Airfly Insights Report

Data Visualization Internship

Submitted By:

Unnati Saxena

At

Infosys Springboard

Table of Content

Introduction

The **Airfly Insights** project is a comprehensive data analytics and visualization study conducted as part of a Data Visualization Internship under *Infosys Springboard*. The goal of the project is to analyze large-scale U.S. domestic airline flight records, uncover operational patterns, understand delay behaviours, and generate meaningful insights that support decision-making for airlines, airports, and aviation stakeholders. Using a combination of statistical analysis, data cleaning, visualization, and dashboard development, the project transforms raw flight data into actionable intelligence.

The dataset contains over **484,000+ flight records** covering multiple years, including detailed information on flight schedules, delays, cancellations, routes, airlines, and airport characteristics. Because flight delays and disruptions directly impact passenger experience, airline reputation, and operational efficiency, understanding these patterns is essential for improving aviation reliability. The project follows an 8-week milestone-based workflow, beginning with data loading and preprocessing, followed by visualization, delay analysis, route exploration, seasonal trends, dashboard creation, and final documentation preparation.

Through visual storytelling, the project highlights key behaviours such as busiest routes, airline performance, cancellation drivers, seasonal variations, and delay causes like Carrier, Weather, and Late Aircraft delays. The findings contribute to identifying inefficiencies, optimizing scheduling, and supporting predictive modelling. Overall, Airfly Insights demonstrates how data visualization and analytics can provide clarity to complex aviation operations and help improve the quality and dependability of airline services.

Problem Statement

The objective of this project is to analyze 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.

Tools and Technologies

The Airfly Insights project makes use of a diverse set of modern tools and technologies that support the full analytics lifecycle—from raw data handling to visualization and presentation. These tools enable efficient data processing, interactive exploration, and professional reporting. Together, they form a comprehensive environment for performing large-scale flight data analysis, identifying operational inefficiencies, and deriving business insights.

1. Python
Python is the core programming language used for analysis due to its extensive library support, readability, and ability to handle large datasets. Throughout the project, Python enables data loading, manipulation, statistical analysis, and feature engineering.

2. Pandas
Pandas provides essential data manipulation and data cleaning capabilities. It is used to load large CSV files, handle missing values, remove duplicates, filter data, convert types, summarize statistics, and create engineered features such as Month, DayNumber, Hour, and Route. Pandas forms the backbone of the project's preprocessing workflow.

3. NumPy
NumPy enhances performance for numerical computations across large datasets. It supports vectorized operations, mathematical transformations, and efficient array handling, making it critical for optimizing performance during data cleaning and feature engineering.

4. Matplotlib Seaborn and Plotly
These visualization libraries are used extensively for exploratory data analysis during the early milestones.

- *Matplotlib* provides low-level customization for charts and graphs.
- *Seaborn* builds on Matplotlib and offers visually appealing, statistically accurate plots.
- *Plotly helps in building interactive visuals.*
  They help visualize delays, distributions, correlations, seasonality, and airline performance insights.

5. Power BI
Power BI is a central tool for building the final dashboards and visual reports. It is used to create interactive dashboards such as airline comparisons, delay breakdowns, cancellation analysis, route-level performance, and flight volume mapping. Its drag-and-drop interface and advanced visualization options make it ideal for converting analytical results into dynamic, easy-to-understand visuals for stakeholders.

6. Databricks
Databricks serves as the cloud-based data analytics environment for running Python scripts, managing datasets, and executing large-scale data transformations. It combines the flexibility of Python with scalable processing power, allowing smooth handling of large flight datasets. Databricks notebooks also support markdown, inline visualizations, and collaborative development, making it a strong environment for end-to-end data analysis and workflow management.

7. Excel / CSV File Handling
CSV files form the primary data format for raw and cleaned datasets. Tools like Microsoft Excel are occasionally used for quick inspections, column checks, and verifying data structures during the early phase of the project.

8. GitHub
GitHub serves as a platform to store notebooks, reports, the README, and the final project documentation, ensuring transparency and reproducibility.

9. PowerPoint
These tools are used in Week 8 to build the final slide deck for project presentation. They help structure findings, display visualizations from Power BI, and communicate the full journey of insights in a clear, professional manner.

10. Report Writing Tools (Word)
Used for compiling the final report, adding formatting, inserting captions, organizing sections such as the Table of Contents and Appendices, and exporting the polished document in PDF format.

Together, these tools and technologies create a powerful and efficient analytics ecosystem. They enable data transformation, in-depth visualization, performance benchmarking, and professional reporting—ultimately helping translate complex flight data into meaningful insights for improving airline operations and decision-making.

Milestone 1: Data Foundation and Cleaning

Week 1:Project Initialization and Dataset Setup

1. Dataset Overview

The dataset contains domestic US airline flight records, including detailed information about flights, delays, airports, and operational metrics.
Originally the dataset had 484,559 rows and 29 columns, and after initial loading and inspection, the cleaned version contains 484,549 rows and 36 columns.

Key Statistics:

- Total flights: 484,549
- Original features: 29
- New features after engineering: 35
- Time period: Multiple years of flight data
- Coverage includes various airlines, airports, and routes across the United States

2. Initial Loading and Exploration (Schema, Types, Size, Nulls)

As part of Week 1, the primary tasks included loading the CSV using pandas, exploring the schema, understanding datatypes, and inspecting dataset size and structure.

This involved:

- Checking all column types
- Understanding the presence of null or inconsistent values
- Inspecting the range and distribution of columns such as delays, times, airport identifiers, carrier information, and more
- Performing sampling to understand data patterns, operational metrics, and overall dataset memory usage
- Conducting memory optimizations where necessary to handle the dataset efficiently

3. Column Descriptions (Original Understanding Before Cleaning)

Time & Scheduling

- DayOfWeek: Day of week (1 = Monday, 7 = Sunday)
- Date: Scheduled flight date
- DepTime: Actual departure time (local, hhmm format)
- ArrTime: Actual arrival time (local, hhmm format)
- CRSArrTime: Scheduled arrival time (local, hhmm format)

Flight Identification

- UniqueCarrier: Unique carrier code
- Airline: Airline company name
- FlightNum: Flight number
- TailNum: Aircraft tail number (specific plane)

Duration & Timing Metrics

- ActualElapsedTime: Actual time from departure to arrival (includes taxi times)
- CRSElapsedTime: Scheduled elapsed time (minutes)
- AirTime: Actual time spent in air (minutes)
- TaxiIn: Time from wheels down to gate arrival (minutes)
- TaxiOut: Time from gate departure to wheels off (minutes)

Delay Information

- ArrDelay: Difference in minutes between scheduled and actual arrival time
- DepDelay: Difference in minutes between scheduled and actual departure time
- CarrierDelay: Delay due to carrier issues (maintenance, crew, cleaning, fueling)
- WeatherDelay: Delay due to weather conditions
- NASDelay: Delay due to National Aviation System
- SecurityDelay: Delay due to security issues
- LateAircraftDelay: Delay caused by late arriving aircraft

Location & Route

- Origin: Origin airport name
- Org_Airport: Origin airport name (duplicate source column)
- Dest: Destination airport name
- Dest_Airport: Destination airport name
- Distance: Distance between airports in miles

Flight Status

- Cancelled: 1 = cancelled, 0 = not cancelled
- CancellationCode: Reason for cancellation (A = carrier, B = weather, C = NAS, D = security, N = not cancelled)

- Diverted: 1 = diverted, 0 = not diverted

This completes all Week 1 deliverables: loading, inspecting, exploring, understanding schema, and identifying issues for preprocessing.

Week 2: Preprocessing and Feature Engineering

1. Data Cleaning Steps (Using Pandas)

Handling Missing Values

- Org_Airport: 1,177 null values fixed
- Dest_Airport: 1,479 null values fixed
- Cancelled: 13 null values fixed
- All missing values were resolved through imputation, producing a clean dataset with no remaining nulls.

Removing Duplicates

- Identified and removed 10 duplicate rows
- Final dataset contains 0 duplicates

Data Type Conversions

- Converted Date column to proper datetime format
- Verified and standardized numeric columns
- Ensured all delay-related columns maintained consistent formats and ranges

2. Feature Engineering

The engineered features created during Week 2 include:

- Month: Extracted from the flight date (1–12)
- DayNumber: Numeric day of week (1 = Monday, 7 = Sunday)
- Hour: Extracted from departure time
- Route: Combined Origin + Dest airport names
- OnTime: Binary column (1 = arrived on time, 0 = delayed)
- DepHour: Extracted from DepTime

After preprocessing and feature creation, the dataset expanded to 36 columns and remained at 484,549 rows.

3. Insights Enabled After Week 2 Processing

(These insights were extracted *after* the preprocessing and feature engineering completed.)

Distance Analysis

- Minimum Distance: 31 miles
- Maximum Distance: 4,502 miles
- Average Distance: 752.14 miles
- Extracted flights >1,000 miles for long-haul analysis

Flight Volume by Day

- Monday: 70,254 flights
- Tuesday: 65,934
- Wednesday: 63,055
- Thursday: 75,011
- Friday: 88,972 (Peak)
- Saturday: 51,330 (Lowest)
- Sunday: 69,995 flights

Operational Performance

- Average Taxi In Time: 6.78 minutes
- Average Taxi Out Time: 19.15 minutes
- Average Departure Delay: 12.34 minutes
- Average Arrival Delay: 14.22 minutes
- Identified top 10 longest flights for detailed analysis
- On-time performance by airline ranged from 0.67 to 0.92

Airport-Level Metrics

- Busiest Origin Airport: Hartsfield–Jackson Atlanta International Airport, Atlanta, GA — 40,213 flights
- Highest Avg Departure Delay: LaGuardia — 25.47 minutes
- Highest Avg Arrival Delay: JFK — 27.36 minutes

Route-Level Insights

- Worst Routes (highest average arrival delay): Example: Boston Logan → Los Angeles International — 34.56 minutes

- Best Routes (lowest arrival delay): Example: San Francisco International → Oakland International — 2.13 minutes

4. Business Insights (After Data Preparation)

- Friday is the busiest travel day (88,972 flights)
- Saturday has the fewest flights (51,330 flights)
- Taxi-out times are nearly 3× longer than taxi-in times
- Long-haul flights (>1,000 miles) contribute significantly to delays
- Certain airports (LaGuardia, JFK) consistently show higher delays
- Clean dataset now supports accurate route-based, seasonal, and airline-level analysis
- Ready for building delay prediction models and operational decision-making tools

Milestone 2: Visual Exploration and Delay Trends

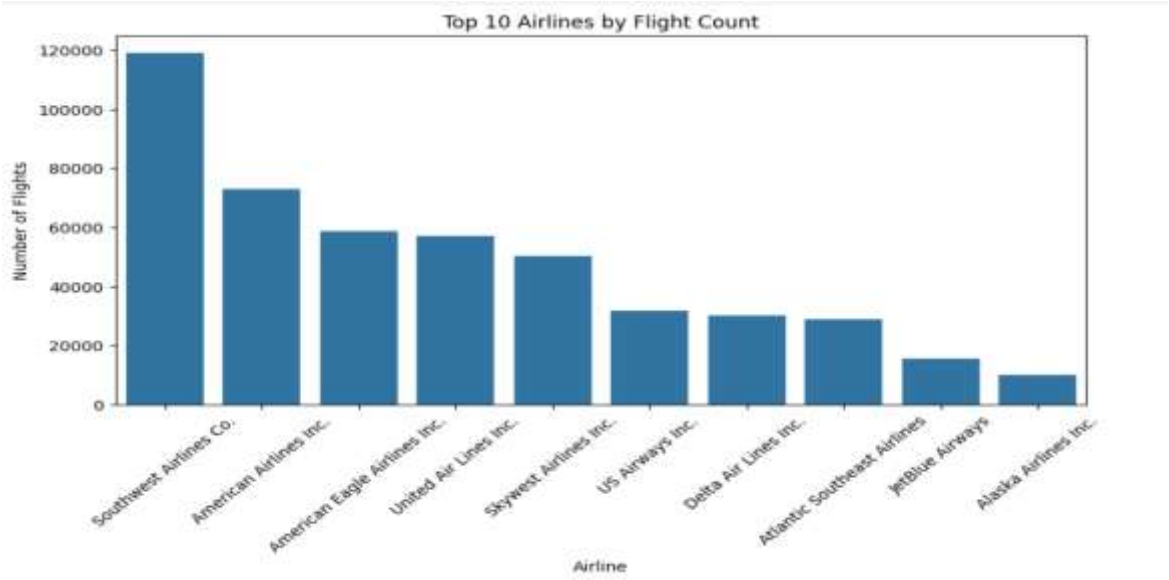
Week 3: Univariate and Bivariate Visual Analysis

- Top airlines, routes, and busiest months
- Flight distribution by day, time, and airport
- Plot bar charts, histograms, boxplots, and line plots

1. Top Airlines, Routes, and Busiest Months

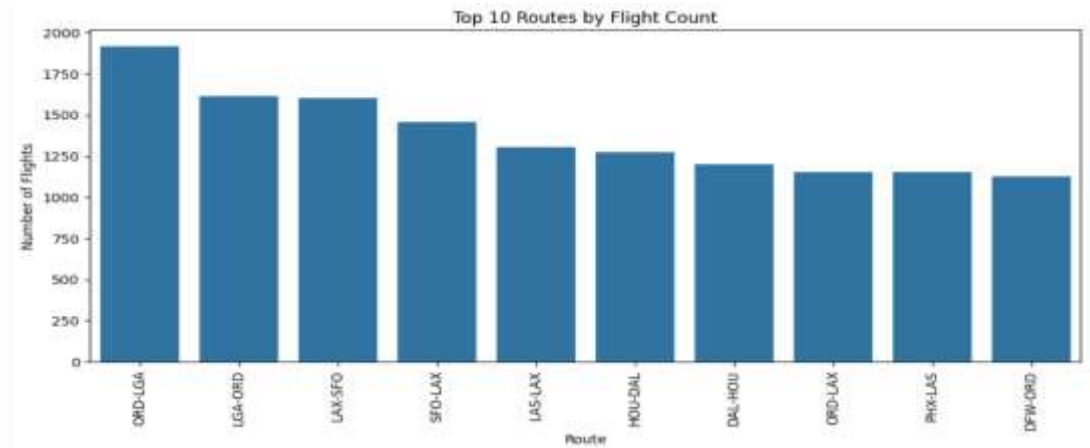
Key Insights:

- Top Airline:



