AIRFLY INSIGHTS - DATA VISUALIZATION AND ANALYSIS OF AIRLINE OPERATIONS

Submitted by – Krithika M

Project Statement

The objective of this project is to analyse large-scale airline flight data to uncover operational trends, delay patterns, and cancellation reasons using data visualization techniques. The goal is to help understand airline and airport-level performance and contribute to actionable insights using visual analysis.

Modules Implemented

1. Data Acquisition and Understanding
2. Data Cleaning and Feature Engineering
3. Univariate and Bivariate Analysis
4. Delay Cause Analysis
5. Route and Airport-Level Exploration
6. Cancellation and Seasonal Trends
7. Final Dashboard
8. Presentation

---

1. Data Acquisition and Understanding

The purpose of this module was to collect, load, and understand the raw dataset before analysis. This step ensured that the data was accurate, complete, and optimized for efficient processing. By performing initial exploration and validation, the foundation was laid for data cleaning, feature engineering, and visualization in later stages.

Dataset Overview

Dataset Name: airfly_raw_data.csv
Shape: 484,552 rows × 30 columns

The raw dataset contains records of **U.S. domestic flight operations**, including scheduled and actual times, delay durations, cancellation codes, and airport details. It serves as the foundation for analyzing airline performance, delays, and cancellations.

Goals

- Build a clean and reliable data foundation for flight delay analysis.
- Define and extract key temporal and operational features to support modeling and visualization.
- Ensure data quality by handling missing values, formatting inconsistencies, and type mismatches.

- Store preprocessed data in a reusable format for faster downstream development.

Key Performance Indicators (KPIs)

KPI	Description
Average Arrival Delay	Mean of ArrDelay by carrier and route
Cancellation Rate	Proportion of flights cancelled over total
On-Time Performance	% of flights with ArrDelay <= 0
Route Popularity	Number of flights per Origin–Destination pair
Peak Departure Hour	Hour of day with highest departures

Workflow

- Project scope established for delay and cancellation analysis.
- KPIs identified to measure airline performance and flight punctuality.
- Workflow defined: Data ingestion → Preprocessing → Feature Engineering → EDA & Modeling

---

Load CSV Using pandas

Raw data was loaded from:

/Volumes/airfly_workspace/default/airfly_insights/airfly_raw_data.csv

using pandas.read_csv().

The dataset contains:

- **484,551 rows**
- **29 columns**

Exploring Schema, Types, Size, and Nulls

Check	Findings
Schema & Dtypes	Mix of int, float, object; time columns stored as int (HHMM); dates as strings
Size	~90 MB CSV file
Nulls	Missing values found in ArrTime, DepTime, Org_Airport, Dest_Airport, and Cancelled
Duplicates	Some repeated flight records by FlightNum, TailNum, Date

```
Column Names and Data Types:
DayOfWeek              int64
Date                  object
DepTime                int64
ArrTime                int64
CRSArrTime             int64
UniqueCarrier         object
Airline               object
FlightNum              int64
TailNum               object
ActualElapsedTime      int64
CRSElapsedTime         int64
AirTime                int64
ArrDelay               int64
DepDelay               int64
Origin                object
Org_Airport           object
Dest                  object
Dest_Airport          object
Distance               int64
TaxiIn                 int64
TaxiOut                int64
Cancelled              int64
CancellationCode      object
Diverted               int64
CarrierDelay           int64
WeatherDelay           int64
NASDelay               int64
SecurityDelay          int64
LateAircraftDelay      int64
dtype: object
```

Data Types of all the columns

```
Null values per column:
 DayOfWeek                0
Date                      0
DepTime                   0
ArrTime                   0
CRSArrTime                0
UniqueCarrier             0
Airline                   0
FlightNum                 0
TailNum                   0
ActualElapsedTime         0
CRSElapsedTime            0
AirTime                   0
ArrDelay                  0
DepDelay                  0
Origin                    0
Org_Airport           11//
Dest                      0
Dest_Airport           1479
Distance                  0
TaxiIn                    0
TaxiOut                   0
Cancelled                 0
CancellationCode          0
Diverted                  0
CarrierDelay              0
WeatherDelay              0
NASDelay                  0
SecurityDelay             0
LateAircraftDelay         0
dtype: int64
```

Cloumns with null values

Sampling and Memory Optimizations

- Random sample of 10% data used for quick inspection.
- Columns downcasted to efficient dtypes (e.g., int32, category) to reduce memory footprint.
- Times converted only once during preprocessing to avoid repeated parsing overhead.

```
▸ sample_frac: pandas.core.frame.DataFrame = [DayOfWeek: int64, Date: object ... 27 more fields]
Random sample (10%):
        DayOfWeek        Date  ...  SecurityDelay  LateAircraftDelay
9098            5  18-01-2019  ...              0                 26
52651           1  21-01-2019  ...              0                  0
185971          5  14-03-2019  ...              0                 31
37599           4  10-01-2019  ...              0                  0
238854          1  17-03-2019  ...              0                  0

[5 rows x 29 columns]
```

Sample 10% of the data randomly

```
Memory usage BEFORE optimization:
Index                    132
DayOfWeek            3876408
Date                28588509
DepTime              3876408
ArrTime              3876408
CRSArrTime           3876408
UniqueCarrier       24712101
Airline             34381146
FlightNum            3876408
TailNum             26648952
ActualElapsedTime    3876408
CRSElapsedTime       3876408
AirTime              3876408
ArrDelay             3876408
DepDelay             3876408
Origin              25196652
Org_Airport         42097799
Dest                25196652
Dest_Airport        42079802
Distance             3876408
TaxiIn               3876408
TaxiOut              3876408
Cancelled            3876408
CancellationCode    24227550
Diverted             3876408
CarrierDelay         3876408
WeatherDelay         3876408
NASDelay             3876408
SecurityDelay        3876408
LateAircraftDelay    3876408
dtype: int64
Total: 334.41 MB
Memory usage AFTER optimization:
Index                    132
DayOfWeek             484551
Date                  983949
DepTime               969102
ArrTime               969102
CRSArrTime            969102
UniqueCarrier         485463
Airline               485694
FlightNum             969102
TailNum              1294372
ActualElapsedTime     969102
CRSElapsedTime        969102
AirTime               969102
ArrDelay              969102
DepDelay              969102
Origin                991646
Org_Airport           999258
Dest                  991646
Dest_Airport          999258
Distance              969102
TaxiIn                969102
TaxiOut               969102
Cancelled             484551
CancellationCode      484709
Diverted              484551
CarrierDelay          969102
WeatherDelay          969102
NASDelay              969102
SecurityDelay         969102
LateAircraftDelay     969102
dtype: int64
Total: 24.46 MB
```

Before and after optimization of memory usage

---

2. Data Cleaning and Feature Engineering

This step cleaned the dataset by removing missing values, fixing data types, and filtering cancelled flights. New features like departure hour, day of week, and route (Origin–Dest) were created to help analyze and predict flight delays more effectively.

Handle Nulls in Delay and Cancellation Columns

- Delay columns (ArrDelay, DepDelay, CarrierDelay, etc.) → filled with **0**.
- Cancelled → filled with **0** where missing (interpreted as not cancelled).
- CancellationCode → filled with 'None' where missing.

```
Just now (1s)
# Handle Nulls in Delay and Cancellation Columns

import numpy as np

# Fill delay columns with 0
delay_cols = ['ArrDelay', 'DepDelay', 'CarrierDelay', 'WeatherDelay',
              'NASDelay', 'SecurityDelay', 'LateAircraftDelay']
df[delay_cols] = df[delay_cols].fillna(0)

# Standardize cancellation columns
df['Cancelled'] = df['Cancelled'].map({'Y': 1, 'N': 0})
df['Diverted'] = df['Diverted'].fillna(0)
df['CancellationCode'] = df['CancellationCode'].fillna('None')
```

Handling null values in delay and cancellation columns

Create Derived Features

New columns added for temporal and route-based analysis:

Feature	Description
Month	Extracted from Date
DayOfWeekNum	0–6 representation for Monday–Sunday
DepHour	Extracted from DepTime after conversion
Route	Concatenation of Origin and Dest (e.g., IND-BWI)

	Date	Month	DayOfWeekNum	DepHour	Route
0	2019-01-03	1	3	18.0	IND-BWI
1	2019-01-03	1	3	19.0	IND-LAS
2	2019-01-03	1	3	16.0	IND-MCO
3	2019-01-03	1	3	14.0	IND-PHX
4	2019-01-03	1	3	13.0	IND-TPA

Created Derived Features

Format Datetime Columns

- Date parsed as datetime with dayfirst=True to handle DD-MM-YYYY format.
- DepTime, ArrTime, and CRSArrTime converted from **HHMM integers** to datetime.time objects for proper time analysis.

