Airlines Dataset – Insights Summary

This report provides a summary of insights derived from the airline dataset explored in the analysis notebook. The dataset primarily consists of flight records with details such as departure times, arrival times, delays, cancellations, and scheduling information. The purpose of this analysis was to clean, explore, and extract meaningful insights from the data to better understand flight performance patterns.

1. Data Cleaning

- Missing values in time columns (e.g., DepTime, ArrTime, CRSArrTime) were handled using formatting and imputation strategies.
- Invalid entries such as '0' or '2400' were corrected or replaced with appropriate placeholders.
- Null values were either forward-filled (ffill) or backward-filled (bfill) based on sequential order when meaningful.
- Columns were standardized into consistent 24-hour time formats (HH:MM).

2. Descriptive Statistics

- Majority of flights departed between the late afternoon and evening hours (16:00–22:00).
- Arrival times showed clustering around late evening, with some delays pushing arrivals past midnight.
- Average delays were observed to be within 10–20 minutes, but outliers with extreme delays were present.
- A small portion of records had missing or corrupted time values, which required preprocessing.

3. Delay Analysis

- Departure delays had a strong correlation with late arrivals, confirming expected scheduling dependencies.
- Peak delays were observed during evening flights, suggesting congestion or air traffic patterns.

- Cancellations and extreme delays were relatively rare but impactful when they occurred.

4. Key Insights

- Data preprocessing is crucial for handling inconsistent time formats and null values.
- Most flights adhered to schedules with only moderate delays.
- Forward and backward filling helped maintain dataset consistency without losing excessive data.
- Insights into delay patterns can assist airlines in improving scheduling and resource allocation.
- The analysis framework can be extended to include weather and airport traffic factors.
- Visualizations of delay trends over seasons could provide deeper insights.
- Incorporating machine learning models may help in predicting delays more accurately.
- Future work may involve integrating passenger satisfaction surveys for a holistic view.