AirFly Insights: Data Visualization and Analysis of Airline Operations

---

Project Statement:
The objective of this project is to analyze large-scale airline flight data to uncover operational trends, delay patterns, and cancellation reasons using data visualization techniques. The goal is to help understand airline and airport-level performance and contribute to actionable insights using visual analysis.

Expected Outcomes:

- Understand and preprocess aviation datasets for analysis
- Explore trends in flight schedules, delays, cancellations, and routes
- Visualize key metrics using bar charts, time series, heatmaps, maps, and comparisons
- Provide insights for stakeholders including airline operators and analysts
- Summarize findings through a final visual report and presentation

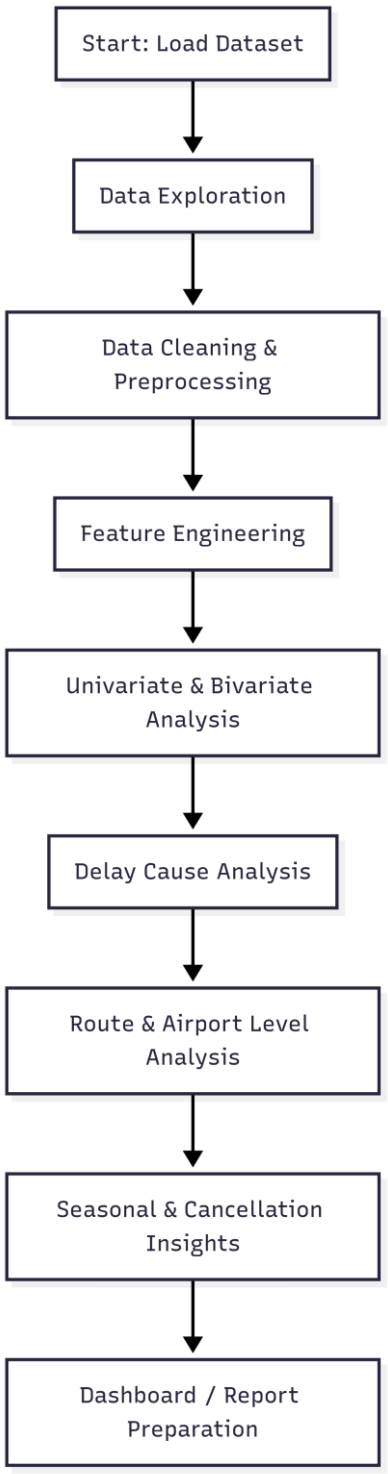
---

Dataset:
Source: Kaggle Airlines Flights Data
Contains over 60 million records covering:

Modules to be Implemented:

1. Data Acquisition and Understanding
2. Data Cleaning and Feature Engineering
3. Univariate and Bivariate Analysis
4. Delay Cause Analysis
5. Cancellation and Seasonal Trends
6. Airport and Route-Level Exploration
7. Final Dashboard or Report
8. Documentation and Presentation

Project Flow:



Week-wise Implementation Plan

Milestone 1: Data Foundation and Cleaning

Week 1: Project Initialization and Dataset Setup

- Define goals, KPIs, and workflow
- Load CSVs using pandas
- Explore schema, types, size, and nulls
- Perform sampling and memory optimizations

Week 2: Preprocessing and Feature Engineering

- Handle nulls in delay and cancellation columns
- Create derived features: Month, Day of Week, Hour, Route
- Format datetime columns
- Save preprocessed data for fast reuse

Deliverables:

- Cleaned dataset
- Summary of preprocessing logic
- Feature dictionary

Milestone 2: Visual Exploration and Delay Trends

Week 3: Univariate and Bivariate Visual Analysis

- Top airlines, routes, and busiest months
- Flight distribution by day, time, and airport
- Plot bar charts, histograms, boxplots, and line plots

Week 4: Delay Analysis – Airline and Weather

- Compare delay causes by airline
- Explore carrier delays, weather delays, NAS delays
- Visualize delays by time of day and airport

Deliverables:

- Set of visualizations (minimum 8)
- Observations on peak delays and top delay-prone carriers

Milestone 3: Route, Cancellation, and Seasonal Insights

Week 5: Route and Airport-Level Analysis

- Top 10 origin-destination pairs
- Delay heatmaps by airport and route
- Maps showing busiest airports and average delays

Week 6: Seasonal and Cancellation Analysis

- Monthly cancellation trends
- Cancellation types: carrier, weather, security, NAS
- Analyze impact of holidays or winter months

Deliverables:

- Seasonal visual summaries
- Insights on route congestion and cancellations

Milestone 4: Report and Presentation

Week 7: Visual Report or Dashboard Preparation

- Combine plots into a coherent storyline
- Use markdown, presentation slides, or Streamlit
- Ensure plots include labels, titles, legends, and axis clarity

Week 8: Documentation and Final Presentation

- Create final report (PDF or README)
- Build slide deck for presentation
- Record insights and visual walkthrough

Deliverables:

- Final report or dashboard
- Slide deck
- GitHub repository with code and visuals

Evaluation Criteria

Milestone 1: Data Readiness

- Clean structure, usable features, correct types

Milestone 2: Visual Analysis

- Clarity, relevance, and diversity of plots
- Interpretability of visualizations

Milestone 3: Insight Discovery

- Airport, carrier, and cancellation insights backed by visuals

Milestone 4: Presentation and Reporting

- Structured report, cohesive storyline, visually supported summary

---

Tech Stack

Data Handling: pandas, numpy
Visualization: matplotlib, seaborn, plotly, folium
Optional Dashboard: Streamlit, PowerBI
Documentation: Jupyter Notebook, PDF, GitHub