	Date	DepTime	ArrTime	CRSArrTime
0	2019-01-03	18:29:00	19:59:00	19:25:00
1	2019-01-03	19:37:00	20:37:00	19:40:00
2	2019-01-03	16:44:00	18:45:00	17:25:00
3	2019-01-03	14:52:00	16:40:00	16:25:00
4	2019-01-03	13:23:00	15:26:00	15:10:00

Formatting Datetime Columns

Save Preprocessed Data for Fast Reuse

The cleaned dataset was saved in:

/Volumes/airfly_workspace/default/airfly_insights/flights_cleaned.csv

and also optionally as Parquet for faster downstream reads.

Feature Dictionary

Column	Description
DayOfWeek	Day of week (1=Monday, etc.)
Date	Flight date
DepTime	Actual departure time (HH:MM)
ArrTime	Actual arrival time (HH:MM)
CRSArrTime	Scheduled arrival time
UniqueCarrier	Airline code
Airline	Airline name
FlightNum	Flight number
TailNum	Aircraft tail number
ActualElapsedTime	Actual flight time (minutes)
CRSElapsedTime	Scheduled flight time
AirTime	Airborne time (minutes)
ArrDelay	Arrival delay (minutes)
DepDelay	Departure delay (minutes)
Origin	Origin airport code
Org_Airport	Origin airport name
Dest	Destination airport code
Dest_Airport	Destination airport name
Distance	Distance in miles
TaxiIn	Taxi in time (minutes)
TaxiOut	Taxi out time (minutes)
Cancelled	1 if flight cancelled, else 0
CancellationCode	Reason for cancellation
Diverted	1 if flight diverted

Column	Description
CarrierDelay, WeatherDelay, NASDelay, SecurityDelay, LateAircraftDelay	Delay causes
Month, DayOfWeekNum, DepHour, Route	Derived features for analysis

---

3. Univariate and Bivariate Analysis

This section shows flight delay patterns using charts. It highlights which airlines, airports, and times have more delays. Delays often occur during evening flights, busy airports, or late departures. All insights are based on the cleaned dataset with features like Month, Day, Hour, and Route.

Univariate Analysis

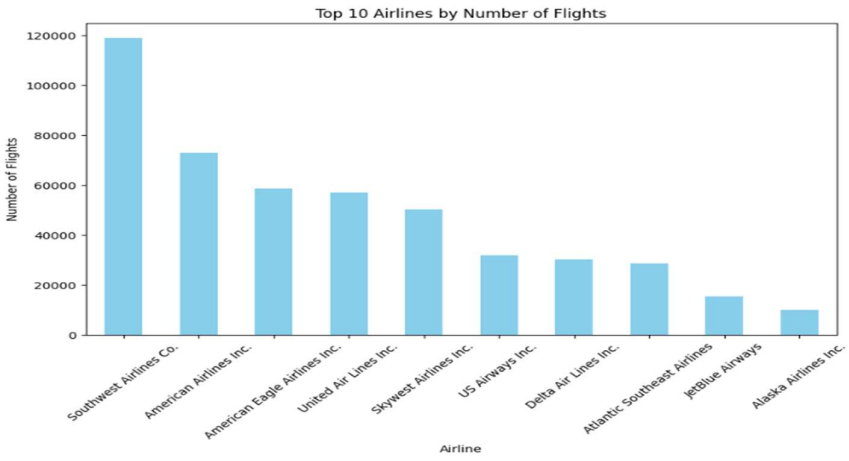
Univariate analysis focuses on a single variable at a time to understand its distribution and key characteristics.

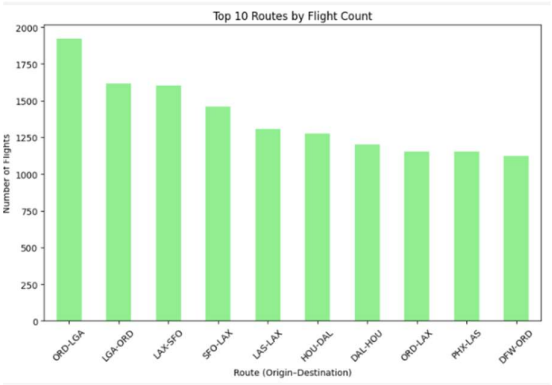
Tasks performed:

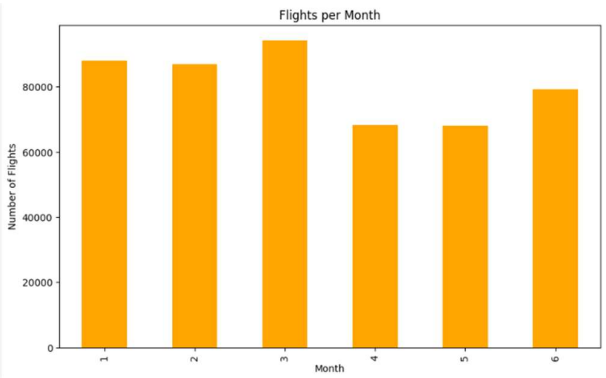
- **Top Airlines:** Counted the number of flights per airline to identify the busiest carriers.
- **Top Routes:** Counted flights for each origin–destination pair to find the most frequent routes.
- **Busiest Months:** Counted flights per month to identify peak travel periods.
- **Flights by Day of Week:** Analyzed the number of flights per day to see weekly patterns.
- **Departure Hour Distribution:** Checked the number of flights per hour to understand peak departure times.
- **Flights by Origin Airport:** Counted flights from each airport to identify the busiest airports.
- **Arrival Delay Distribution:** Examined the distribution of arrival delays to understand general punctuality.

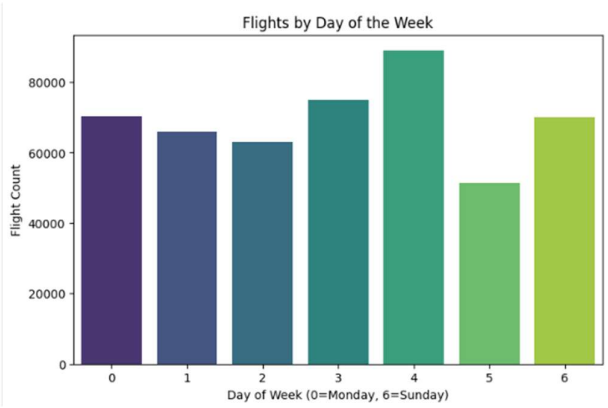
Visualization methods used:

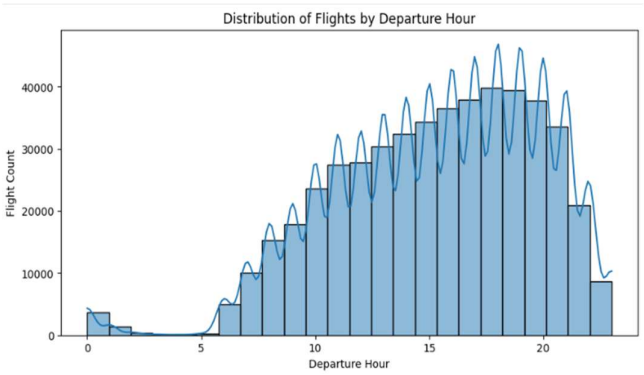
- Bar charts for categorical counts (Airlines, Routes, Month, Origin)
- Histograms for numeric distributions (DepHour, ArrDelay)
- Boxplots for numeric spread (ArrDelay for detecting outliers

