

Flight Delay Data Analysis Report

Dataset: Flight_delay.csv Total Records: 484,551

1. Introduction and Data Overview

This report summarizes the structure and key insights derived from the provided flight delay dataset, which contains information on **484,551** flight records. The dataset includes details on flight schedules, actual times, various delay components, and airline/airport information.

1.1. Dataset Structure and Integrity

The dataset contains **29 columns** covering all aspects of flight operations and delays.

Column Category	Example Columns	Description	Data Type
Flight/Time	Date, DepTime, ArrTime, DayOfWeek	Scheduling and actual time of flight segments.	Mixed
Identifiers	UniqueCarrier, Airline, FlightNum, TailNum	Codes and names for the airline and aircraft.	Object/Int
Performance	ActualElapsedTime, CRSElapsedTime, AirTime	Measured and scheduled flight duration in minutes.	int64
Delays	ArrDelay, DepDelay	Total arrival and departure delay in minutes.	int64
Locations	Origin, Org_Airport, Dest, Dest_Airport	Airport codes and full names for departure and arrival.	Object
Delay Causes	CarrierDelay, WeatherDelay, NASDelay, SecurityDelay, LateAircraftDelay	Breakdown of delay into five contributing factors.	int64
Status	Cancelled, Diverted	Binary flags for flight status (0=No, 1=Yes).	int64

2. Key Summary Statistics

This section provides descriptive statistics for critical numerical metrics, focusing on flight delays, durations, and distance.

2.1. Delay Statistics (in Minutes):

Metric	Arrival Delay (Ar	Departure Delay (DepDelay)	
Average (μ)	60.91	57.5	
Median (50th %ile)	42	40	
Standard Deviation (σ):	56.98	55.99	
Maximum	1707	1710	

- The average flight in this dataset experienced an **arrival delay of approximately 61 minutes**.
- The minimum delays are positive, indicating the dataset focuses primarily on flights that were already delayed.

2.2. Flight Time and Distance:

Metric	Actual Elapsed	Scheduled Elaps	Distance (miles)
Average (μ)	134.81	131.4	752.14
Median (50th %i	116	114	599

3. Detailed Delay Analysis

3.1. Primary Causes of Delay:

This analysis breaks down the average delay into its five contributing components.

Delay Component	Average Delay (min)	Approximate Percentage of Total Delay		
Late Aircraft Delivery	26.65	≈43.7%		
Carrier Delay	17.42	≈28.6%		
NAS Delay (Airspace)	13.6	≈22.3%		
Weather Delay	3.15	≈5.2%		
Security Delay	0.08	≈0.1%		

- **Late Aircraft Delay** is the dominant factor, accounting for nearly 44% of the average total delay, signifying that late arrivals are often caused by preceding flights arriving late.
- **Carrier Delay** (airline-specific issues like maintenance) and **NAS Delay** (National Airspace System issues like traffic congestion) are the next most significant causes.

3.2. Airline Performance Metrics

- Top 5 Airlines by Flight Volume:

Airline Name	Flight Count
Southwest Airlines	119,048
American Airlines	73,053
American Eagle	58,698
United Air Lines	56,896
Skywest Airlines	50,384

- Average Arrival Delay by Airline (Top & Bottom 5):

Worst Performing	Avg. Arrival Delay (min)	Best Performing	Avg. Arrival Delay (min)
JetBlue Airways	72.87	Frontier Airlines	41.97
United Air Lines	69.67	Southwest Airlines	51.03
American Airlines	65.73	Hawaiian Airlines	55.66
Skywest Airlines	65.19	Alaska Airlines Inc.	57.56
American Eagle	64.28	US Airways Inc.	58.45

- Frontier Airlines Inc. has the lowest average delay (41.97 minutes), while JetBlue Airways has the highest (72.87 minutes).

4. Airport Traffic Analysis

- This section summarizes the origin airports with the highest volume of flights recorded in the dataset.
- Top 5 Origin Airports by Flight Volume:

Airport Name	City/State	Flight Count
Chicago O'Hare	Chicago, IL	46,945
Dallas/Fort Worth	Dallas/Fort Worth	33,027
Hartsfield-Jackson	Atlanta, GA	28,834
Denver International	Denver, CO	23,543
Los Angeles International	Los Angeles, CA	17,194

- Chicago O'Hare International Airport is the busiest origin airport in this dataset, indicating it is a significant hub for flight operations and potential source of delays.

Conclusion of Flight Delay Analysis:

The dataset reveals that the analyzed flights experienced significant delays, with an overall average arrival delay of approximately 61 minutes.

The single most critical driver of this systemic lateness is the Late Aircraft Delay, which accounts for nearly 44% of the average delay. This indicates that operational efficiency is primarily hampered not by individual flight issues, but by a cascading effect where a late arrival of one flight segment causes subsequent flights using the same aircraft to also be delayed. This points to a need for better buffer times or improved aircraft scheduling and turnaround procedures.

- Operational Performance: High-volume carriers like Southwest Airlines Co. manage to maintain a below-average delay, while JetBlue Airways and United Air Lines Inc. demonstrate the highest average delays, suggesting room for improvement in their operational resilience and recovery strategies.
- Infrastructure Impact: The high traffic volume at major hubs, particularly Chicago O'Hare International Airport, correlates with the significance of NAS Delay (National Airspace System/congestion) as a top-three delay factor, suggesting that air traffic control and airport capacity constraints contribute to the overall delay problem.