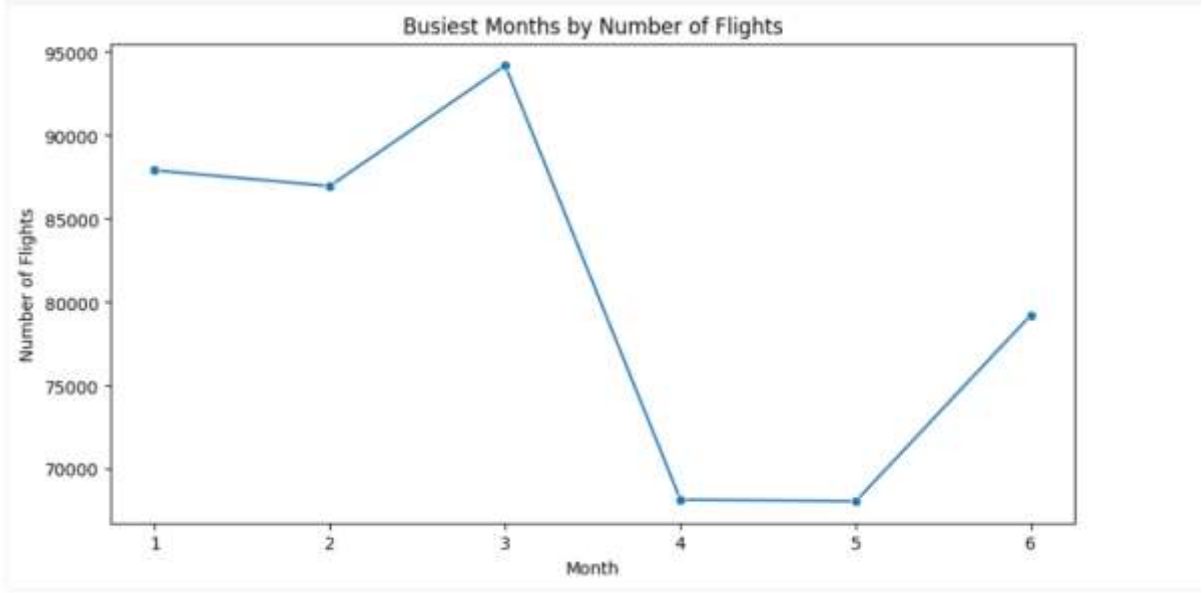
Southwest Airlines Co. has the most flights (~120,000), much more than any other airline.

- Top Route:



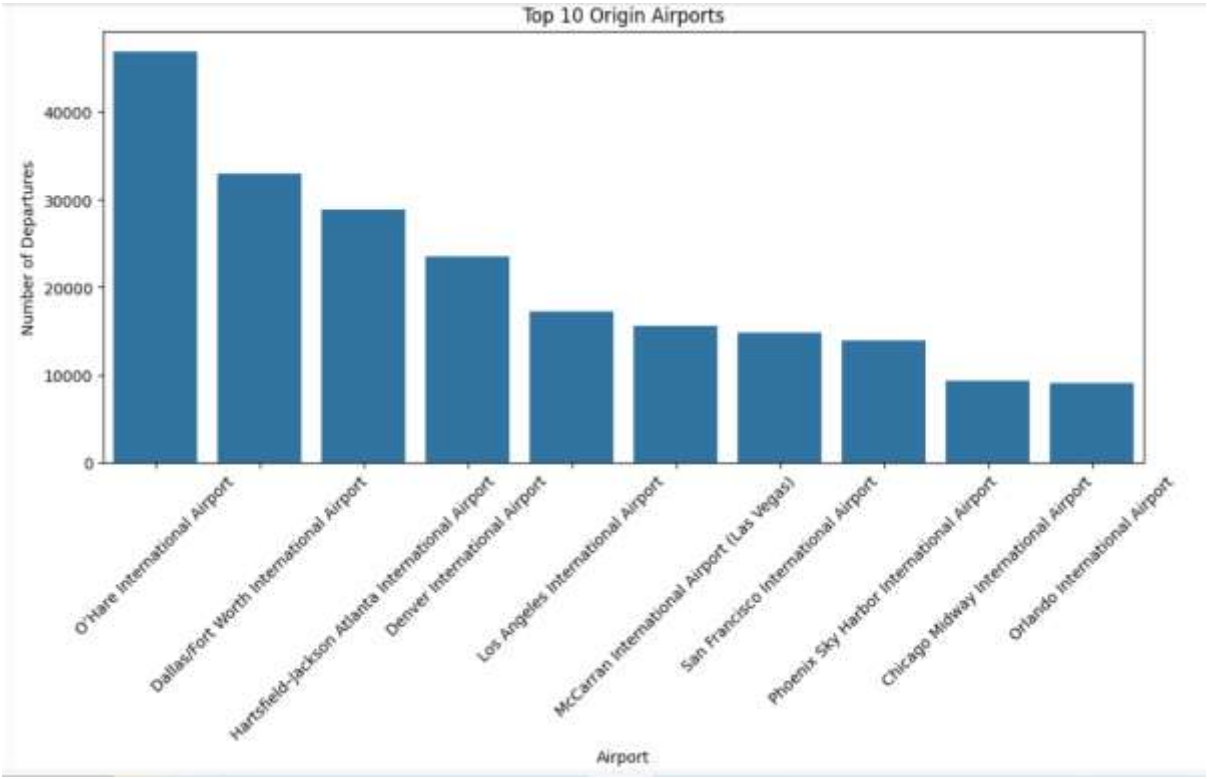
The busiest flight path is O'Hare (ORD) → LaGuardia (LGA).

- Busiest Month:



March (Month 3) has the most flights (around 94,500). Flight numbers drop sharply in April and May.

- Top Origin Airport:



O'Hare International Airport (ORD) has the highest number of departing flights.

- Pattern:
  Air traffic is highest in early spring and focused around major city routes.

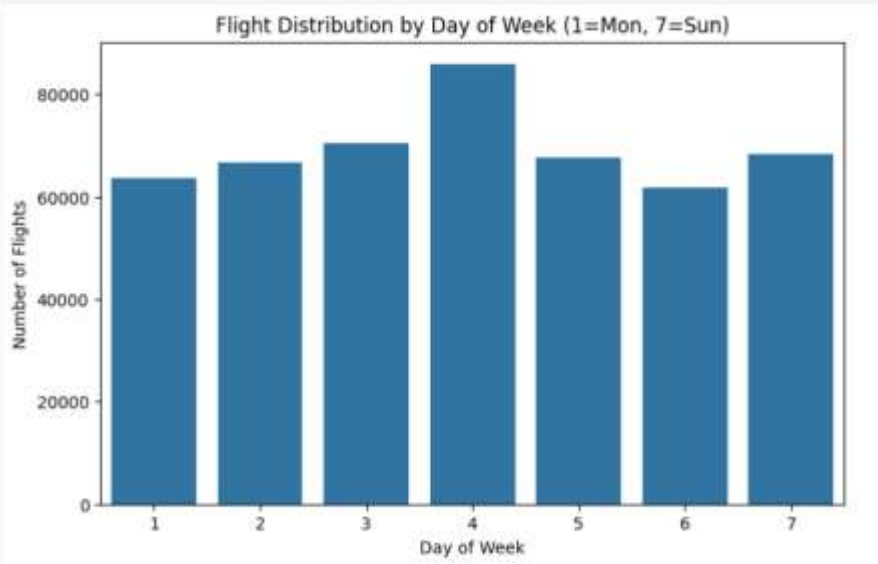
Possible Measures:

- Add more flights or larger aircraft on popular routes (like ORD–LGA).
- Plan for extra staff and airport operations during March.
- Use quieter months (April–May) for maintenance or training.

2. Flight Distribution by Day, Time, and Airport

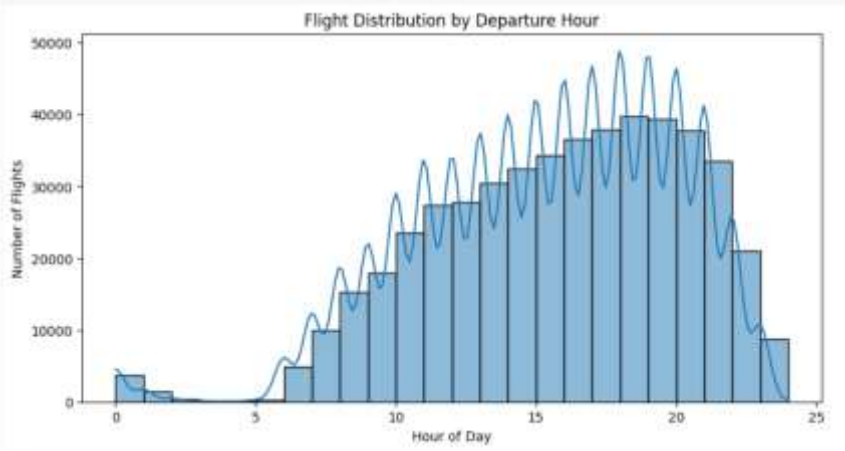
Key Insights:

- Busiest Day:



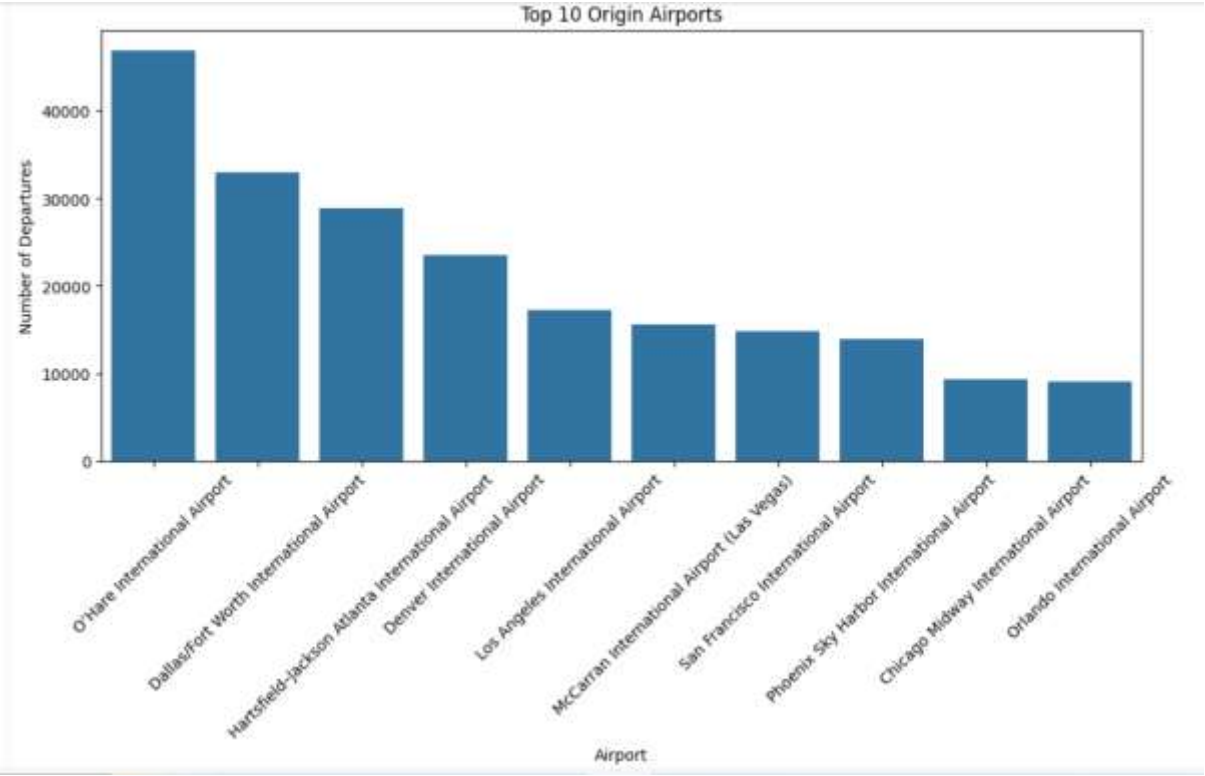
Thursday (Day 4) has the highest number of flights.

- Busiest Time:



Flights mostly depart between late afternoon and early evening, peaking around 7 PM (19th hour).

- Busiest Origin Airport:



O’Hare (ORD) again leads with the most departures.

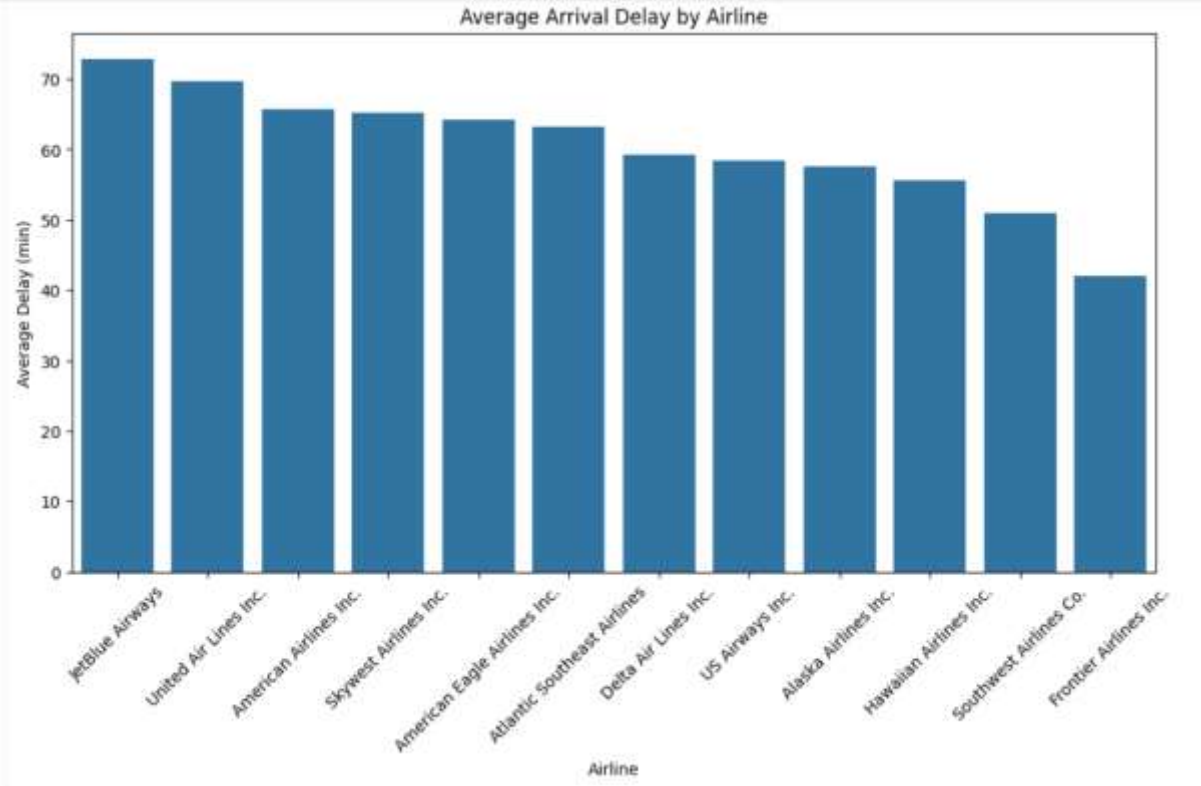
- Pattern:
  Weekdays, especially mid-to-late week, are busier than weekends.

