2nd_Milestone_(Visual_Exploration_and_Delay_Trends)

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.

Metrics and Insights

1. Dataset Overview

The airline delay dataset was explored to gain insights into flight trends, delay behavior, and performance metrics.

Initial Dataset: 484,551 rows × 29 columns

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

Total Data Points: 14,051,979

2. Data Cleaning Steps (Using Pandas)

Duplicate Rows Removed: 2 duplicates dropped → 0 remaining

Missing Values (Raw Data):

`Org_Airport`: 1,177 null values

`Dest_Airport`: 1,479 null values

Missing Values (After Cleaning): 0

Flight Volume by Airline

- **Southwest Airlines Co.** leads with **119,048 flights**, the largest market share.

- **American Airlines Inc.** and **American Eagle Airlines Inc.** follow closely, operating **73,053** and **58,698 flights**respectively.

- **Regional airlines** like Skywest and Atlantic Southeast contribute significantly to domestic traffic.

- **Smaller carriers** such as JetBlue and Alaska Airlines handle a limited but consistent flight volume.

📊 **Insight:**
The data indicates that a few large carriers dominate U.S. air traffic, while smaller regional airlines maintain niche routes.

Flight Data Analysis Report

1. Number of Flights per Airline

Airline	Number of Flights
Southwest Airlines Co.	119048
American Airlines Inc.	73053
American Eagle Airlines Inc.	58698
United Air Lines Inc.	56896
Skywest Airlines Inc.	50384
US Airways Inc.	31755
Delta Air Lines Inc.	30220
Atlantic Southeast Airlines	28678
JetBlue Airways	15364
Alaska Airlines Inc.	10000

2. Number of Flights by Day of the Week

Day of Week	Number of Flights
Monday	63761
Tuesday	66659
Wednesday	70353
Thursday	85912
Friday	67718
Saturday	61679
Sunday	68467

📊 **Insights:**

- **Thursday** experiences the **highest number of flights**, possibly due to business and weekend travel overlap.

- **Saturday** has the **lowest flight volume**, typical for reduced business travel.

- Traffic remains stable across weekdays, indicating consistent demand.

3. Number of Flights by Month

Month	Number of Flights
1	87910
2	86967
3	94196
4	68169
5	68071
6	79236

📊 **Insights:**

- **March** records the **highest flight count**, correlating with spring travel and holidays.

- **April and May** show lower numbers — possible off-peak travel months.

- **June** begins a summer rise in flight demand.

4. Top 10 Origin Airports by Number of Flights

Airport	Number of Flights
Chicago O'Hare International Airport	46945
Dallas/Fort Worth International Airport	33027
Hartsfield-Jackson Atlanta International Airport	28834
Denver International Airport	23542
Los Angeles International Airport	17194
McCarran International Airport	15528
San Francisco International Airport	14825
Phoenix Sky Harbor International Airport	13873
Chicago Midway International Airport	9318
Orlando International Airport	9043

5. Top 10 Routes by Number of Flights

Route	Number of Flights
ORD-LGA	1920
LGA-ORD	1615
LAX-SFO	1603
SFO-LAX	1457
LAS-LAX	1305
HOU-DAL	1276
DAL-HOU	1200
ORD-LAX	1154
PHX-LAS	1152
DFW-ORD	1125

📊 **Insights:**

- The **ORD–LGA (Chicago to New York)** route is the **most traveled** corridor.

- High-frequency routes mostly connect **business hubs** or **major metros**.

- Many top routes show **reciprocal traffic patterns**, e.g., LAX–SFO and SFO–LAX.

🛫 6. Top 10 Routes by Number of Flights

Route	Flights
ORD–LGA	1,920
LGA–ORD	1,615
LAX–SFO	1,603
SFO–LAX	1,457
LAS–LAX	1,305
HOU–DAL	1,276
DAL–HOU	1,200
ORD–LAX	1,154
PHX–LAS	1,152
DFW–ORD	1,125

7. Average Arrival Delay vs Departure Hour

- Average delay time tends to **increase during peak hours** (late afternoon and evening).

- **Morning flights (6–9 AM)** usually have **shorter delays**.

8. Total Delay Time by Type

Delay Categories:

- **CarrierDelay** → Airline-related (maintenance, crew issues).

- **WeatherDelay** → Weather-related conditions.

- **NASDelay** → National Air System (air traffic, airport ops).

- **SecurityDelay** → Security checks or incidents.

- **LateAircraftDelay** → Aircraft arriving late from previous trip.

📊 **Insights:**

- **Carrier Delay** and **Late Aircraft Delay** are **primary contributors** to total delay time.

- **Weather delays** are **less frequent** but **more severe** when they occur.

- **Security delays** are rare and have minimal impact on total delay time.

9. Average Arrival Delay vs Departure Hour

- Early morning departures (**6–9 AM**) experience **least delay**, benefiting from unclogged airspace.

- **Late afternoon and evening flights (4–10 PM)** face **higher delays**, largely due to cumulative daily disruptions.

- **Peak-hour congestion** and **aircraft reuse schedules** amplify late-day delay durations.

📊 **Strategic Implication:**
Scheduling optimization and buffer planning during peak hours could reduce average delay times.

Insights Summary

- Southwest Airlines operates the highest number of flights (119,048), showing strong market presence.

- Flight traffic peaks on **Thursdays (85,912 flights)**, followed by Wednesdays and Fridays.

- **March (94,196 flights)** records the highest monthly flight count, indicating seasonal travel spikes.

- **Chicago O'Hare International Airport (ORD)** is the busiest origin airport, followed by Dallas/Fort Worth (DFW) and Atlanta (ATL).

- The **most frequent route** is **ORD–LGA** (Chicago to New York), highlighting a major business corridor.

- Delay data shows varied causes — **Carrier Delay**, **Weather Delay**, **NAS Delay**, **Security Delay**, and **Late Aircraft Delay** — influencing total performance.

- Distribution analysis shows both departure and arrival delays vary with time of day, implying congestion patterns.

- Dataset is clean, enriched with delay categories and ready for visualization and deeper trend modeling.