Milestone 2: Visual Exploration and Delay Trends

Week 3 – Univariate and Bivariate Visual Analysis

Objective

To perform exploratory data analysis (EDA) on flight delay data by examining univariate and bivariate patterns. The goal is to identify top-performing airlines and routes, temporal flight patterns, delay distributions, and relationships among numerical features.

Dataset Overview

Dataset Name: Flight_delay_final.csv
Rows: 484,549
Columns: 33
Duration: Year 2019
Key Columns: Airline, Route, Month, DayOfWeek, Hour, Org_Airport, DepDelay, ArrDelay, AirTime, CRSElapsedTime

Univariate and Bivariate Visual Analysis

Top Airlines by Number of Flights

Southwest Airlines has the maximum number of flights, followed by Delta, American, and United Airlines.

Top 10 Busiest Routes

Routes between ATL–ORD, LAX–LAS, and DFW–DEN appear most frequently, representing hub-to-hub routes with heavy traffic.

Flight Distribution by Month

Flights remain fairly consistent across months, with a slight rise during summer (May–July).

Flight Distribution by Day of Week

Thursday and Friday record the highest flight volumes, while Saturday has the fewest flights.

Flights by Hour of Departure

Flight frequency peaks around 8–10 AM and 5–7 PM, typical for business travel patterns.

Top 10 Busiest Origin Airports

ATL (Atlanta), ORD (Chicago O'Hare), DFW (Dallas/Fort Worth), and DEN (Denver) are top origin airports.

Average Arrival Delay by Airline

Frontier and Spirit Airlines have the highest average delays, while Delta and Alaska Airlines maintain lower averages.

Correlation Heatmap (Continuous Features)

Departure delay and arrival delay are highly correlated (~0.95), showing that delays at departure often propagate to arrival.

Summary of Key Insights

Aspect	Key Finding
Top Airlines	Southwest Airlines operates the highest number of flights.
Busiest Routes	ATL–ORD, LAX–LAS, and DFW–DEN are top routes.
Monthly Pattern	Slight increase in flights during summer months (May–July).
Weekly Pattern	Thursdays and Fridays are busiest; Saturday least busy.
Daily Pattern	Peak hours: 8–10 AM and 5–7 PM.
Airports	ATL, ORD, DFW, and DEN dominate departures.
Delays	Frontier and Spirit have highest average delays.

Correlations	Departure delays strongly influence arrival delays.

Conclusion

The exploratory analysis reveals that the U.S. flight network is hub-dominated and schedule-driven. Delays show systemic propagation — once a flight departs late, arrival delays are almost certain. Operational peaks occur in summer months and weekdays, coinciding with business and leisure travel demand. Major hub airports experience heavier congestion, aligning with higher route frequencies.