Possible Measures:

- Increase airport and ground crew capacity during Thursday evenings.
- Spread flight schedules to reduce evening congestion.
- Provide more passenger support and services during peak hours.

3. Delays and Performance

Key Insights:



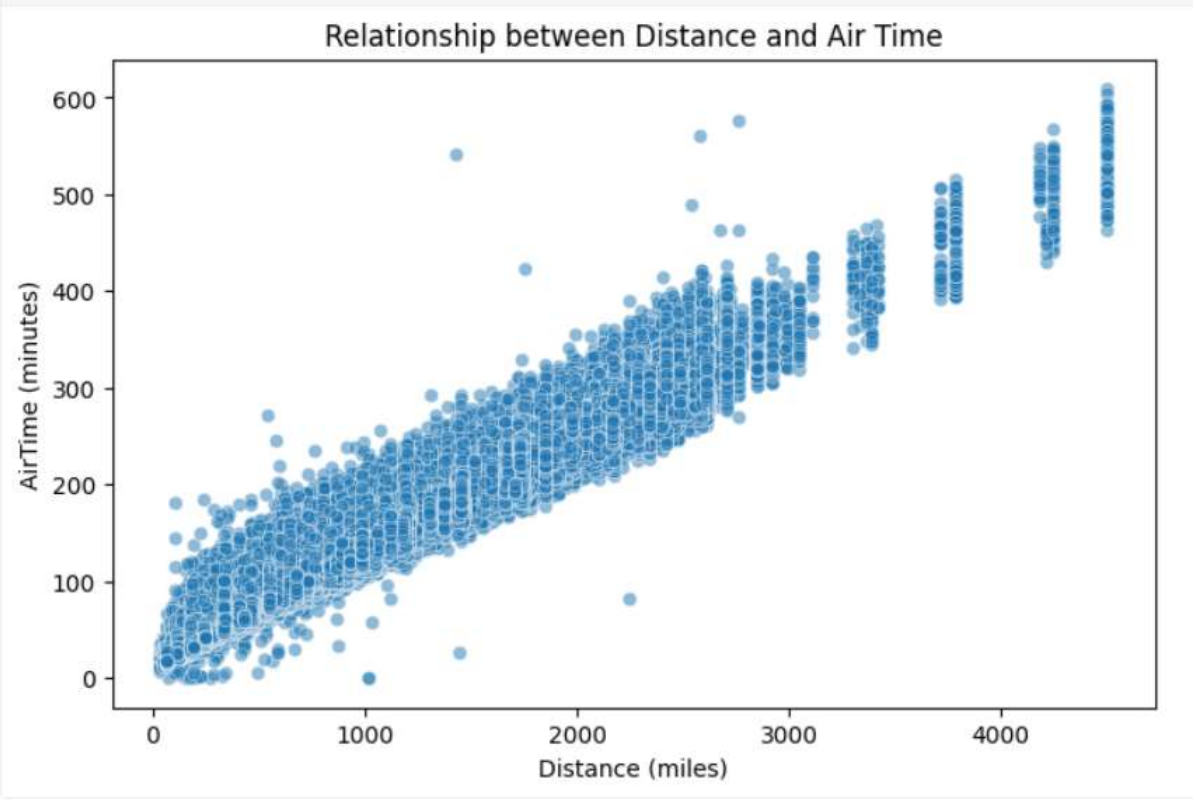
- Most Delayed Airline:
  JetBlue Airways – highest average arrival delay (over 70 minutes).
- Best Performer:
  Frontier Airlines Inc. – lowest average arrival delay (~42 minutes).
- Efficient Large Airline:
  Southwest Airlines manages the most flights but still has low delays.
- Delay Patterns:
  Average delays stay between 0–100 minutes, but some flights have extreme delays (up to 3,000 minutes).

Possible Measures:

- Airlines like JetBlue should review crew, maintenance, and turnaround processes.
- Reduce flight congestion by spreading departure times.
- Study efficient airlines (like Southwest) to apply their best practices.

4. Distance and Air Time

Key Insights:



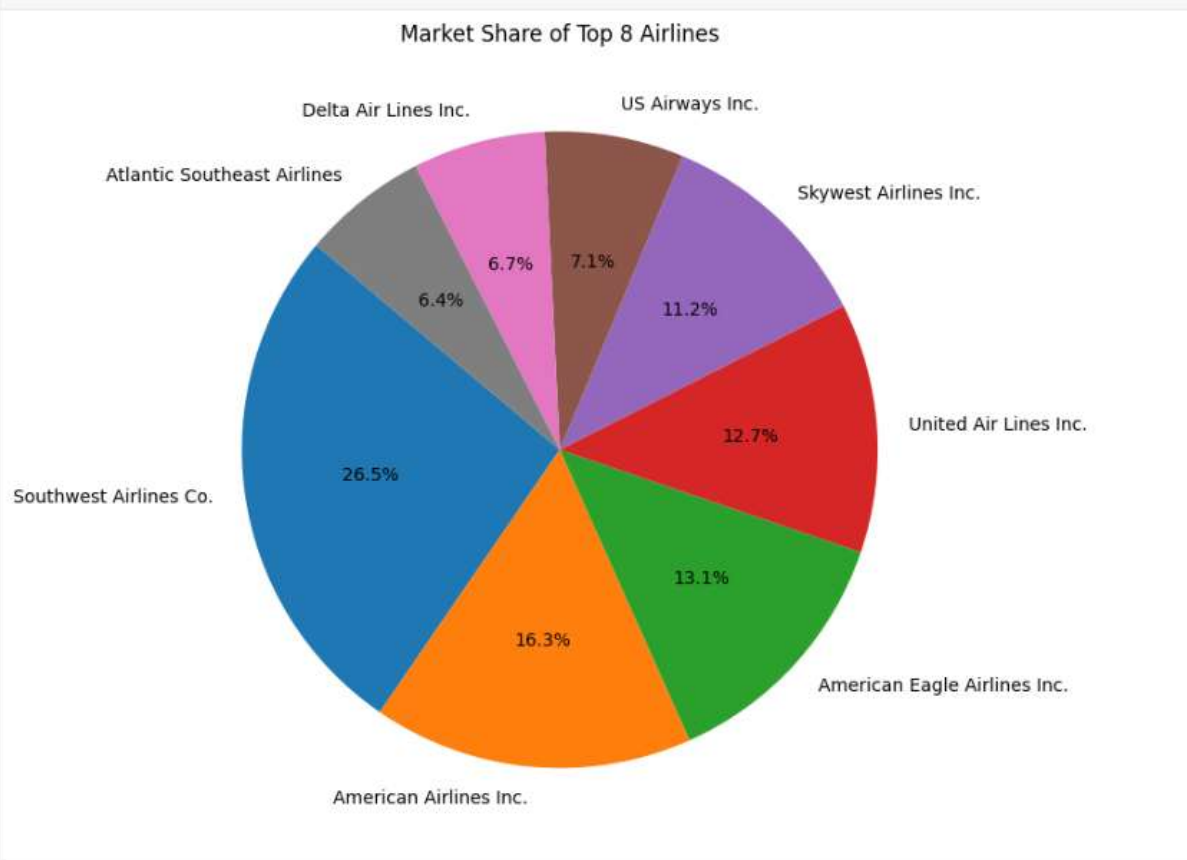
- Strong Relationship:
  The scatter plot shows a clear positive correlation — longer distances mean longer air times.
- Data Quality:
  The close pattern of points along a line shows reliable data with no major errors.

Possible Measures:

- Use this relationship to predict flight times more accurately.
- Check future data against this trend to catch possible data entry issues.

5. Market Share and Airport Activity

Key Insights:



- Largest Market Share:
  Southwest Airlines holds about 26.5% of total flights.
- Second Place:
  American Airlines holds around 16.3%.
- Busiest Arrival Airport:
  O'Hare (ORD) again is the busiest destination airport.
- Distance Distribution:
  Most flights are short-haul (200–300 miles); a smaller group are medium-haul (1,500–2,000 miles).
- Flight Duration:
  Short flights (50–100 minutes) are most common.

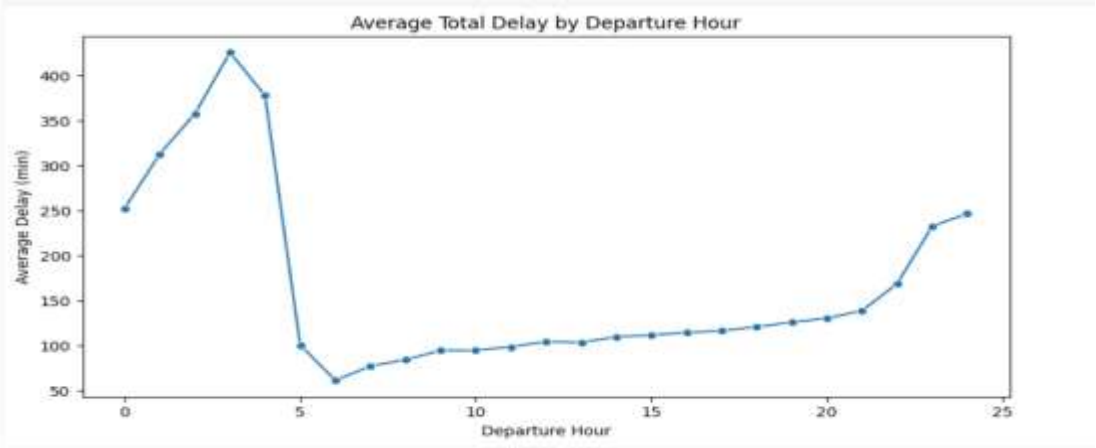
Possible Measures:

- Expand medium-haul routes to balance traffic.
- Improve infrastructure and scheduling at O'Hare.
- Smaller airlines could focus on niche or underserved routes.

6. Delays and Flight Timing

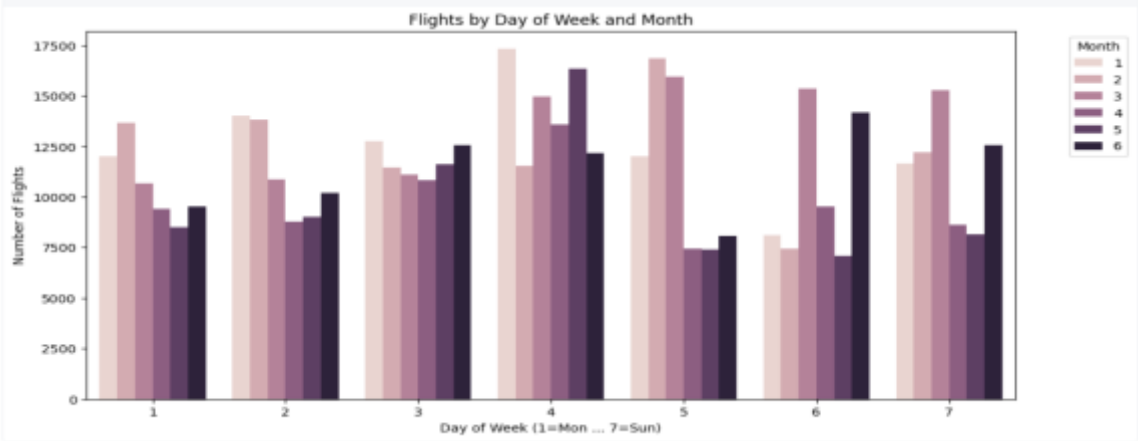
Key Insights:

- Peak Delay Times:



Flights at 4 AM have the highest delays (~400 minutes). Delays drop around 6 AM (60 minutes) and then slowly rise during the day.

- Volume Patterns:



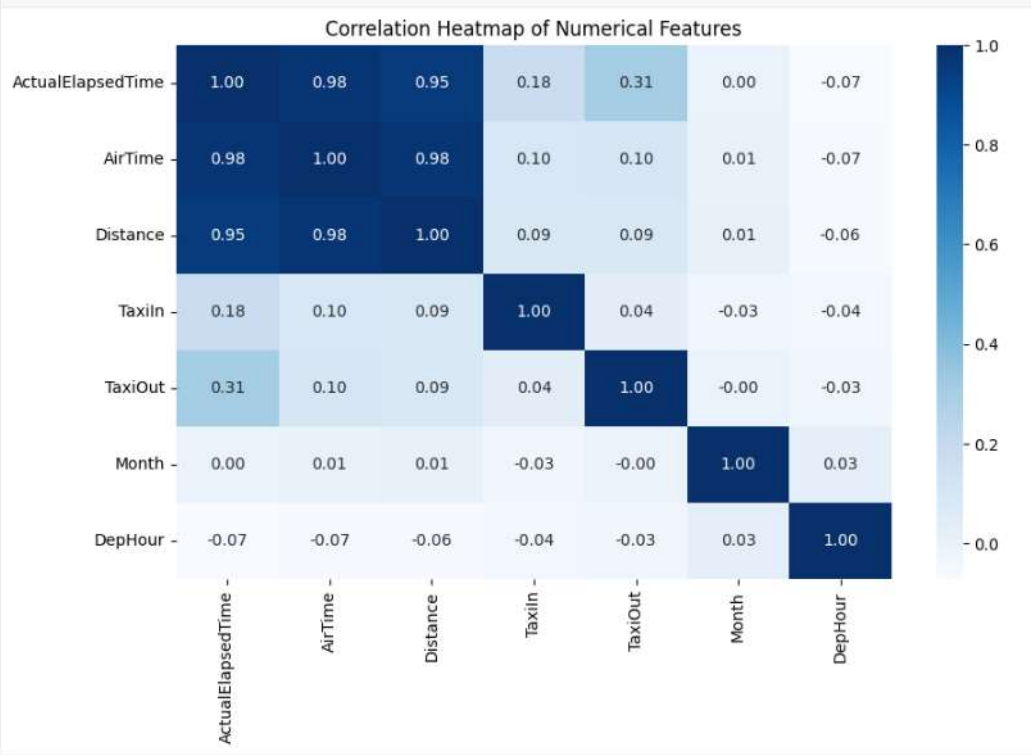
  - March (Month 3) has the highest flight activity every week.
  - Fridays (Day 5) are generally busy across all months.
  - Thursdays (Day 4) show strong activity in April and May.