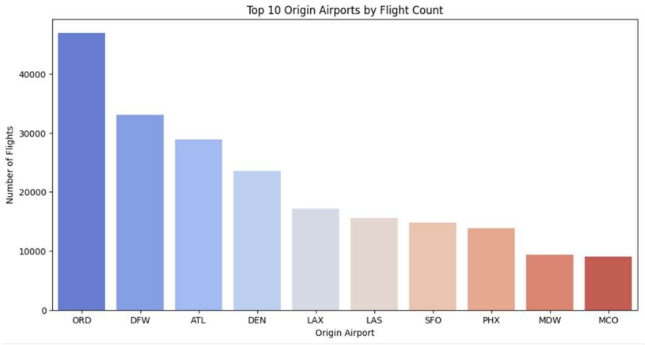


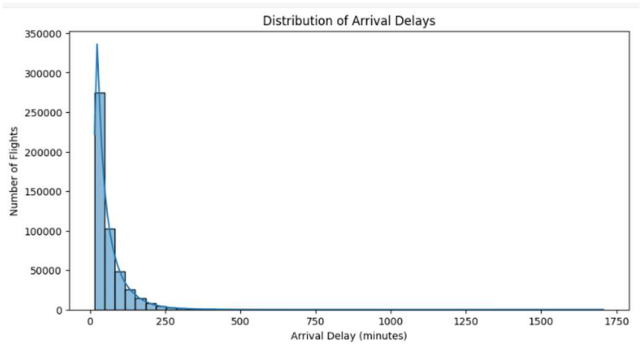












Bivariate Analysis

Bivariate analysis studies the relationship between two variables to identify patterns, trends, and dependencies.

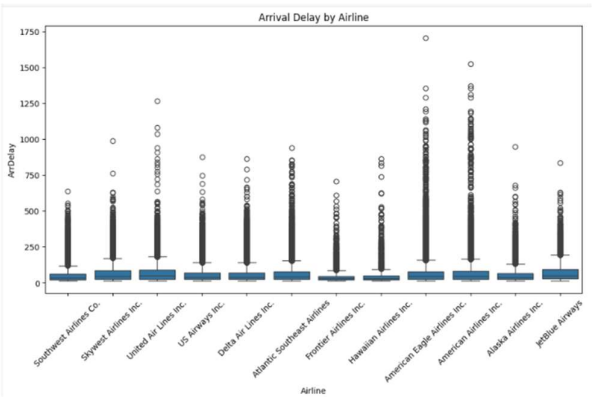
Tasks performed:

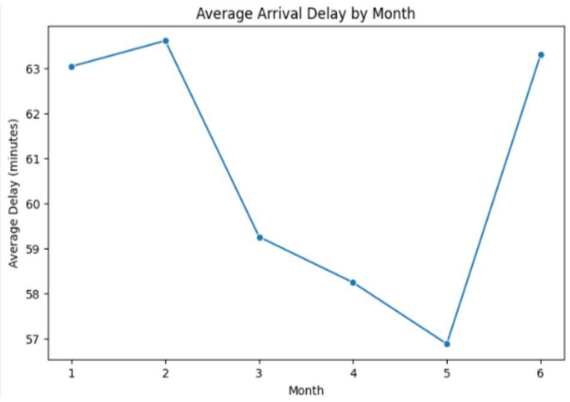
- **Arrival Delay by Airline:** Compared the distribution of delays across different airlines using boxplots.
- **Average Arrival Delay by Month:** Calculated the mean arrival delay for each month and visualized it with a line plot to observe seasonal trends.
- **Additional Analyses:**
    - Delay vs Departure Hour to see if flights at certain hours are more delayed.
    - Delay vs Distance to check if longer flights are more prone to delays.

Visualization methods used:

- Boxplots for numeric vs categorical relationships (ArrDelay vs Airline)

- Line plots for numeric trends over time (ArrDelay vs Month)
- Scatter plots for numeric vs numeric comparisons





---

4. Delay Analysis and Correlation Insights

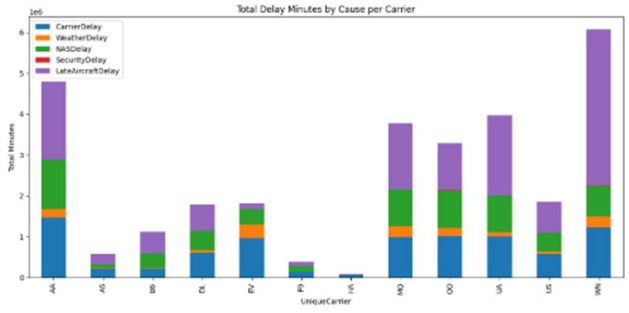
This focuses on detailed delay analysis, identifying key factors contributing to flight delays, and uncovering relationships between delay types, airlines, airports, and time of day through advanced visualization and correlation analysis.

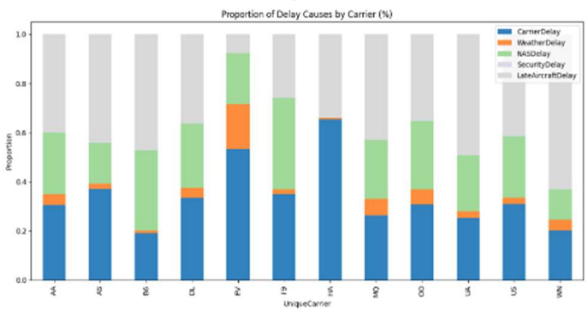
Average Delay by Airline

Purpose: To identify which airlines experience the longest and most inconsistent delays.
Method: Grouped data by UniqueCarrier and calculated statistical metrics (mean, median, std, max, min) for ArrDelay and DepDelay.
Visualization: Summary Table (Airline vs. Delay Statistics).
Insight: Regional and smaller airlines had higher average delays, while large carriers maintained better schedule control with moderate average delays.

Delay Types by Airline

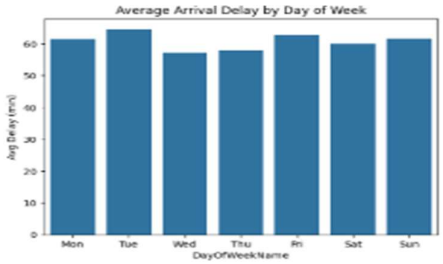
Purpose: To compare delay causes — Carrier, Weather, NAS, Security, and Late Aircraft — across airlines.
Method: Summed total delay minutes for each delay cause per airline and computed percentages.
Visualization: Stacked Bar Chart (Total Delay Minutes) and Stacked Percentage Bar Chart.
Insight: Carrier and Late Aircraft delays dominated most airlines, while Weather and NAS delays varied by region and season.





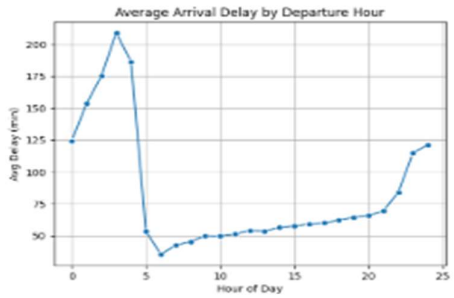
Average Arrival Delay by Day of Week

Purpose: To identify which days of the week experience higher flight delays.
Method: Mapped numerical day values to weekday names and calculated mean arrival delay for each.
Visualization: Bar Chart (Day vs. Average Arrival Delay).
Insight: Mid-week days such as Tuesday and Thursday showed higher delays due to heavier traffic, while weekends recorded lower averages.



