Flight Delay Analysis Report

1. Dataset Description

Source:
• The dataset is available on Kaggle as Flight Delay Prediction dataset.
• It contains data about flights, including details of airlines, routes, departure, arrival, delays, and weather conditions.
• The dataset typically contains several million records with flight-specific features.

Columns:

Column Description
flight_date Date of the flight
airline Airline carrier (e.g., Delta, United, American)
origin_airport Departure airport code
destination_airport Arrival airport code
departure_time Scheduled departure time
arrival_time Scheduled arrival time
departure_delay Minutes of delay at departure
arrival_delay Minutes of delay at arrival
distance Distance between origin and destination
weather Variables capturing weather conditions (temperature, wind, precipitation)

2. Data Quality

• Missing values / Nulls — Some flights may not have complete weather or delay information.
• Outliers — Some delays are extremely large (e.g., more than 1000 minutes) and need to be validated.
• Data type consistency — Times are often stored as strings and need to be converted to datetime; delay columns should be numeric.
• Duplication — Possible duplicate entries for the same flight number and date need to be checked.
• Skewness / distribution — Delay variables are highly skewed since most flights depart on time, but a few have very long delays.
• Data reliability / bias — Delays may be influenced by reporting standards of airlines or missing weather stations.
• Range checks & validity — Negative delay values (early departures/arrivals) must be validated and handled.

3. Operations Performed

Loading the data (via pandas.read_csv)
Inspecting data types and converting types (e.g., time to datetime format, numeric conversion for delays)
Handling missing values:
• Dropping rows with too many missing features
• Imputing missing weather data using mean/median of nearby records

Removing duplicates
Outlier detection and treatment:
• Using IQR and z-score methods to cap extreme delay values

Feature creation / derived metrics:
• Delay flag (Delayed vs. On-time)
• Total travel time = arrival_time – departure_time
• Delay ratio = arrival_delay / scheduled_duration

Exploratory Data Analysis (EDA):
• Univariate analysis (histograms for delay distributions)
• Bivariate analysis (scatter plots: weather vs. delays, distance vs. delays)
• Grouped aggregations (by airline, airport, season)
• Correlation analysis of delay factors

Visualizations: bar charts of airline delays, heatmaps for airport delays, line charts for seasonal delay trends.

4. Key Insights

Airline-wise performance
• Some airlines consistently perform better with fewer delays, while others show higher average delays.

Airport-specific patterns
• Certain hub airports have more delays due to congestion and traffic.
• Weather-affected airports (snow-prone or storm-prone regions) experience higher delay frequencies.

Seasonal & temporal effects
• Delays are higher during peak travel seasons (summer, holidays).
• Evening flights tend to have higher delays compared to morning flights.

Impact of weather
• Strong correlation between adverse weather (storms, snow, heavy rain) and increased delays.
• Clear-weather days show significantly fewer disruptions.

Distance vs. delays
• Longer flights tend to have smaller proportional delays compared to short-haul flights.

Outliers & extremes
• A small subset of flights face extreme delays (over 5–10 hours), heavily affecting averages.

5. Recommendations

1. Optimize scheduling
Airlines should build buffer times into peak-hour schedules to reduce cascading delays.

2. Improve weather prediction integration
Real-time weather data should be integrated into scheduling and passenger notifications.

3. Airport congestion management
High-traffic airports should implement better slot management to minimize ground delays.

4. Seasonal readiness
Airlines should prepare contingency plans for peak holiday seasons and adverse weather months.

5. Passenger communication
Providing timely updates through apps and SMS reduces passenger dissatisfaction during unavoidable delays.

6. Focused operational improvements
Airlines with consistently high delays should review operational processes (crew scheduling, turnaround times).

7. Policy-level interventions

Regulators may enforce performance benchmarks for airlines and airports with high delay records.