Possible Measures:

- Avoid scheduling many flights between 2–5 AM, when delays are longest.
- Spread flights evenly throughout the day.
- Increase air traffic control and ground support during busy periods (Thursdays, Fridays).

7. Flight Duration and Correlation

Key Insights:



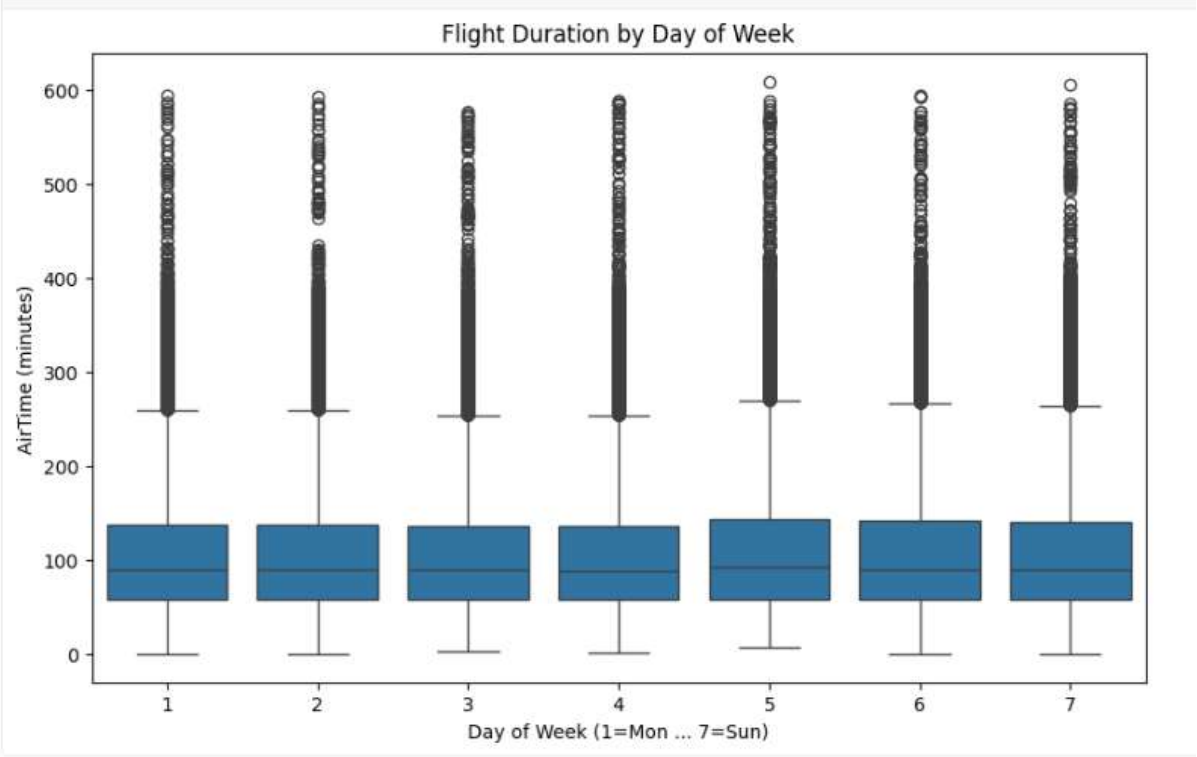
- Longest Average Duration:
  JetBlue Airways (~175 minutes).
- Shortest Average Duration:
  SkyWest Airlines Inc. (~100 minutes).
- Longest Routes:
  HNL–ATL (Honolulu–Atlanta) and ATL–HNL, both over 4,250 miles.
- Strong Correlation:
  Flight time depends mainly on distance (high correlation).
- Moderate Correlation:
  Small link between TaxiOut time and total duration (0.31).
- Weak Correlation:
  Month and departure hour don't affect flight time much.

Possible Measures:

- Use correlation data to forecast travel time and plan connections.
- Review taxiing processes to reduce ground time.
- Analyze long-haul flights (like Hawaii routes) for fuel and crew efficiency.

8. Flight Duration by Day of Week

Key Insights:



- Consistency:
  Flight durations are about the same every day (~100 minutes).
- No Major Changes:
  Median times and outliers are stable across all days, meaning day of week doesn't affect flight length.

Possible Measures:

- Keep uniform staffing and scheduling across all days — demand is steady.
- Use this stability for predictable crew rotations and maintenance planning.

9. Overall Summary

Key Takeaways:

- Thursdays, evenings, and March are the busiest times for flights.
- Southwest Airlines leads in both flight count and efficiency.
- O'Hare (ORD) is the central hub for both departures and arrivals.
- Most flights are short-haul and consistent across days.
- Delays are time-dependent (especially early morning) but not day-specific.
- Flight data shows strong internal consistency between distance and air time.

Overall Recommendations:

- Adjust schedules to avoid early-morning congestion.
- Focus resources on busy routes, airports, and months.
- Adopt efficiency strategies from low-delay airlines.
- Continue monitoring patterns seasonally for better planning.

Week 4: Delay Analysis – Airline and Weather

- Compare delay causes by airline
- Explore carrier delays, weather delays, NAS delays

- Visualize delays by time of day and airport

Introduction

This report analyzes different types of flight delays across airlines, airports, and times of day. The purpose is to identify the main reasons behind these delays and suggest practical measures to reduce them.

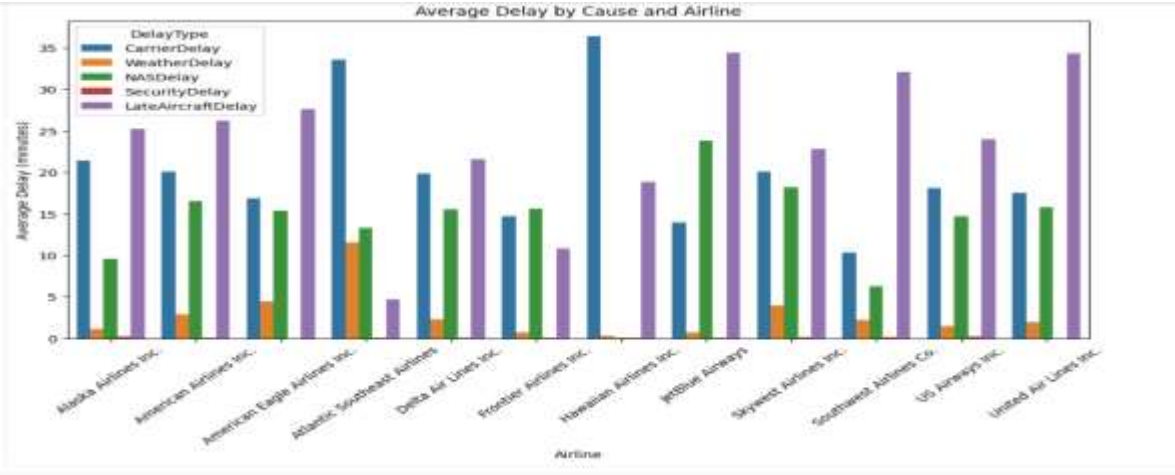
Types of Flight Delays

Before discussing the results, here are the main delay types included in the analysis:

- **Carrier Delay:** Caused by the airline itself, such as maintenance problems, crew delays, baggage handling, or fueling issues.
- **Weather Delay:** Caused by bad weather conditions such as fog, storms, or snow that make it unsafe to fly.
- **NAS (National Airspace System) Delay:** Caused by air traffic control congestion, airport capacity limits, or route restrictions.
- **Security Delay:** Caused by security checks or incidents that delay boarding or departure.
- **Late Aircraft Delay:** Happens when an aircraft arrives late from a previous flight and causes the next flight to also be late.

Insights and Recommended Measures

1. Major Causes of Delays



Insight:

Average Delay by Cause and Airline

- **"Late Aircraft Delay"** is often the **biggest cause** of average delays for many airlines, shown by the tall purple bars.
- **"Carrier Delay"** (blue bars) and **"Late Aircraft Delay"** (purple bars) are the most significant average delays across most airlines.
- **American Eagle Airlines Inc.** and **Frontier Airlines Inc.** stand out with particularly high average "Late Aircraft Delays".
- **JetBlue Airways** has a high average **"Carrier Delay"**.
- **Weather, NAS** (National Airspace System), and **Security** delays are generally much smaller on average than Carrier and Late Aircraft delays for most airlines.

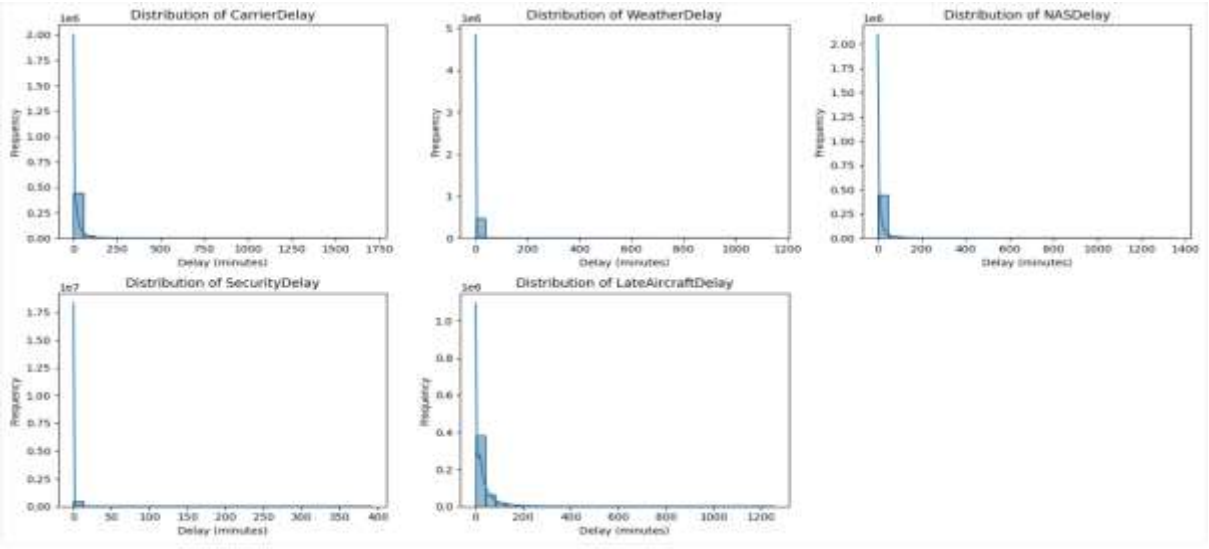
Distribution of Major Delay Types (Box Plot, Bottom):

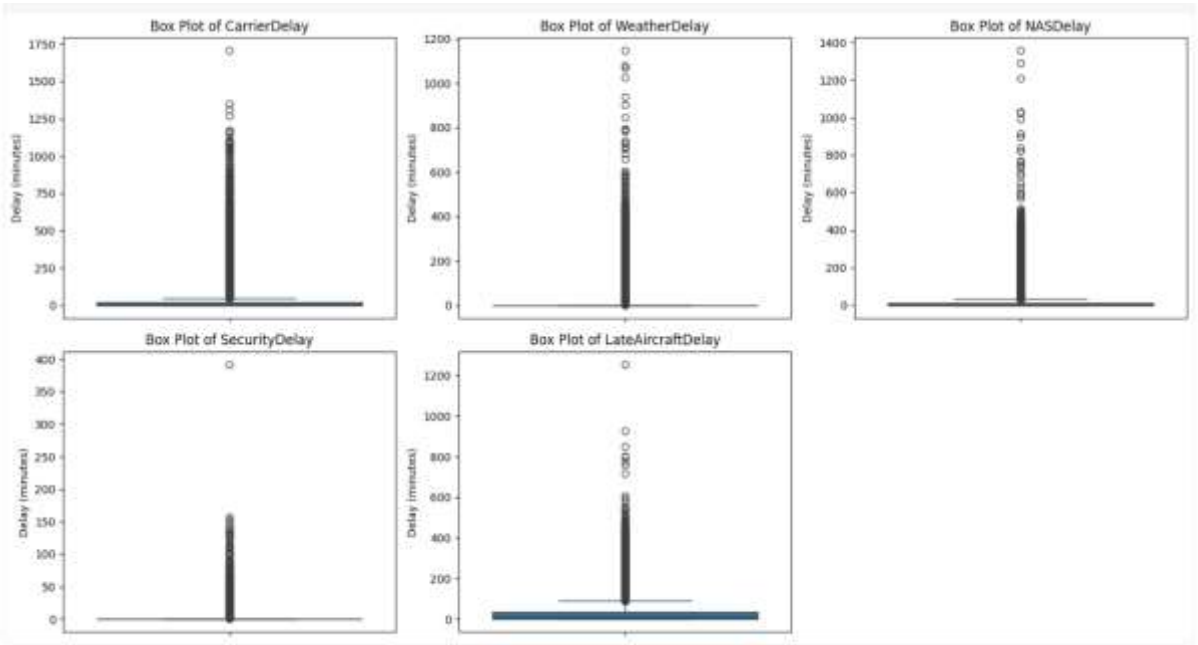
- All three major delay types—**CarrierDelay, WeatherDelay, and NASDelay**—have a **lot of very long, unusual delays** (outliers), shown by the scattered dots above the main boxes.
- The **median** (middle) delay time for all three is very low, near zero, meaning *most* delays are short or non-existent.

Measures:

- **Add buffer time** between flights, especially in early morning schedules, to prevent small delays from affecting later flights.
- **Improve ground operations** such as baggage handling and refueling to reduce Carrier Delays.
- **Ensure enough maintenance staff and reserve crew** to handle unexpected issues quickly.