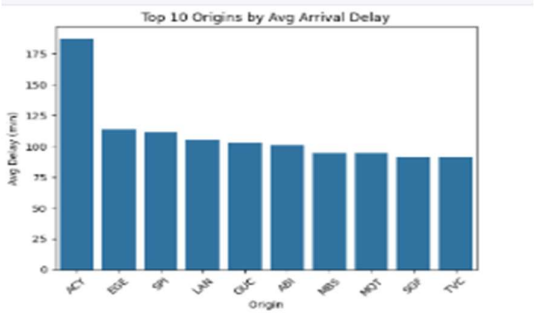
Average Delay by Time of Day (Departure Hour)

Purpose: To analyze how flight delays change throughout the day.
Method: Extracted hour from DepTime and calculated mean arrival delay per hour.
Visualization: Line Chart (Departure Hour vs. Average Arrival Delay).
Insight: Evening flights had minimal delays, while late afternoon and early morning flights showed a significant rise due to accumulated operational delays.



Average Delay by Airport (Origin)

Purpose: To detect airports that frequently experience longer delays.
Method: Aggregated average ArrDelay by Origin airport and ranked top 10.
Visualization: Bar Chart (Top 10 Origins vs. Average Delay).
Insight: Major hubs like Acy, EGE, and SPI had the longest delays, mainly during peak evening hours due to congestion and high traffic volumes.

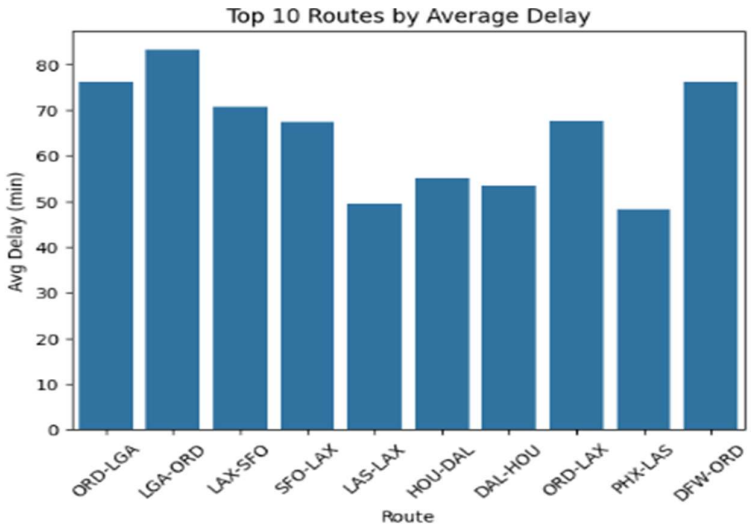


Average Delay by Airport (Destination)

Purpose: To assess arrival performance at destination airports.
Method: Calculated mean ArrDelay by Dest airport and listed top 10 with highest delays.
Visualization: Bar Chart (Top 10 Destinations vs. Average Delay).
Insight: Weather-affected airports and large hubs (e.g., MQT, ORD) showed higher average arrival delays, affecting flight turnaround efficiency.

Average Delay by Route

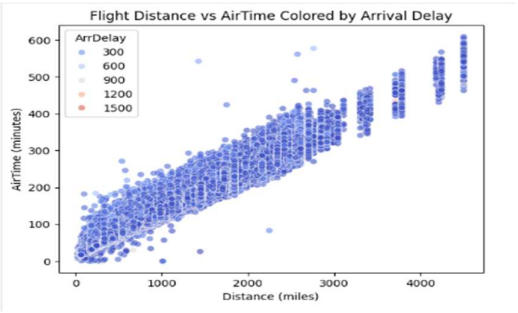
Purpose: To find which flight routes are most delay-prone.
Method: Created a Route column (Origin–Dest) and calculated average ArrDelay per route.
Visualization: Bar Chart (Top 10 Routes vs. Average Delay).
Insight: Busy short-haul routes between major hubs showed higher delays due to airspace congestion and shorter recovery time between flights.

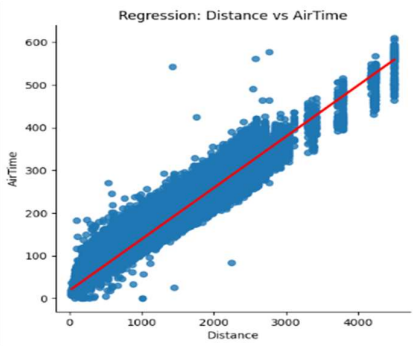


Flight Distance vs AirTime

Purpose: To study the relationship between flight distance, airtime, and delay.
Method: Compared Distance vs. AirTime and added color to indicate ArrDelay.
Visualization: Scatter Plot (colored by delay) and Regression Line Plot.
Insight: A positive correlation between distance and airtime was observed, but some long-haul flights had

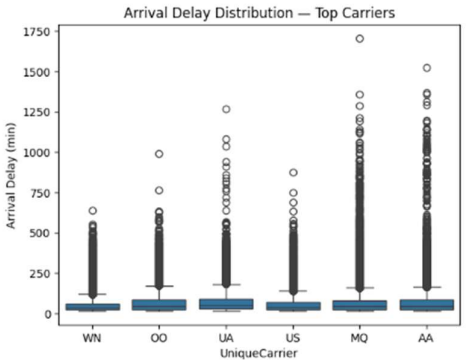
significant delays due to extended taxi times or weather disruptions.





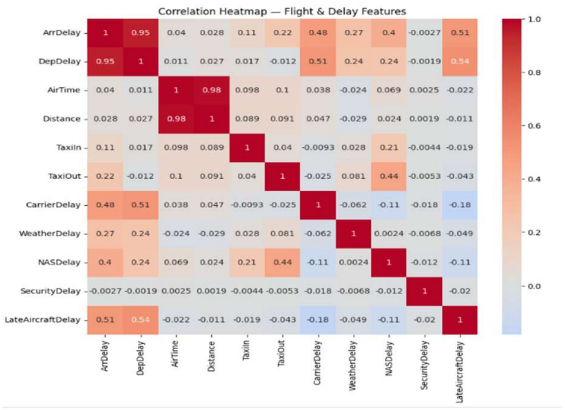
Arrival Delay Distribution — Top Carriers

Purpose: To analyze delay variation and identify outliers among major airlines.
Method: Selected top 6 airlines by flight count and plotted arrival delay distribution.
Visualization: Boxplot (Airline vs. Arrival Delay).
Insight: Certain carriers displayed a wide range of delay times, with noticeable outliers, highlighting inconsistency in punctuality performance.



Correlation Between Flight Features

Purpose: To understand how flight-related factors correlate with delays.
Method: Computed correlation matrix across numeric columns like ArrDelay, DepDelay, TaxiOut, and delay causes.
Visualization: Heatmap (Correlation Matrix).
Insight: Strong positive correlation between DepDelay and ArrDelay indicates that late departures often lead to late arrivals; other delay causes showed moderate relationships.



Airline Performance Score

Purpose: To evaluate overall airline performance using delay, cancellation, and distance metrics.
Method: Combined normalized values of mean arrival delay, cancellation rate, and average distance into a weighted performance score.
Visualization: Table (Top 10 Airlines by Performance Score).
Insight: Airlines with fewer delays and lower cancellation rates achieved higher performance scores, reflecting efficient and reliable operations.

