Airfly\_Insights\_Report

**1. Dataset Overview**

The airline delay dataset was explored to gain insights into flight trends, delay behavior, and performance metrics. Initially, the dataset consisted of 484,551 rows and 29 columns. Following data cleaning and feature engineering, the processed dataset now contains 484,549 rows and 33 columns.

**2. Data Cleaning Steps (Using Pandas)**

The following steps were performed to clean and preprocess the dataset: **1. Dataset Overview**

Initial Dataset: 484,551 rows × 29 columns

Final Dataset (after cleaning & feature engineering): 484,549 rows × 33 columns

Total Data Points: 14,051,979

**2. Data Types (Selected Columns)**

• Integer Columns-> ActualElapsedTime, CRSElapsedTime, AirTime, ArrDelay, DepDelay, Distance, TaxiIn, TaxiOut, Cancelled, Diverted, CarrierDelay, WeatherDelay, NASDelay, SecurityDelay, LateAircraftDelay

Object Columns-> Origin, Org\_Airport, Dest, Dest\_Airport, CancellationCode

**3. Missing Values (Raw Data)**

• Org\_Airport → 1,177 null values

Dest\_Airport → 1,479 null values

**5. Distance Statistics**

• Minimum Distance: 31 miles

• Maximum Distance: 4,502 miles

Average Distance: 752.14 miles

**3. Metrics and Insights**

**Raw Dataset Shape:** (484,551, 29)

**Cleaned Dataset Shape:** (484,549, 33)

**Null Values (RAW):** Org\_Airport (1177), Dest\_Airport (1479)

**Null Values (CLEANED):** 0

**Duplicate Rows (RAW):** 2

**Duplicate Rows (FINAL):** 0

**Extra Derived Columns:** Month, DayNumber, Hour, Route

**Minimum Distance:** 31 miles

**Maximum Distance:** 4,502 miles

**Average Distance:** 752.14 miles

**Flights with Distance > 1,000 miles:** Extracted for separate analysis (FlightNum, Origin, Dest, Distance)

**Flights by Day of Week:** 1 → 70,254, 2 → 65,934, 3 → 63,055, 4 → 75,011, 5 → 88,972, 6 → 51,330, 7 → 69,995

**Top 10 Longest Flights:** Extracted with columns Airline, FlightNum, Origin, Dest, Distance

**Average Taxi In Time:** 6.78 minutes

**Average Taxi Out Time:** 19.15 minutes

**Date Column:** Successfully converted to datetime format

**Insights**

Dataset now has extra time-based features (Month, DayNumber, Hour, Route).

All null values are fixed.

Duplicate rows removed.

Minimum Distance: 31 miles, Maximum Distance: 4,502 miles, Average Distance: 752.14 miles.

Flights counted by day of week – highest on Day 5 (88,972 flights).

Long flights (Distance > 1,000 miles) extracted for analysis.

Average Taxi In: 6.78 minutes, Average Taxi Out: 19.15 minutes.

Date column converted to datetime format.

Top 10 longest flights identified.

Final dataset is clean and ready for analysis.