2. Delay Patterns





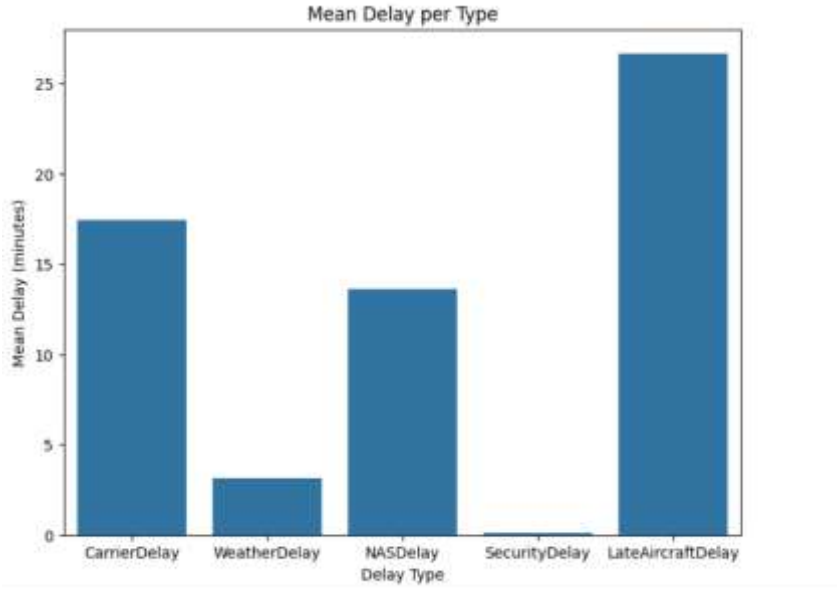
Insight:

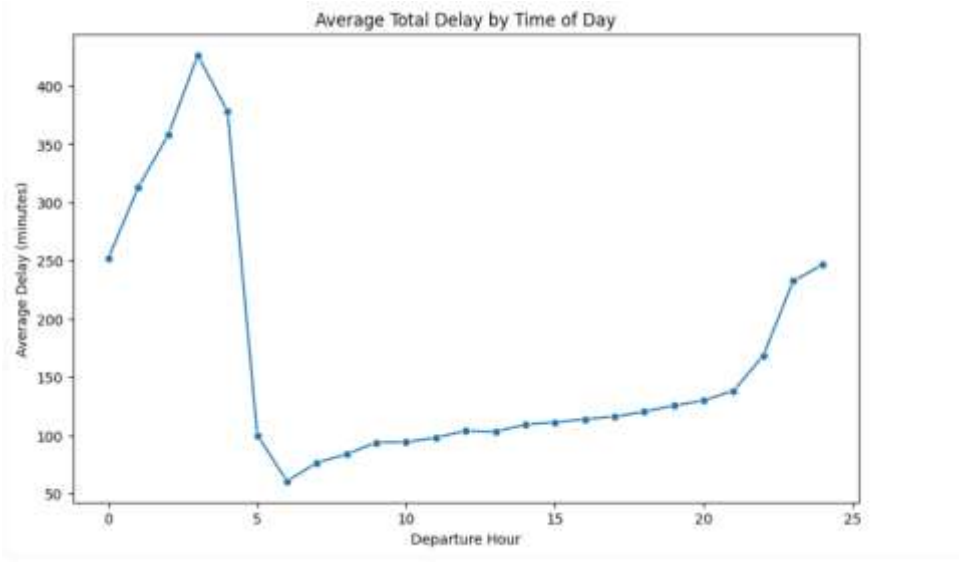
- **Delay Distributions:**
  - For all delay types, the vast majority of delays are very short, close to 0 minutes, indicated by the huge bars near the y-axis.
  - This confirms that extreme, long delays are rare, but they do happen.
- **Delay Box Plots:**
  - The box plots for all five delay types (Carrier, Weather, NAS, Security, Late Aircraft) confirm that most delays are concentrated near the bottom (0 minutes), but there are many extreme outliers (long delays) stretching up the graph.

Measures:

- **Monitor outliers closely:** Identify and study flights with unusually long delays to find specific causes.
- **Set up early-warning systems** to flag flights that exceed normal turnaround times so staff can act faster.

3. Time of Day Impact





Insight:

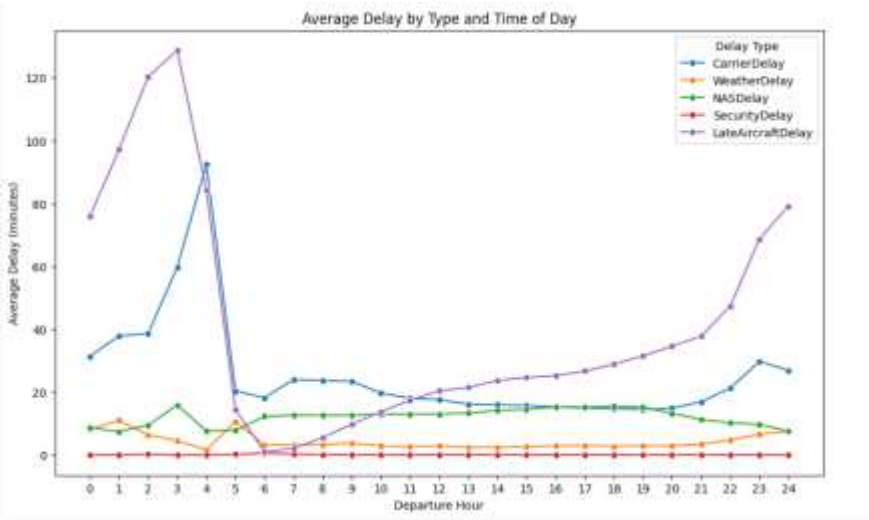
- **Mean Delay per Type (Bar Chart, Top):**
    - "Late Aircraft Delay" has the highest average (mean) delay time, around 26 minutes.
    - "Carrier Delay" is the second highest, around 17.5 minutes.
    - "Security Delay" has the lowest average delay, near 0.

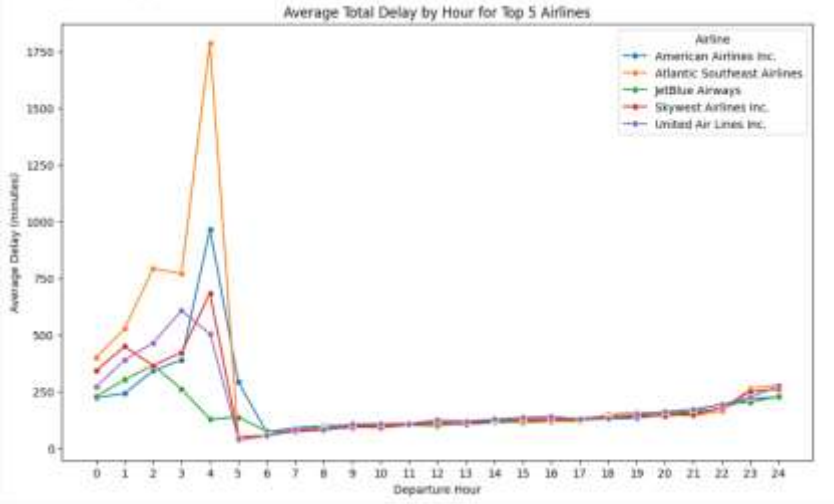
- **Average Total Delay by Time of Day (Line Chart, Bottom):**
    - The total average delay is highest in the early morning hours (Departure Hours 3 to 4), peaking around 400 minutes (nearly 7 hours).
    - Delays drop sharply around hour 5, and then stay low and stable through the main hours of the day (roughly 7 to 18).
    - Delays start to increase again late at night (from hour 21 to 24).

Measures:

- **Reschedule early morning flights** at major airports like O'Hare and LAX to slightly later times to reduce congestion.
- **Prepare a late-night recovery plan** by assigning standby crews and maintenance teams to handle aircraft arriving late at night.

4. Airline and Airport Performance





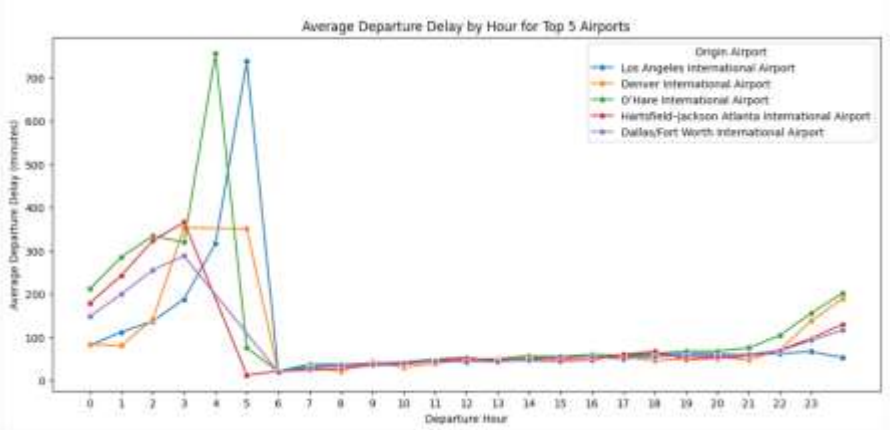
Insight:

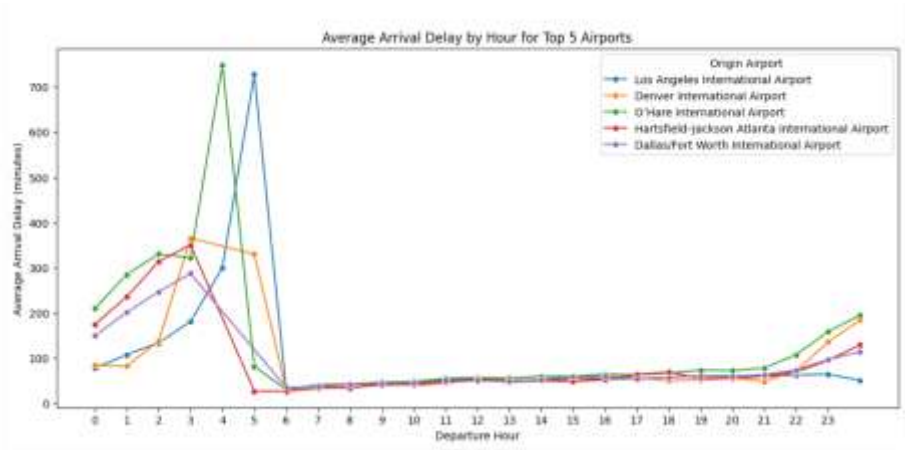
- **Average Delay by Type and Time of Day:**
    - The **early morning peak** (Hours 3 to 5) in total delay is almost entirely driven by **"Late Aircraft Delay"** and **"Carrier Delay"**.
    - During the stable daytime hours (around 7 to 20), **"NAS Delay"** has the highest average, though it's still low.
    - The **"Late Aircraft Delay"** starts to build up again late in the day, causing the late-night increase in total delay.
- **Average Total Delay by Hour for Top 5 Airlines:**
    - **Atlantic Southeast Airlines** and **American Airlines Inc.** experience extremely high average total delays during the **early morning** (Hours 3 to 5), with Atlantic Southeast reaching over 1750 minutes (nearly 30 hours!).
    - Delays are generally **low and similar** for all top airlines during the core daytime hours.

Measures:

- **Target problem airlines and airports** with specific improvement programs.
- **Increase early morning staffing levels** at O'Hare and LAX to handle high-delay periods.
- **Review airline schedules** to avoid having too many flights departing at the same time during peak hours.

5. Top Delay-Prone Airports





Insight:

Departure and Arrival Delays by Hour for Top 5 Airports:

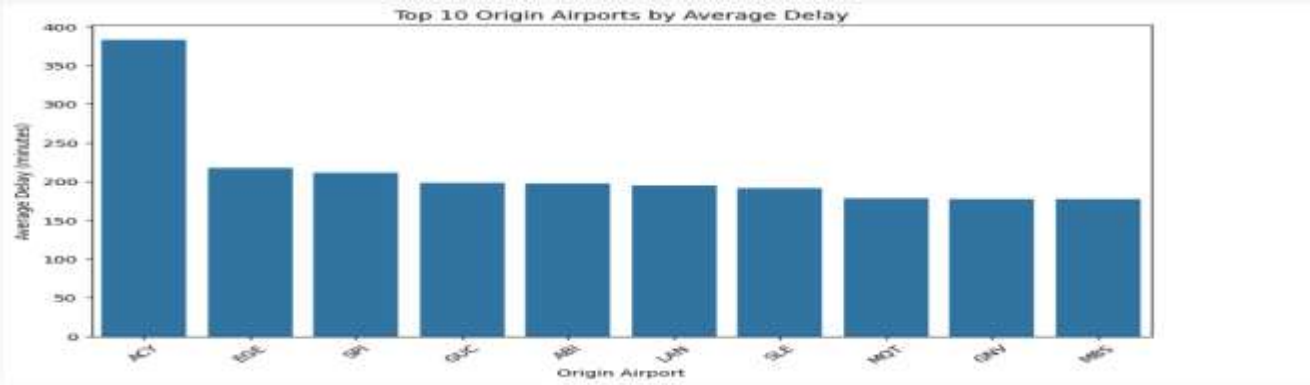
- Similar to the airline data, all top airports (Los Angeles, Denver, O'Hare, Hartsfield-Jackson Atlanta, Dallas/Fort Worth) see their **highest average departure and arrival delays** in the **early morning** (Hours 3 to 5).
- **O'Hare International Airport** and **Los Angeles International Airport** show the most **extreme peaks** in both departure and arrival delays during this early morning period.
- Delays are **low and stable** during the day and start to increase again late at night.

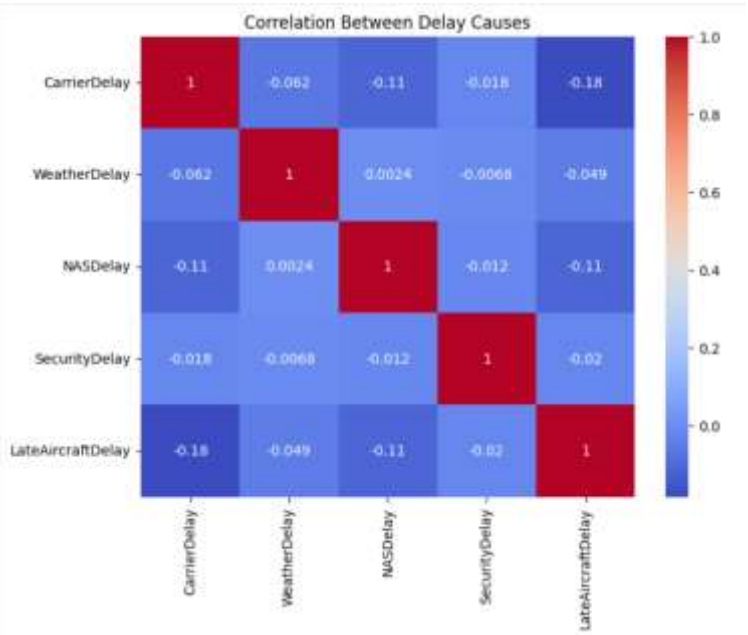
Measures:

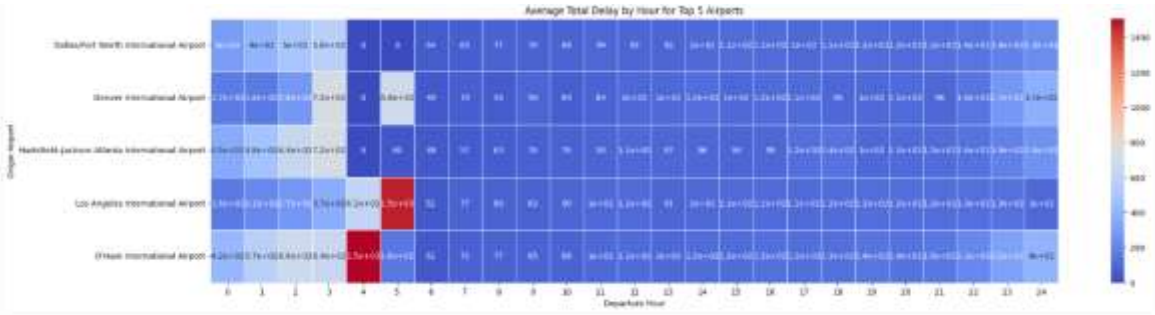
- **Investigate operational issues** at ACY to understand the cause of long delays.

