# Airfly Insights Report

## 1. Dataset Overview

The Flight Delay dataset contains detailed information on domestic U.S. flights, including airline, flight number, origin and destination airports, scheduled and actual departure/arrival times, and various types of delays (carrier, weather, NAS, security, late aircraft). After cleaning and preprocessing, the dataset is ready for exploratory data analysis (EDA) and predictive modeling.

## 2. Data Cleaning Steps (Pandas)

The following data cleaning and preprocessing steps were performed using Pandas:  
1. Removed duplicate records using drop\_duplicates().  
2. Handled missing values:  
 - Delay-related columns (ArrDelay, DepDelay, CarrierDelay, WeatherDelay, NASDelay, SecurityDelay, LateAircraftDelay) were filled with 0.  
 - Categorical columns (Airline, Origin, Dest, TailNum) were filled with 'Unknown'.  
 - Rows with missing Date or FlightNum were dropped.  
3. Converted data types:  
 - Converted Date column to datetime format.  
 - Reformatted DepTime and ArrTime into proper HH:MM format.  
4. Feature Engineering:  
 - Created TotalDelay column = sum of all delay types.  
 - Created DelayFlag = 1 if ArrDelay > 15 minutes, else 0.  
 - Extracted DayOfWeek, Month, and Year from Date.  
5. Outlier Handling:  
 - Capped extreme delay values (e.g., above 99th percentile) to reduce skewness.

## 3. Metrics and Insights

From the cleaned dataset, the following insights and metrics were generated:

• Overall Delay Trends: A significant portion of flights experience delays, with arrival delays often being higher than departure delays.  
• Airline-Level Patterns: Different airlines show varying on-time performance. Some carriers consistently report higher average delays.  
• Airport & Route Analysis: Congested hubs (e.g., ATL, ORD, LAX) typically show longer delays compared to regional airports.  
• Time & Seasonal Variations: Delays are more frequent during late evenings and weekdays, with strong seasonal peaks during winter storms and holiday travel periods.  
• Delay Contribution Breakdown: Carrier-related issues and Late Aircraft are the largest contributors to delays, while Weather delays, though less frequent, tend to cause the longest disruptions.

## 4. Suggested Visualizations

• Distribution of Delays: Histogram of ArrDelay and DepDelay.  
• Airline & Airport Performance: Bar chart of average delay per airline, heatmap for origin-destination delays.  
• Temporal Patterns: Line chart for hourly delays, bar chart for weekday vs weekend delays, and monthly trends.  
• Delay Cause Analysis: Stacked bar chart for different delay types, pie chart for delay cause proportions.  
• On-Time Performance Dashboard: KPIs showing % on-time flights, average delay minutes, and top 5 most delayed routes.