Flight Delay Analysis Report

1. Dataset Description

1.1 Source:

• Dataset obtained from **Kaggle – Flight Delay Prediction Dataset**.
• Contains large-scale flight operation data with details on airlines, airports, routes, departure and arrival delays, and weather conditions.
• Dataset includes millions of flight records across various airlines and airports.

1.2 Columns:

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 – Weather data (temperature, wind, precipitation)

1.3 Data Quality:

• Some missing values in weather and delay columns.
• Outliers detected in delay columns (values > 1000 minutes).
• Time columns converted to datetime format for consistency.
• Duplicate records removed using flight number and date combination.
• Delay columns converted to numeric; negative values (early arrivals) validated.

2. Operations Performed

2.1 Data Cleaning & Preparation:

• Loaded dataset using pandas and inspected schema.
• Converted time fields to datetime objects and delays to numeric values.
• Removed duplicates and handled missing weather data using mean/median imputation.
• Detected and capped outliers using IQR and z-score methods.
• Derived new metrics: Delay_Flag, Total_Travel_Time, Delay_Ratio.

2.2 Exploratory Data Analysis (EDA):

• Histograms for delay distributions (departure and arrival).
• Scatter plots for Distance vs Delay and Weather vs Delay.
• Airline-wise and airport-wise group aggregations.
• Correlation matrix of delay, distance, and weather.

2.3 Visualizations:
• Bar charts of average delay per airline.
• Heatmaps for airport delay frequencies.
• Line charts for seasonal delay trends.
• Boxplots for delay distributions by distance.

3. Key Insights

3.1 Airline Performance:
• Certain airlines consistently maintain lower average delays.
• Some carriers face frequent delays, indicating scheduling issues.

3.2 Airport Trends:
• Hub airports experience higher delays due to congestion.
• Weather-prone airports show seasonal spikes in delay duration.

3.3 Temporal & Seasonal Effects:
• Delays increase during holiday and summer seasons.
• Evening flights face higher delay probabilities.

3.4 Weather Impact:
• Strong correlation between adverse weather (storms, snow, rain) and increased delays.
• Clear-weather days show minimal disruptions.

3.5 Distance vs Delay:
• Short-haul flights have higher delay ratios than long-haul ones.
• Longer routes are generally more punctual due to better scheduling buffers.

3.6 Outliers & Extremes:
• A small percentage of flights experience extreme delays (5–10 hours).

4. Recommendations

4.1 Scheduling Optimization:
• Introduce buffer times during peak hours.
• Adjust flight slots dynamically based on real-time air traffic.

4.2 Weather Integration:
• Integrate real-time weather APIs into flight scheduling.
• Provide early alerts to minimize disruptions.

4.3 Airport Congestion Management:

• Improve slot allocation in high-traffic airports.
• Encourage off-peak scheduling for short routes.

4.4 Seasonal Planning:
• Develop contingency strategies for high-delay seasons.
• Allocate extra staff during holidays and summer.

4.5 Passenger Communication:
• Use automated notifications for delay updates.
• Increase transparency to reduce dissatisfaction.

4.6 Operational Improvements:
• Review crew scheduling, turnaround, and ground handling efficiency.

4.7 Future Analytics Opportunities:
• Build predictive models for delay forecasting.
• Use ML for on-time vs delayed flight classification.
• Develop real-time dashboards for monitoring and prediction.