- **Enhance coordination** between airlines and airport authorities to speed up ground handling and departures.
- **Introduce stricter performance monitoring** at airports with high average delays.

6. Correlation Between Delay Causes







Insight:

- **Top 10 Origin Airports by Average Delay (Bar Chart, Top):**
    - **ACY** (Atlantic City International Airport) has the **highest average delay** among the top 10 origin airports, at over 350 minutes.

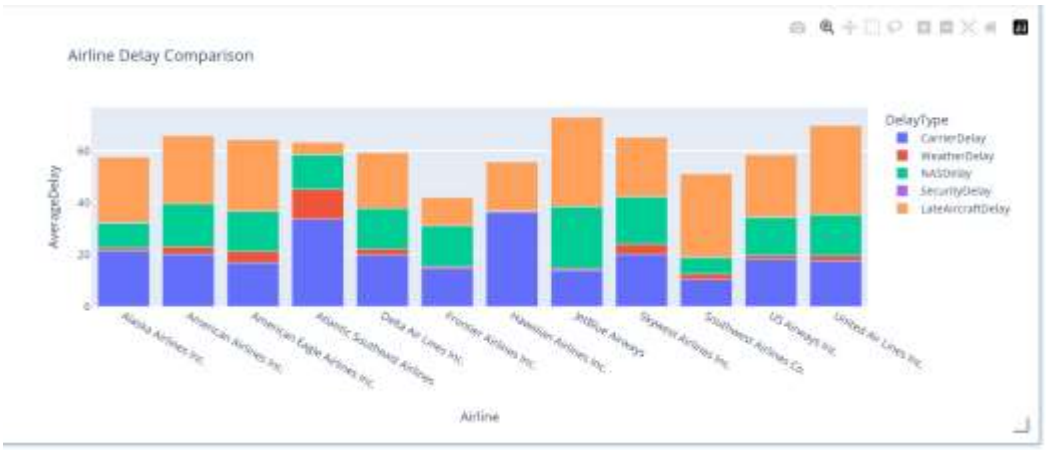
- The other top airports have average delays between 175 and 225 minutes.

- **Correlation Between Delay Causes (Heatmap, Bottom):**
    - All delay causes show **very weak negative or near-zero correlation** with each other.
    - This suggests that an increase in one type of delay (e.g., Weather Delay) is **unlikely to directly cause** or be linked to a predictable change in another type of delay (e.g., Security Delay).

Measures:

- **Create separate management plans** for each delay type instead of a one-size-fits-all approach.
- **Prepare weather-specific strategies** (like standby runways or diversion plans) since weather delays are independent and unpredictable.

7. Airline Comparison



Insight:

- **Airline Delay Comparison (Stacked Bar Chart, Bottom):**
    - The **total average delay** varies between airlines, but **Hawaiian Airlines Inc.** appears to have the highest total average delay.
    - **Carrier Delay** (blue) and **Late Aircraft Delay** (purple) are the **largest components** of the average total delay for almost every airline.
    - **JetBlue Airways** has a particularly large average **Carrier Delay** component.

Measures:

- **Work with high-delay airlines** like Hawaiian and JetBlue to review maintenance and crew scheduling.
- **Encourage sharing of best practices** from low-delay airlines to help others improve.
- **Introduce performance-based incentives** to motivate airlines to reduce their delay times.

Conclusion

The analysis shows that **Late Aircraft** and **Carrier Delays** are the main causes of total delay time, especially during early morning hours.
By improving scheduling, ground operations, staff readiness, and weather planning, airlines and airports can greatly reduce both the frequency and duration of flight delays.
A focus on the worst-performing airlines and airports, along with better coordination and planning, will lead to more reliable and on-time flight operations.

Milestone 3: Route, Cancellation, and Seasonal Insights

Week 5: Route and Airport-Level Analysis

- Top 10 origin-destination pairs
- Delay heatmaps by airport and route
- Maps showing busiest airports and average delays

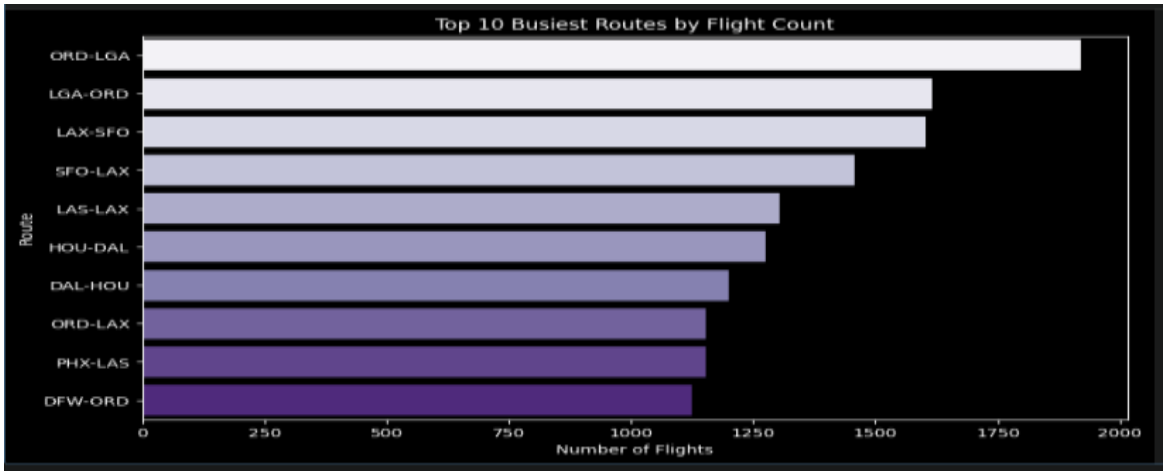
Overview

In Week 5, the analysis focused on understanding flight delays at the **route and airport level**. The main goal was to identify which airports and flight routes experience the most delays, the times and days when delays are highest, and the causes of these delays. This was done using various visualizations, including **heatmaps, bar charts, box plots, and maps**, which allowed us to examine patterns, extreme delays (outliers), and the relationship between flight volume and delays. The analysis also highlights key areas for operational improvement to reduce delays.

Insights and Measures

1. Top 10 Busiest Routes by Flight Count

Process:
A bar chart was used to show the Busiest Route.

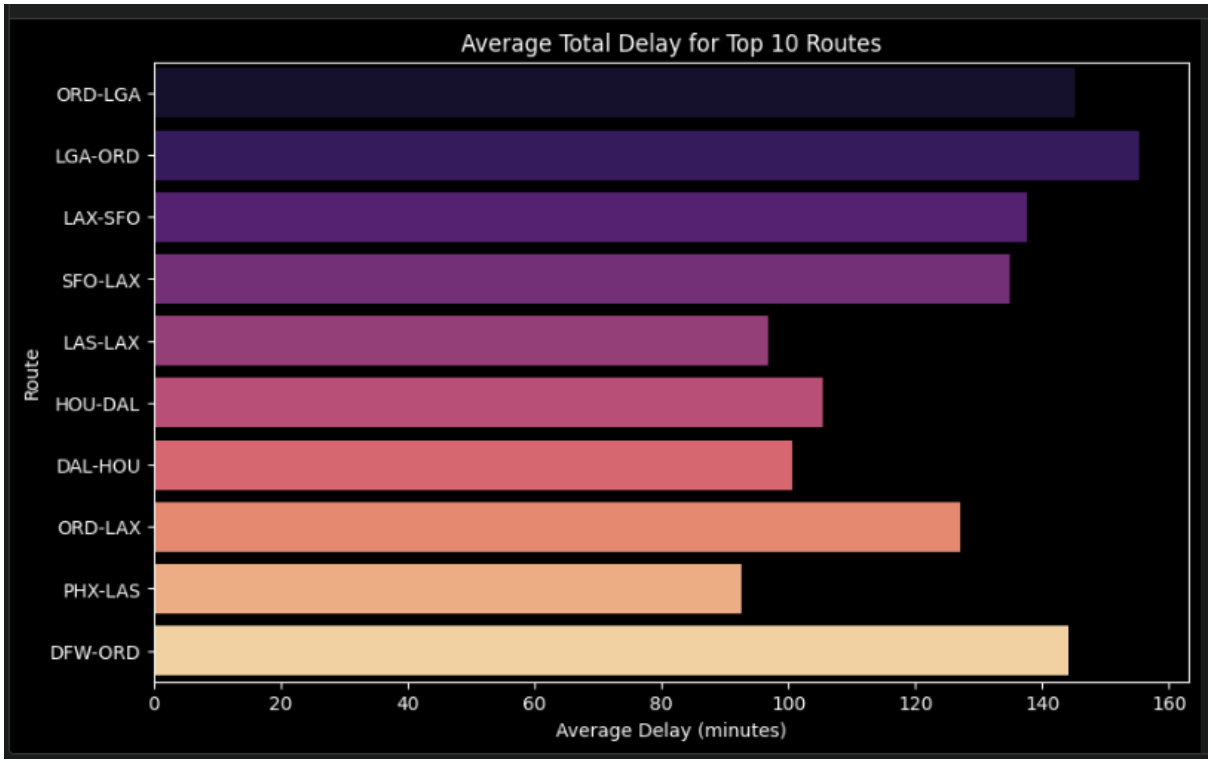


Insight:
• The ORD–LGA(Chicago O'Hare International Airport to LaGuardia Airport) route is the busiest, with nearly 2,000 flights, showing heavy traffic between Chicago and New York.
• Major city pairs like LAX–SFO(Los Angeles International Airport to San Francisco International Airport) and HOU–DAL(William P. Hobby Airport to Dallas Love Field) also dominate, highlighting strong regional and business travel demand.

Measures:
• Optimize scheduling and gate management on top routes to reduce congestion.
• Coordinate flight timings on reciprocal routes (e.g., ORD–LGA and LGA–ORD) for better efficiency.

2. Average Total Delay for Top 10 Routes

Process:
A bar chart was used to visualize the average delay for the top 10 routes.



Insight:

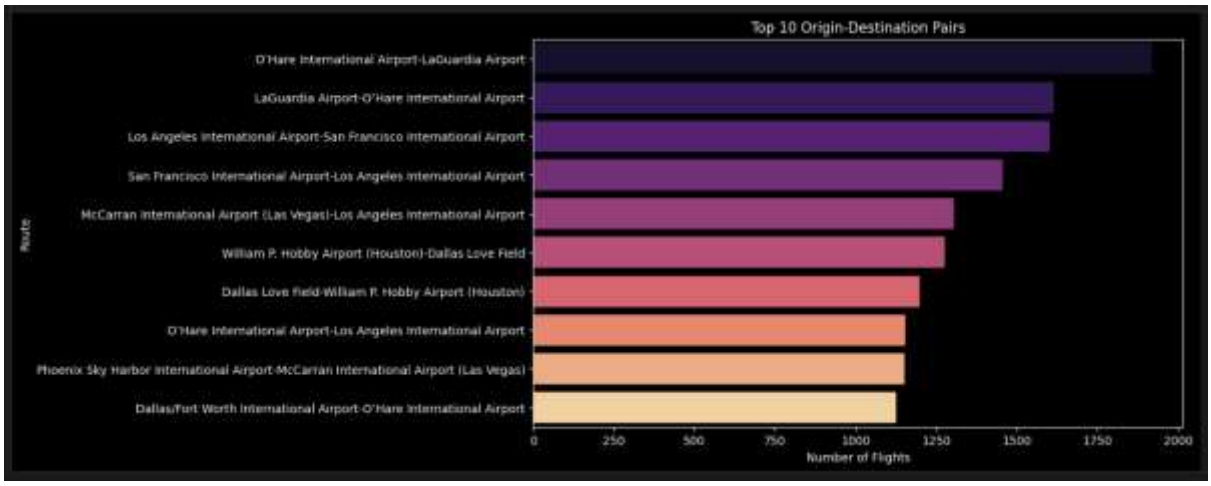
- Longest delays occur on LGA-ORD(LaGuardia Airport to Chicago O'Hare International Airport) (155 min) and ORD-LGA(Chicago O'Hare International Airport to LaGuardia Airport) (145 min).
- Delays range from 95 to 155 minutes across top routes.

Measures:

- Implement proactive monitoring on high-delay routes.
- Schedule buffer time or backup aircraft to reduce ripple effects of delays.

3. Top 10 Origin-Destination Pairs and Busiest Routes

Process:
A table and bar chart were used to rank the top 10 most-flown routes and origin-destination pairs.



