# Airfly Insights Report

## Metrics and Insights

The analysis of the flight delay dataset provided the following key metrics and insights:

1. Average delay time per airline, airport, and route.

2. Percentage of delayed flights.

3. On-time performance rate.

4. Delay distribution by time of day, day of week, and season.

5. Feature importance in delay prediction models.

Insights included identification of delay patterns across different airlines, variation of delays by day of the week and time of the day, cancellation rates, and performance comparison of airlines.

## Data Cleaning Steps

The following data cleaning steps were performed using Pandas:

- Converted 'Date' column to datetime format for time-based analysis.

- Extracted month from the date column to allow seasonal trend analysis.

- Checked for null values and handled them by filling missing airport codes with 'Unknown'.

- Removed duplicate records based on key identifiers such as UniqueCarrier, FlightNum, and AirNum.

- Sampled 10% of the data for exploratory analysis to reduce computation time.

- Optimized memory usage by checking column data types and memory allocation.

These steps ensured the dataset was clean, consistent, and ready for analysis and modeling.