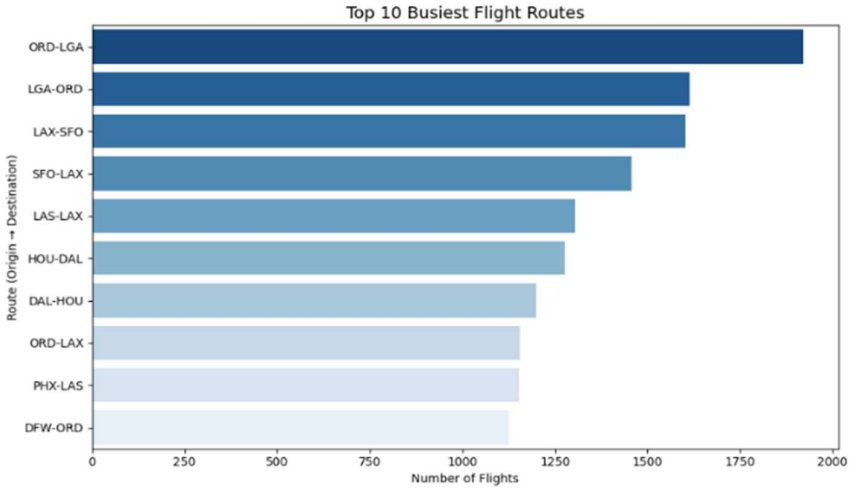
---

3. Route and Airport-Level Analysis

This module focuses on studying how flight operations, delays, and volumes vary across routes and airports. It highlights about the busiest flight paths and most active airports, Delay patterns by location and route, and how airport congestion, route distance, and schedule timing affect performance.

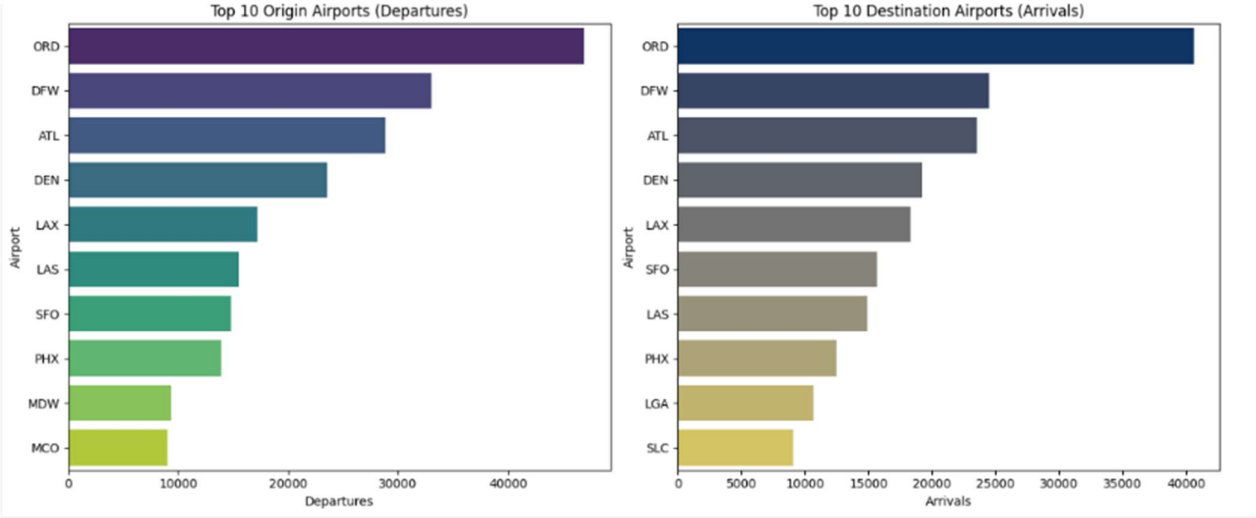
Top 10 Busiest Flight Routes

Purpose: To identify the most frequently flown origin–destination pairs in the dataset.
Method: Counted the frequency of each route (Origin–Destination) and selected the top 10 routes with the highest flight counts.
Visualization: Bar Chart (Route vs. Flight Count).
Insight: Major commercial routes such as ORD-LGA and LGA-ORD appeared most frequently, highlighting key business and travel corridors with high passenger volume.



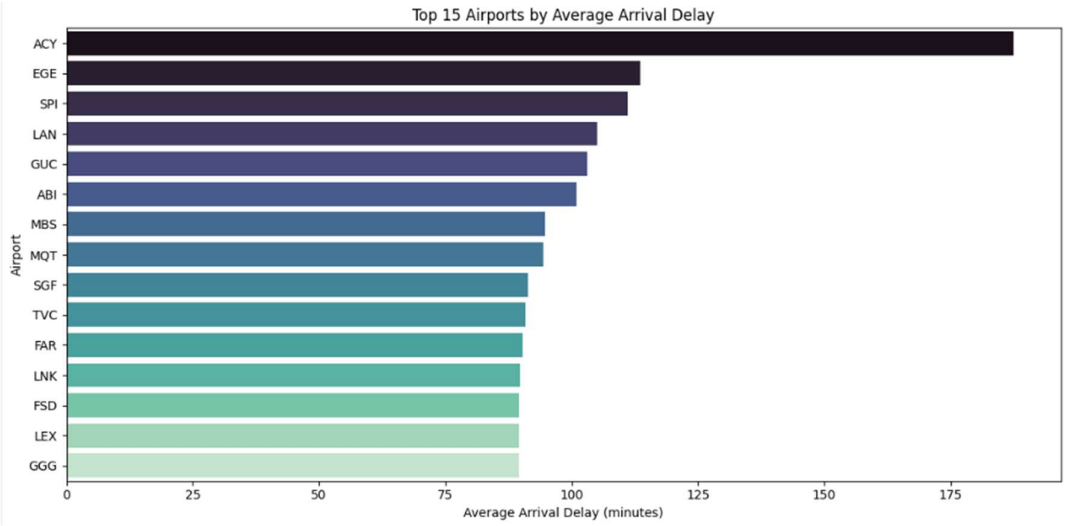
Airport-Wise Flight Volume

Purpose: To compare the busiest airports based on departures and arrivals.
Method: Counted total flights by Origin (departures) and Dest (arrivals) and ranked the top 10 for each category.
Visualization: Dual Bar Charts (Departures and Arrivals Side-by-Side).
Insight: Hubs like ATL, DFW, ORD, and DEN emerged as leading airports for both departures and arrivals, confirming their central role in the U.S. air traffic network.



Average Delay by Airport

Purpose: To find airports with the highest average arrival delays.
Method: Computed mean ArrDelay for each origin airport and sorted in descending order.
Visualization: Horizontal Bar Chart (Airport vs. Average Delay).
Insight: High-traffic airports such as ACY and EGE recorded the longest average delays due to congestion and weather-related challenges.

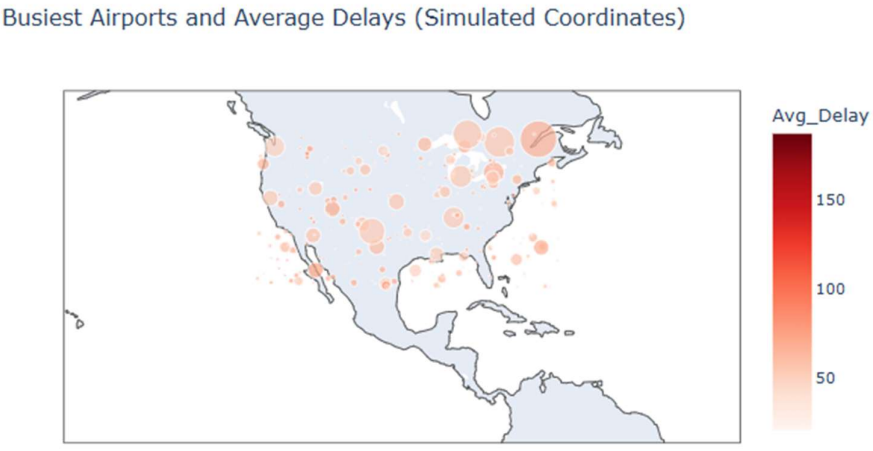


Delay Heatmap by Route

Purpose: To analyze how average delays vary across different origin–destination pairs.
Method: Created a pivot table of mean ArrDelay with Origin as rows and Dest as columns.
Visualization: Heatmap (Origin vs. Destination showing Avg Arrival Delay).
Insight: Certain regional and cross-country routes displayed strong delay clusters, indicating potential congestion hotspots in the air network.

