Final Report – Global Airline Operations (2022–2023)

1. Introduction

Airline disruptions such as delays and cancellations are major challenges in the aviation industry. These events not only cause passenger dissatisfaction but also result in significant operational costs. This project investigates the causes of flight disruptions based on a dataset sourced from Kaggle, covering airline operations between 2022 and 2023. The objective is to analyze flight status patterns and provide actionable insights for performance improvement.

2. Data Scoping

The project focuses on identifying the causes and patterns of flight disruptions (delays and cancellations) in the airline industry.

The scope was narrowed to include relevant attributes such as `Flight Status`, `Departure Date`, `Country`, `Continent`, and `Passenger ID`.

These features were selected to allow temporal and geographical analysis. The goal was to highlight seasonal trends and regional disparities using a publicly available dataset from Kaggle (2022–2023).

3. Data Curation & Cleaning

The dataset required several cleaning steps. First, irrelevant columns such as `First Name`, `Last Name`, and `Pilot Name` were removed. Missing values were identified and either dropped or imputed depending on the variable. Two engineered features were created:

- `TotalDelay`: which combines multiple delay components into one variable.
- `IsWeekend`: a binary variable to differentiate weekday vs. weekend flights.

This process ensured the data was ready for exploratory and visual analysis in Tableau.

4. Data Analysis & Computation

• Dataset: Airline Dataset (Kaggle – 2022–2023)

• Tools Used: Python (Pandas), Tableau, Excel

• Data Wrangling: Removed irrelevant columns, cleaned missing values, engineered features (TotalDelay, IsWeekend, etc.)

Exploratory Data Analysis (EDA) Highlights:

- Distribution of Flight Status: Cancelled, Delayed, On-time

- Histogram of Delay Durations

- Monthly Trend of Delays

- Geographic visualization of delays and cancellations

5. Statistical Analysis & Predictive Modeling

No predictive models were developed in this project. The focus was on exploratory and descriptive analysis to support operational decision-making.

6. Challenges and Solutions

- Challenge: Dataset had multiple irrelevant or redundant columns

- Solution: Performed deep cleaning and data profiling;

- Challenge: Tableau visual rendering and filters

- Solution: Manual configuration of calculated fields and improved formatting.

7. Description of Dashboard

Link: https://public.tableau.com/shared/NN7B4C94F?:display_count=n&:origin=viz_share_link

• Purpose: To enable users to interactively explore flight disruptions over time and geography

• Key Features:

  - Filters by status, date, region

  - Pie chart, line chart, map visualization

• Utility: Understand seasonal and regional disruption patterns; guide airline operations strategy

8. Conclusions and Future Work

Conclusions:

- Over 66% of flights experienced disruptions (delays or cancellations)

- Peak delays often align with holiday periods, suggesting planning issues

- Some carriers are more prone to late aircraft or weather-related issues

Future Work:

- Enrich the dataset with weather and traffic conditions

- Develop predictive models to forecast disruptions

- Drill down analysis for individual airline performance

9. References & Acknowledgements

- Dataset: https://www.kaggle.com/datasets/iamsouravbanerjee/airline-dataset

- Tools: Tableau, Excel

- Templates: Genigraphics Poster Presentation: https://www.genigraphics.com/templates