Insight:

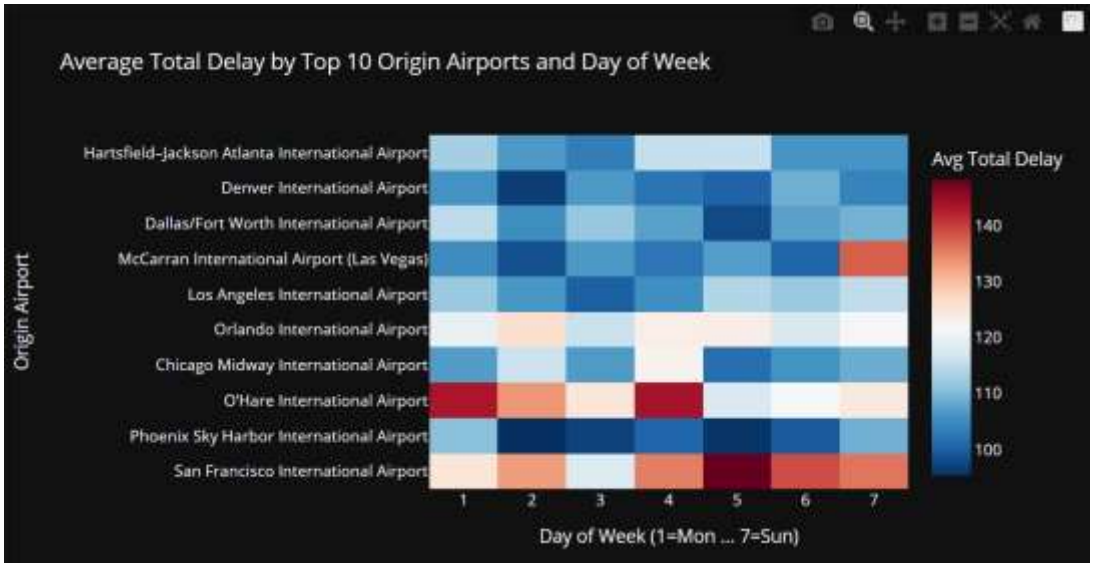
- ORD-LGA(Chicago O'Hare International Airport to LaGuardia Airport) is the most frequent route, followed by LGA-ORD.
- Other high-frequency routes include LAX-SFO(Los Angeles International Airport to San Francisco International Airport) and DFW-HOU(Dallas/Fort Worth International Airport to William P. Hobby Airport).
- The busiest routes also show **high median delays**, indicating that typical flights are often delayed.

Measures:

- Prioritize resources (ground staff, gates, and crews) for busiest routes.
- Monitor aircraft rotation to prevent delays from cascading to subsequent flights.
- Implement contingency plans for extreme delays on high-frequency routes.

4. Average Total Delay by Top 10 Origin and Destination Airports and Day of Week

Process:
Heatmaps were used to show delays for departures at top airports by day of the week.



Insight:

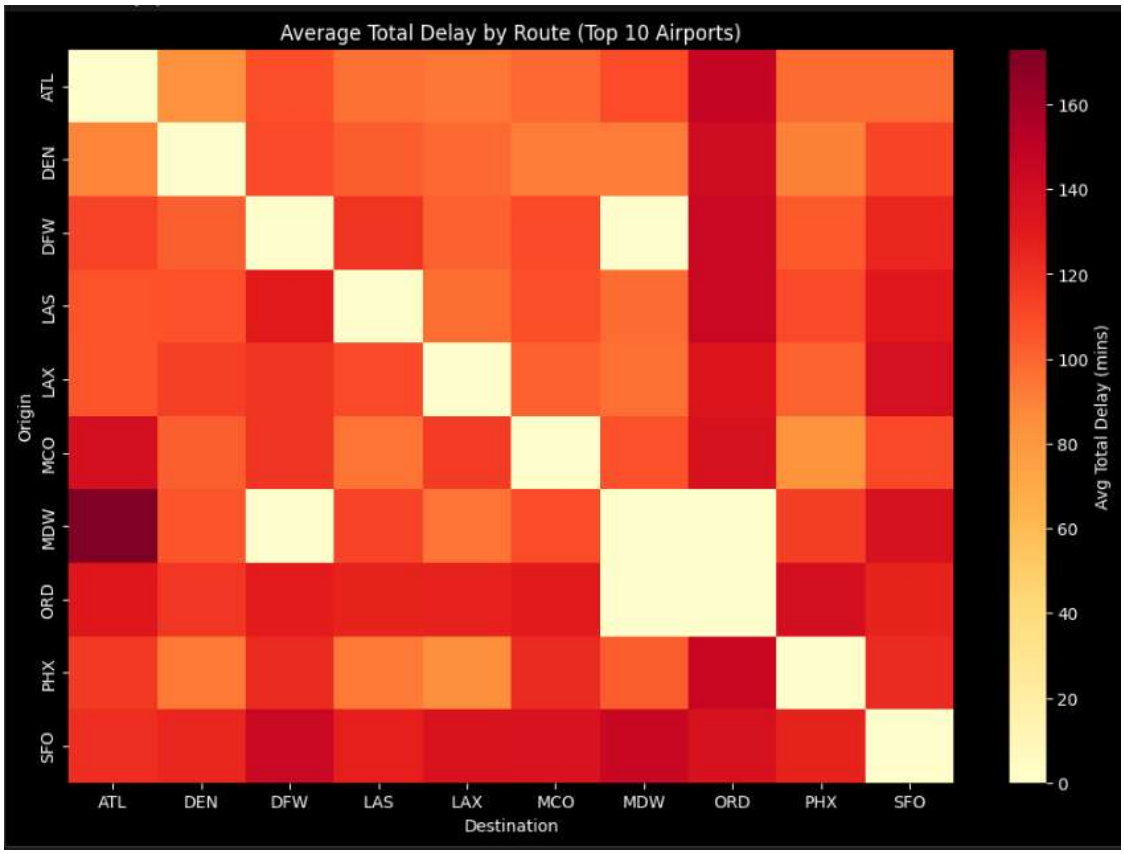
- LAS(Harry Reid International Airport) has highest delays on Sunday (Day 7), while ORD(Chicago O'Hare International Airport) shows high delays on Monday and Thursday.
- Delay patterns vary across the week.

Measures:

- Increase staffing and gate readiness on high-delay days.
- Review weekly schedules to redistribute flights and reduce peak congestion.

5. Average Total Delay by Route (Top 10 Airports)

Process:
A heatmap was created showing average delays for flights between the top 10 airports. Dark red represented long delays, light yellow represented short delays.



Insight:

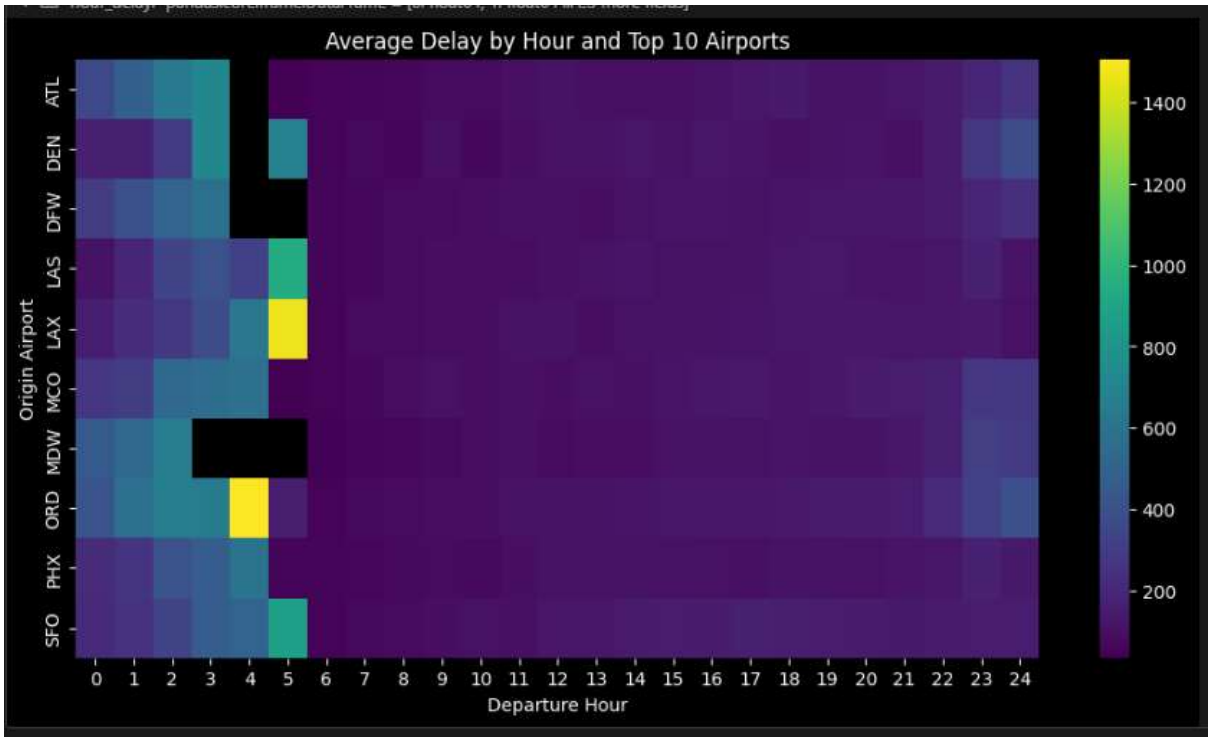
- Longest delays occur on high-traffic routes such as MDW-ATL(Chicago Midway International Airport to Hartsfield–Jackson Atlanta International Airport) and ATL-ORD(Hartsfield–Jackson Atlanta International Airport to Chicago O'Hare International Airport).
- Some routes, like ORD-MDW(Chicago O'Hare International Airport to Chicago Midway International Airport) and PHX-SFO(Phoenix Sky Harbor International Airport to San Francisco International Airport), experience minimal delays.

Measures:

- Target high-delay routes for operational improvements, such as prioritizing timely departures.
- Monitor aircraft rotation to prevent Late Aircraft Delays along busy routes.
- Adjust flight schedules to spread flights more evenly during the day.

6. Average Delay by Hour and Top 10 Airports

Process:
A heatmap was created showing average flight delays at the top 10 busiest airports by departure hour (0–24). The color scale ranged from dark (short delays) to bright yellow (long delays).



Insight:

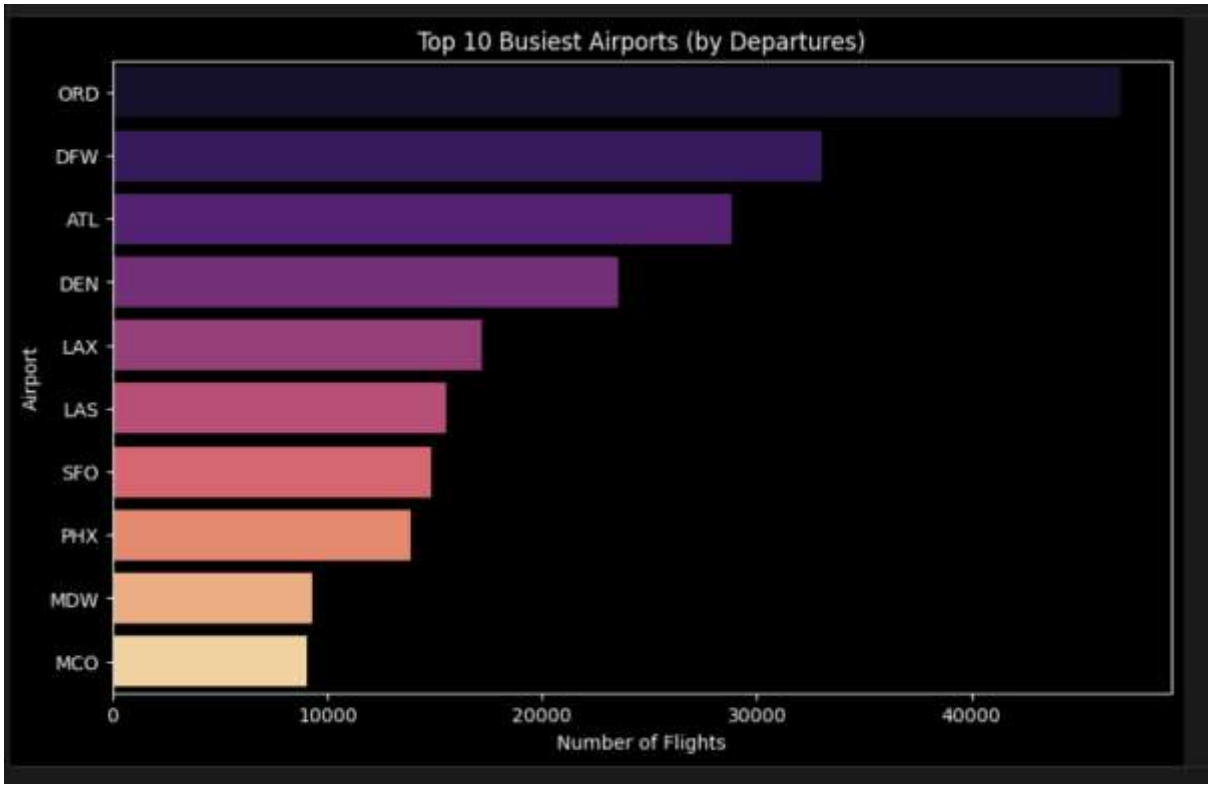
- The longest delays occur very early in the morning (4–6 AM) at ORD (Chicago O'Hare), LAX (Los Angeles), and LAS(Las Vegas).
- Delays are highest in the first few hours of the day and drop significantly during the main day, showing that early flights are more prone to problems.
- Large black blocks indicate hours with very few or no flights.

Measures:

- Adjust early morning schedules at high-delay airports to reduce congestion.
- Increase staffing and ground support for early morning flights of the day.
- Add buffer time between early departures to prevent Late Aircraft Delays from cascading.

7. Top 10 Busiest Airports (by Departures)

Process:
A bar chart was used to rank airports as busiest based on the number of departing flights.