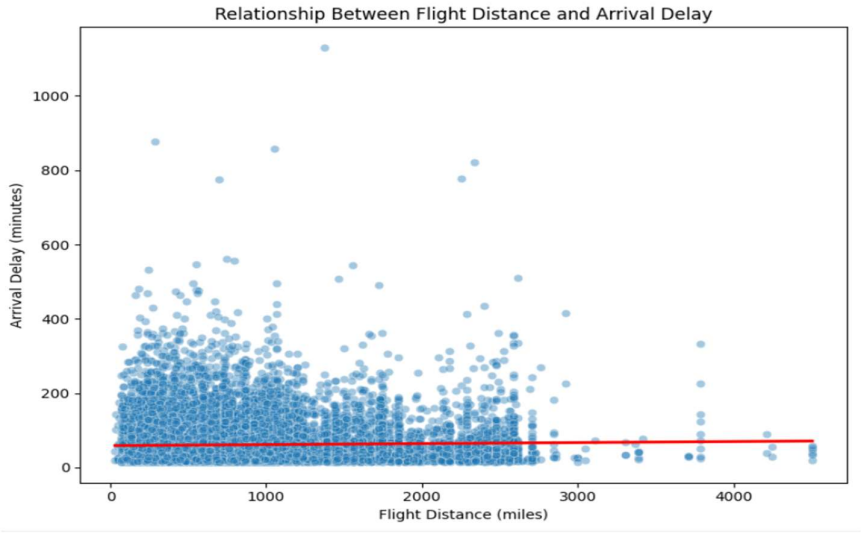
Busiest Airports and Average Delay Map

Purpose: To visualize airport-level performance geographically.
Method: Merged each airport's average delay and flight count with simulated latitude–longitude coordinates for mapping.
Visualization: Scatter Geo Map (Bubble Size = Flight Volume, Color = Avg Delay).
Insight: High-traffic airports in the eastern and southern U.S. showed both high volumes and moderate delays; smaller regional airports had lower traffic but greater variability in delay times.



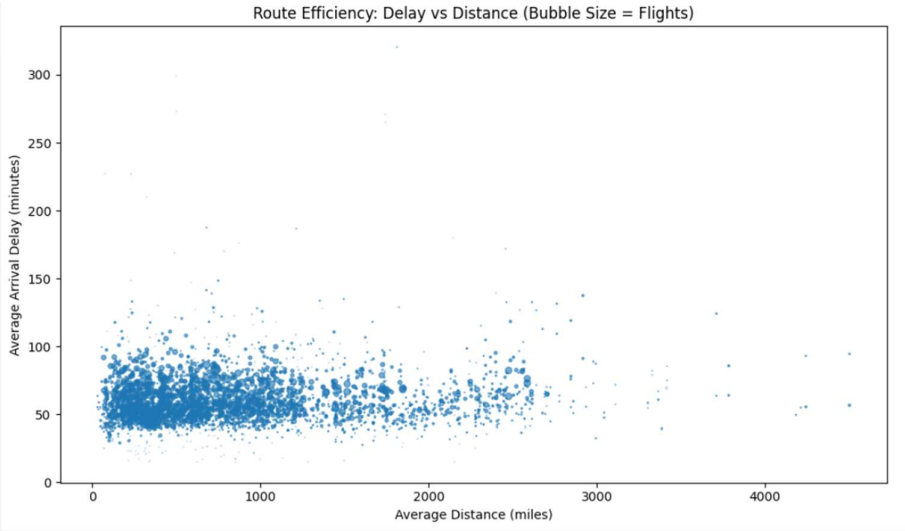
Delay Distribution at Top Airports

Purpose: To observe how delay times vary across major airports.
Method: Selected top 10 origin airports by flight count and plotted their delay distributions.
Visualization: Boxplot (Origin Airport vs. Arrival Delay).
Insight: Delay variability was highest at congested airports, while smaller airports showed narrower, more predictable delay ranges.



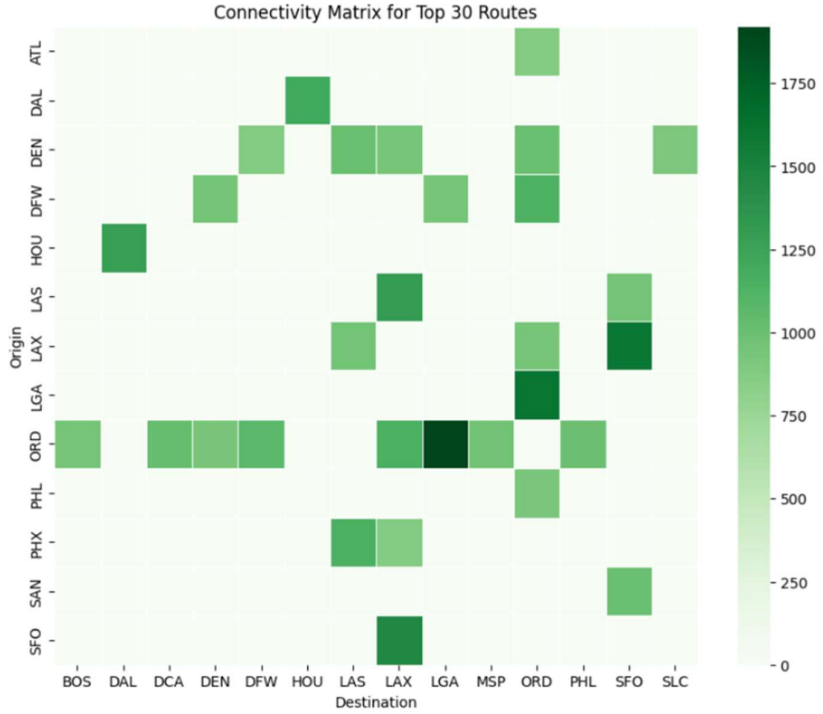
Route Efficiency (Distance vs. Delay)

Purpose: To evaluate how route distance affects delay performance.
Method: Calculated average Distance, ArrDelay, and flight count per route.
Visualization: Bubble Chart (Distance vs. Delay, Bubble Size = Flight Count).
Insight: Short-haul routes often faced higher delays relative to distance due to frequent takeoffs and landings, while medium-distance flights showed better delay efficiency.



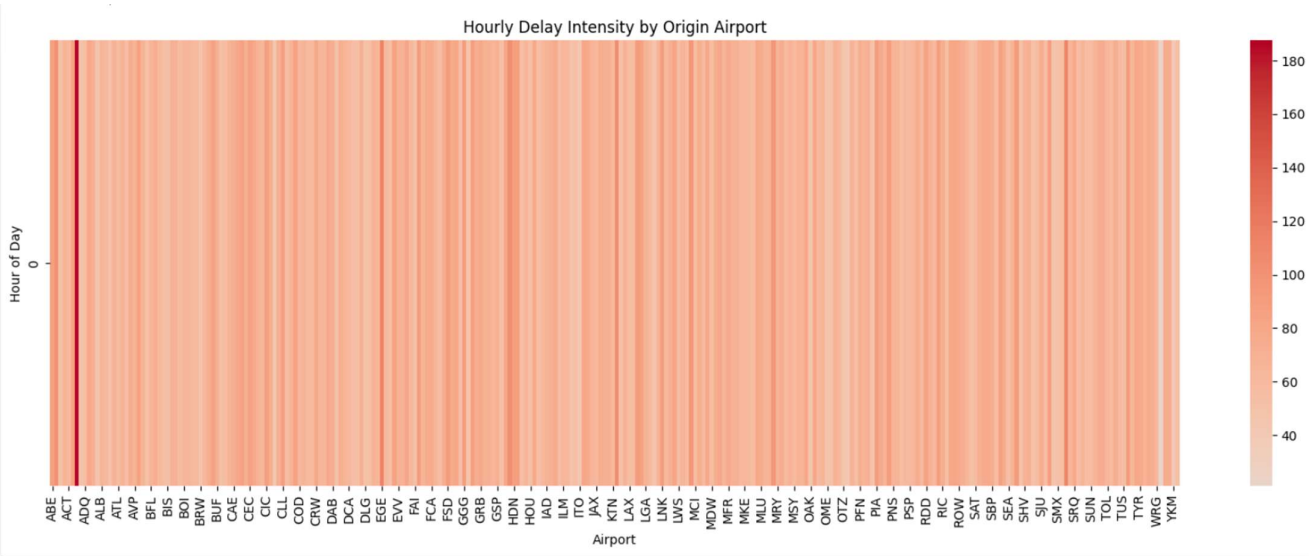
Connectivity Matrix (Top Routes)

Purpose: To visualize the strength of airport connections.
Method: Generated a crosstab between top 30 origin–destination pairs to approximate airport connectivity.
Visualization: Heatmap (Origin vs. Destination).
Insight: Highly connected airports formed dense clusters, reflecting their importance as transfer or hub locations in the flight network.



