google Colab enhanced productivity and collaboration. Each week's milestone contributed to developing a deeper understanding of data preparation, visualization, and insight communication. The internship helped strengthen technical expertise in Python and Power BI while improving the ability to interpret and narrate data-driven stories effectively.

By the end of this internship, I successfully built a data visualization pipeline, conducted analytical evaluations, and produced a professional dashboard summarizing airline operational performance. This experience provided both technical learning and analytical maturity essential for future roles in data science and visualization.