

## Week3\_AirFly\_Insights

### Objectives

- Perform exploratory data analysis (EDA) on the cleaned flight dataset.
- Produce visualizations illustrating delay distributions, busiest airports/routes, carrier performance, and temporal patterns.
- Create reusable plotting functions for consistent styling and fast iteration.
- Save plots to disk for inclusion in presentations and reports (optional).
- Provide insights from visual analysis that inform modeling choices in later weeks.

### Summary of Code Snippets (1-50)

Provided sequentially in the accompanying chat content. Key areas covered:

- Data loading & quick checks
- Distribution plots (delays, distances)
- Time series and seasonal plots
- Carrier and route comparisons
- Heatmaps and pivot tables for origin-destination analyses
- Boxplots and violin-like summaries (using matplotlib)
- Plot saving and lightweight dashboards
- Utility plotting helpers for reuse

### Key Visual Findings

- Delay distributions are right-skewed with a long tail of extreme delays; median is lower than mean.
- Peak delays cluster around particular hours of the day and certain high-volume routes.
- Some carriers show consistently lower median delays but higher variance, suggesting occasional extreme disruption.
- Cancellation patterns often align with specific routes and may spike on certain days/seasons.

## Challenges & Notes

- For reproducible visuals, ensure the cleaned parquet or CSV is the same dataset used in Week 2 transformations.
- Matplotlib was used exclusively (no seaborn) and each plot is produced in its own cell as required.
- Some plots (heatmaps) require pivot tables that could be large — consider sampling or aggregating by week/month for clarity.