Hourly Delay Intensity by Airport

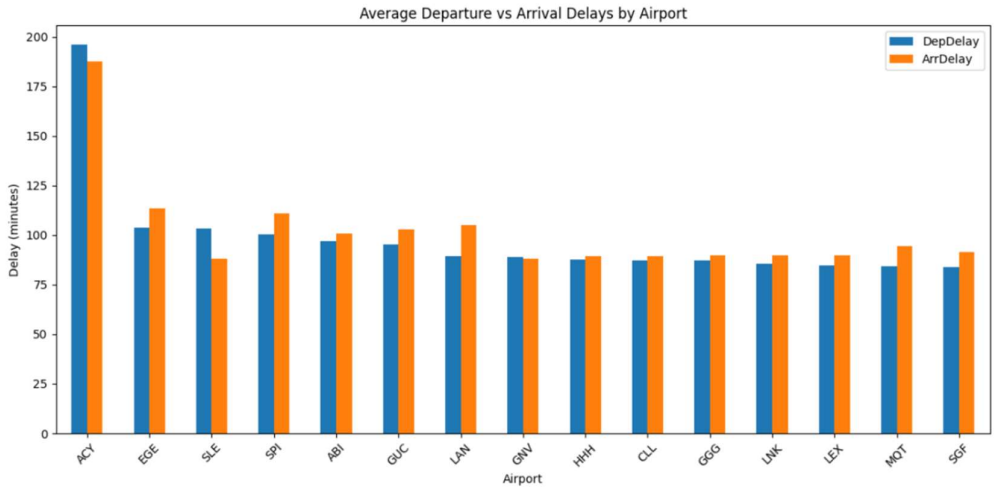
Purpose: To analyze how delay severity changes by time of day and airport.
Method: Created a pivot table of average ArrDelay grouped by hour and airport.
Visualization: Heatmap (Hour vs. Origin Airport).
Insight: Most airports showed peak delay intensity during late afternoons and evenings, aligning with high flight congestion periods.



Departure vs Arrival Delay Comparison by Airport

Purpose: To compare how departure delays impact arrival delays across airports.
Method: Calculated mean DepDelay and ArrDelay for each origin airport.
Visualization: Grouped Bar Chart (Airport vs. Avg Delay).

Insight: A strong positive relationship was observed — airports with longer departure delays tended to have proportionally higher arrival delays.



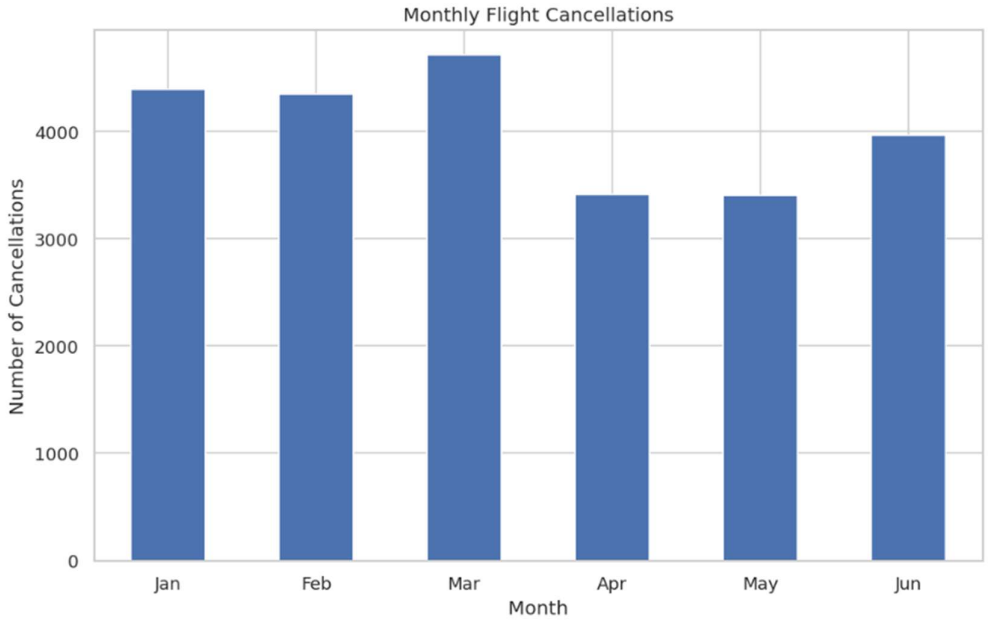
---

6. Cancellation and Seasonal Trends

Seasonal and Cancellation Analysis focuses on identifying when and why flight cancellations occur throughout the year. It explores, monthly and seasonal patterns in cancellations,impacts of holidays and winter weather, and causes of cancellations such as carrier, weather, NAS, and security delays.

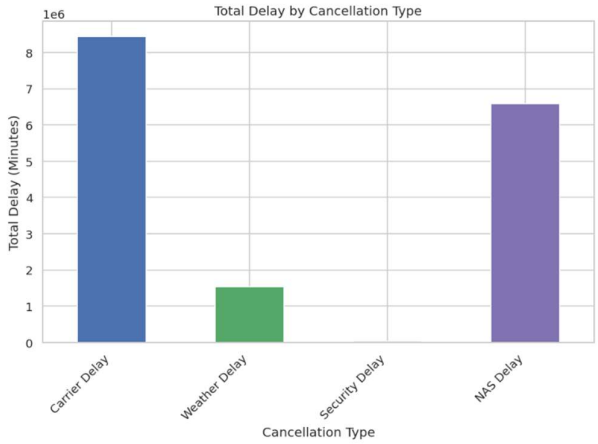
Monthly Flight Cancellations

Purpose: To examine how flight cancellations vary throughout the year.
Method: Grouped data by month and counted total cancelled flights.
Visualization: Bar Chart (Month vs. Number of Cancellations).
Insight: Winter months such as January and February recorded the highest cancellations, suggesting strong seasonal impacts due to weather conditions.



Total Delay by Cancellation Type

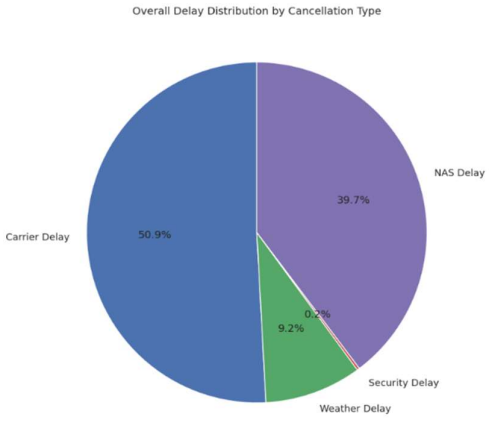
Purpose: To analyze the overall contribution of each delay cause to total delay time.
Method: Summed total minutes for Carrier, Weather, NAS, and Security delays.
Visualization: Vertical Bar Chart (Cancellation Type vs. Total Delay Minutes).
Insight: Carrier and NAS delays contributed the most to total delay time, while security-related delays were minimal.




```
Total Delay Duration by Cancellation Type (in Minutes)
Carrier Delay     8440607
Weather Delay     1527927
Security Delay      39749
NAS Delay         6589613
dtype: int64
```

Overall Delay Distribution by Cancellation Type

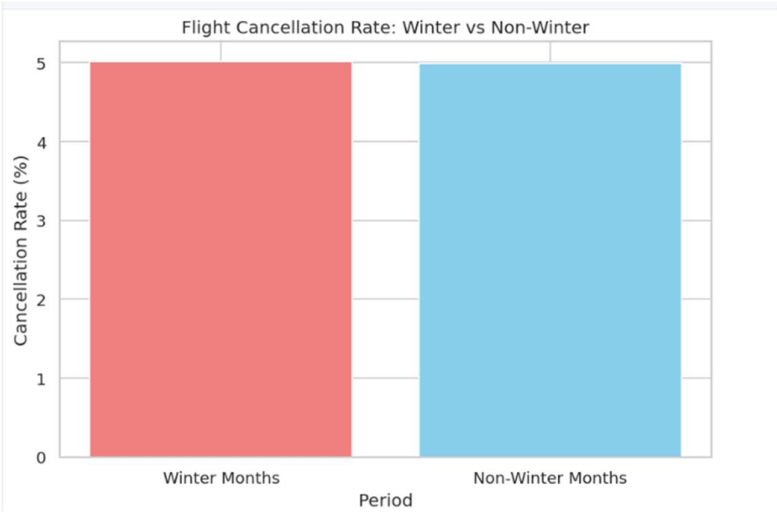
Purpose: To compare how different delay categories contribute proportionally to total delay time.
Method: Calculated the percentage share of each delay type from total delay minutes.
Visualization: Pie Chart (Proportion of Delay Causes).
Insight: Carrier delays dominated the delay distribution, followed by NAS and Weather delays, indicating airline-related issues as the main cause.



Winter vs Non-Winter Cancellation Rate

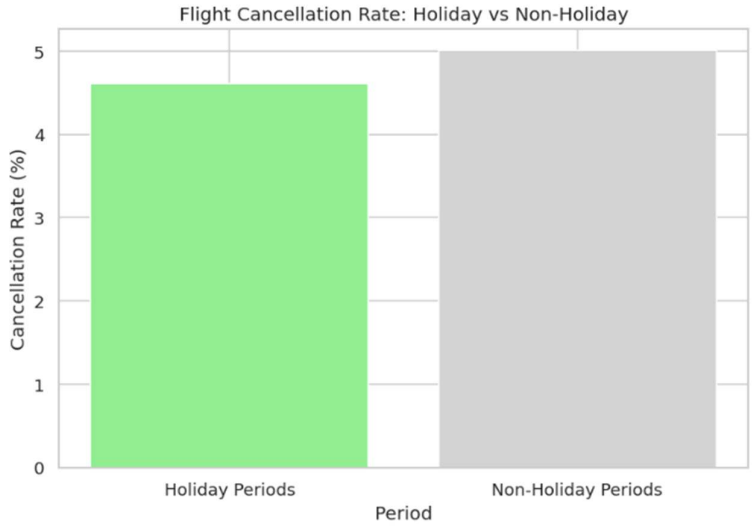
Purpose: To measure the impact of cold-season weather on flight cancellations.
Method: Compared the average cancellation rate between winter months and non-winter months.
Visualization: Bar Chart (Winter vs. Non-Winter Cancellation Rate).

Insight: Winter months had a noticeably higher cancellation rate, confirming the influence of harsh weather on flight operations.



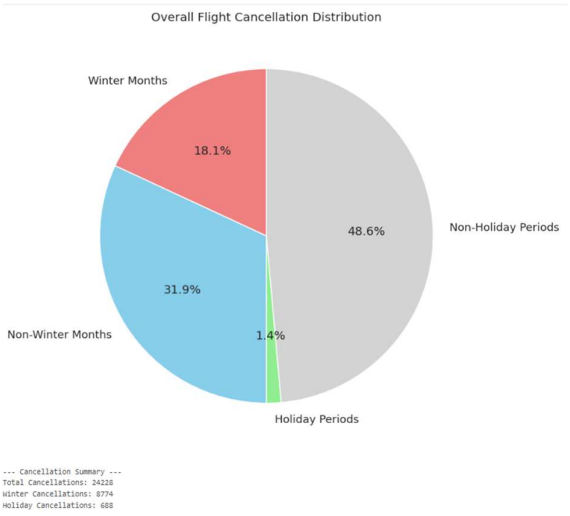
Holiday vs Non-Holiday Cancellation Rate

Purpose: To assess how holiday periods affect flight cancellations.
Method: Compared cancellation rates during major holidays (e.g., Christmas, New Year) with regular days.
Visualization: Bar Chart (Holiday vs. Non-Holiday Cancellation Rate).
Insight: Holiday periods showed a slight increase in cancellations, likely due to high demand and overbooked schedules leading to operational strain.



Overall Flight Cancellation Distribution

Purpose: To visualize the overall share of cancellations across different time periods.
Method: Summed cancellations for Winter, Non-Winter, Holiday, and Non-Holiday flights.
Visualization: Pie Chart (Proportion of Cancellations by Season and Period).
Insight: Winter and holiday months collectively represented a significant portion of total cancellations, emphasizing the effect of both seasonal weather and passenger demand surges.



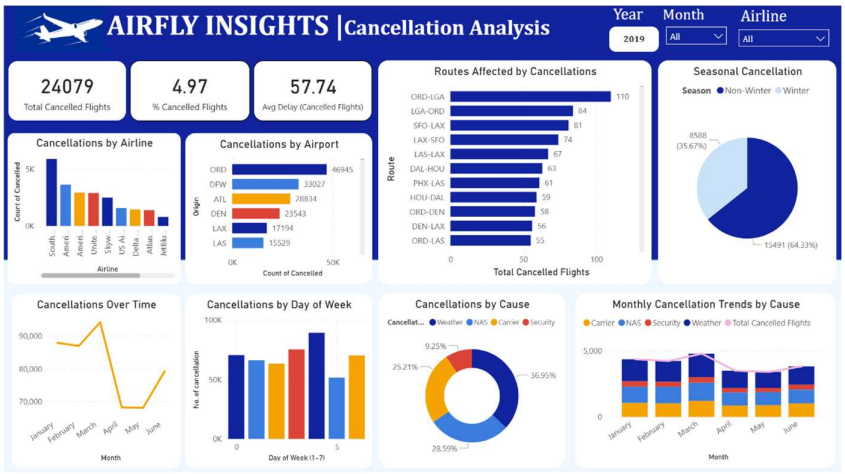
---

7. Final Dashboard

To combine all major analyses—delays, cancellations, routes, and seasonal trends—into an interactive visual dashboard for better understanding and decision-making. The final Power BI dashboard brings together cleaned and processed airline data into dynamic visuals and KPIs. It allows users to explore flight performance across airlines, airports, routes, and time periods with filters and drill-down options.

Steps to Create Dashboard in Power BI

1. **Import Data:** Load the cleaned dataset into Power BI.
2. **Clean & Transform:** Use Power Query to fix data types and create new fields (Month, Route, Delay Type).
3. **Model Data:** Build relationships and calculated columns if needed.
4. **Create Visuals:** Add charts like bar, line, map, and pie for key metrics.
5. **Add KPIs:** Show total flights, % cancelled, and average delay.
6. **Apply Filters:** Use slicers for Airline, Airport, Month, and Cause.
7. **Design Layout:** Arrange visuals neatly with clear titles and colors.
8. **Publish & Share:** Upload to Power BI Service for interactive viewing.



---

8. Presentation

The presentation summarized the entire project — covering objectives, dataset overview, methodology, key analyses, visual insights, and the final Power BI dashboard. It showcased major findings on delays, cancellations, and route performance through clear visuals and explained how the dashboard supports data-driven decision-making.

Conclusion

The analysis successfully identified major delay causes, high-risk routes, and seasonal impacts, helping improve decision-making for airline operations and performance optimization.