Insight:

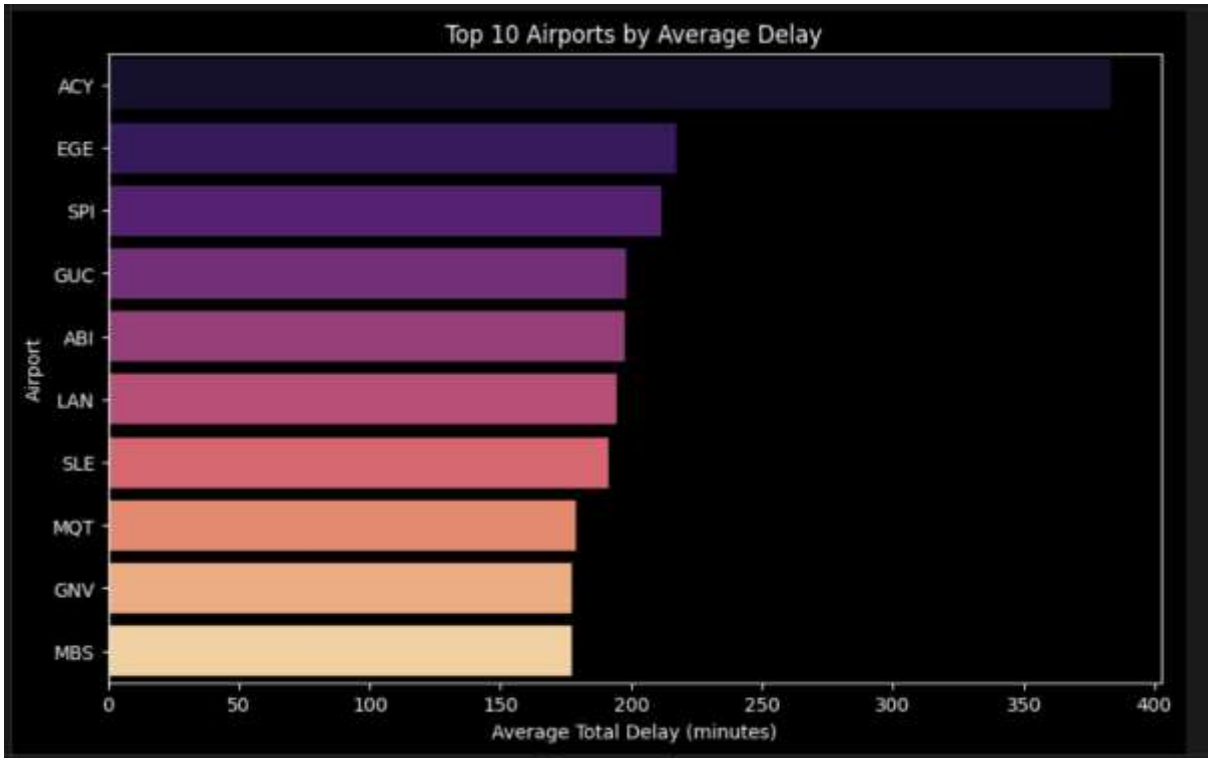
- ORD(Chicago O'Hare International Airport) is the busiest (~45,000 departures), followed by DFW(Dallas/Fort Worth International Airport) (~32,000) and ATL(Hartsfield–Jackson Atlanta International Airport) (~30,000).
- These airports are major hubs, where small disruptions can quickly lead to large-scale delays.

Measures:

- Focus operational improvements on major hubs to ensure high-volume flights are on time.
- Optimize turnaround times and gate assignments to reduce congestion during peak hours.

8. Top 10 Airports by Average Delay

Process:
A bar chart was used to show the average delay per airport.



Insight:

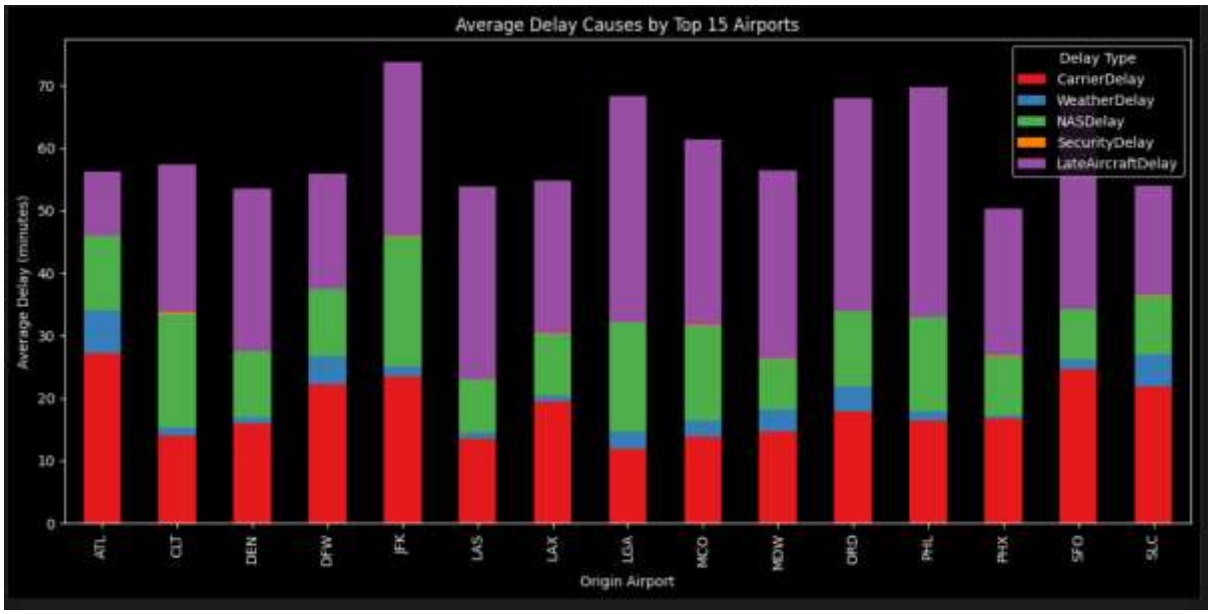
- ACY (Atlantic City) has the highest average delay (~380 min), followed by EGE(Eagle County Regional Airport) (~200 min) and SPI(Abraham Lincoln Capital Airport) (~200 min).
- Smaller airports tend to have higher delays, while larger hubs generally have moderate delays.

Measures:

- Investigate operational inefficiencies at small, high-delay airports.
- Provide targeted support for these airports (crew planning, maintenance).

9. Average Delay Causes by Top 15 Airports

Process:
Stacked bar charts were created to show the contribution of different delay types (Late Aircraft, Carrier, NAS, Weather, Security) for top airports.



Insight:

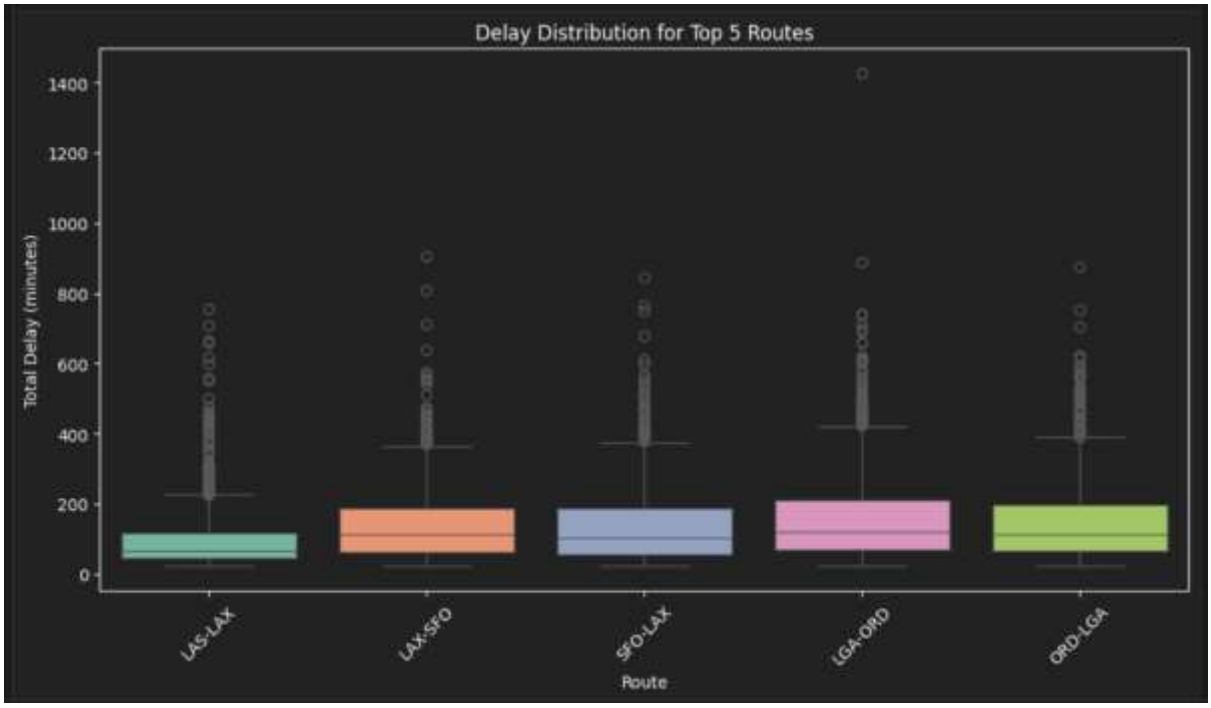
- Late Aircraft Delay is the largest contributor for nearly all airports.
- Carrier Delay is the second-largest, particularly at ATL(Hartsfield–Jackson Atlanta International Airport), DFW(Dallas/Fort Worth International Airport), and PHX(Phoenix Sky Harbor International Airport).
- JFK(John F. Kennedy International Airport) has the highest total delay, with large contributions from Late Aircraft, Carrier, and NAS Delays.

Measures:

- Improve flight rotation and turnaround efficiency to reduce Late Aircraft Delays.
- Strengthen airline operations (crew, maintenance) to reduce Carrier Delays.
- Invest in NAS systems to minimize airspace-related delays at busy airports.

10. Delay Distribution for Top 5 Routes

Process:
Box plots were used to analyze the spread of delays on the top 5 busiest routes.



Insight:

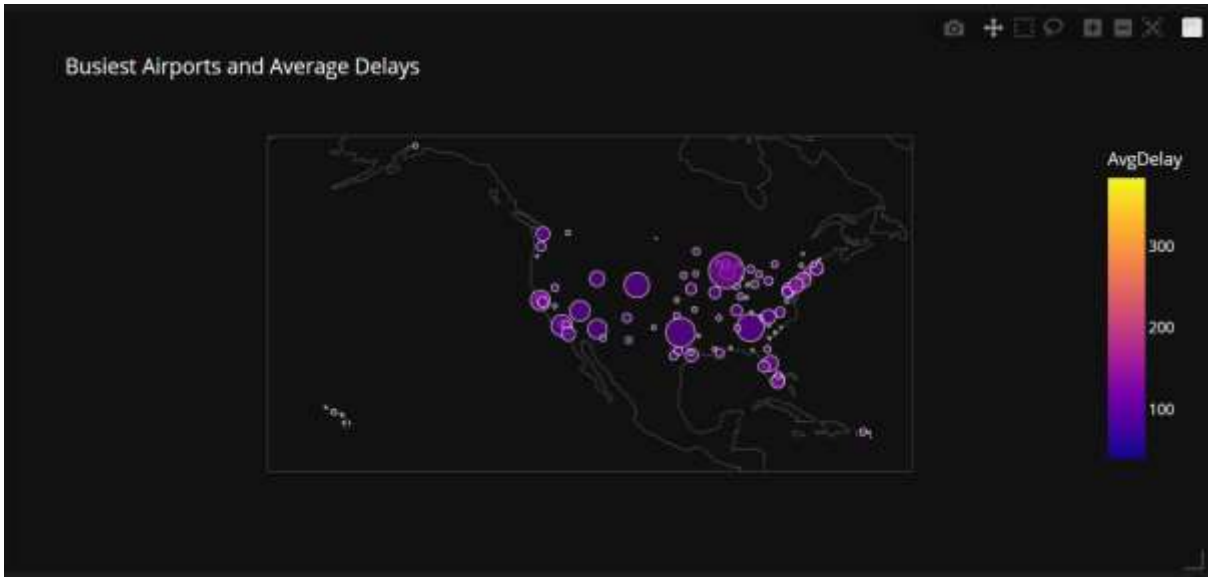
- Median delays for busiest routes are around 100–125 minutes.
- LAS-LAX(Harry Reid International Airport to Los Angeles International Airport) has lower median delays and smaller variation.
- Extreme outliers exist on all routes, with delays over 23 hours (LGA-ORD).

Measures:

- Focus improvement efforts on top delayed routes.
- Implement alert systems for flights approaching extreme delays.
- Schedule buffer flights or backup aircraft to recover from major delays.

11. Busiest Airports and Average Delays (Geographical Map)

Process:
A map visualized airport location, flight volume (circle size), and average delay (color).



Insight:

- Largest airports are clustered in Eastern and Midwestern US.
- East Coast and Midwest airports generally have higher average delays.
- West Coast airports usually have lower delays, though some exceptions exist.

Conclusion:

Overall, the route and airport-level analysis shows that flight delays are most severe during early morning hours, at specific high-traffic airports, and along major routes such as O'Hare–LaGuardia. Late Aircraft and Carrier Delays continue to be the biggest contributors. By improving scheduling, increasing early-morning staffing, and focusing on operational efficiency at key airports and routes, airlines can greatly reduce total delay times and improve on-time performance.

Week 6 Report: Seasonal and Cancellation Analysis

- Monthly cancellation trends
- Cancellation types: carrier, weather, security, NAS
- Analyze impact of holidays or winter months

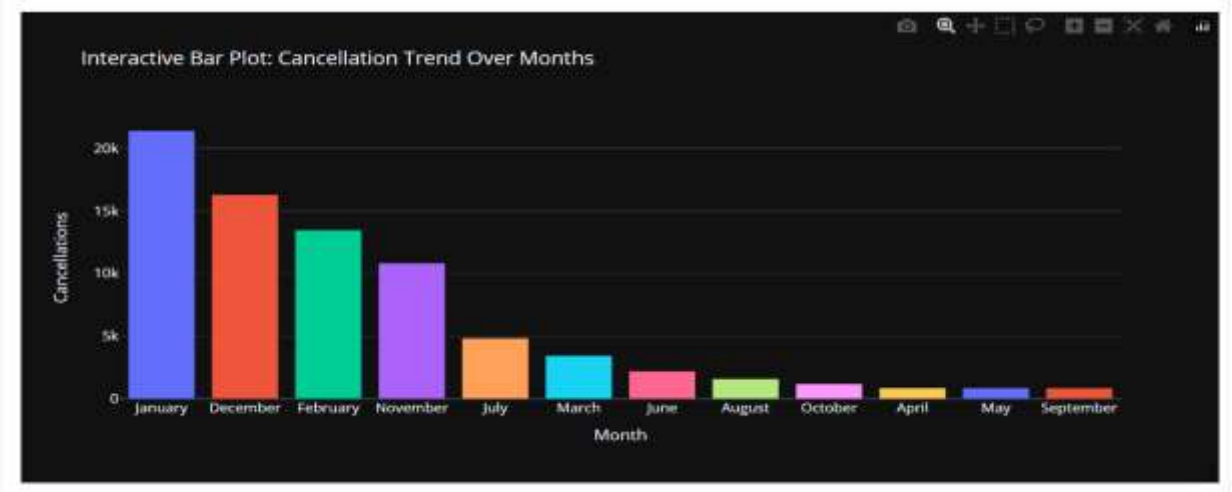
Overview

This week's task focused on analyzing flight cancellation trends with respect to **seasonal patterns** and **types of cancellation reasons**. The goal was to identify when and why cancellations occur most frequently, assess the impact of winter months and holidays, and suggest data-driven measures to minimize disruptions.

The analysis used multiple visualizations — including bar charts, heatmaps, and comparative graphs — to uncover insights into monthly trends, major causes of cancellations, and airline-specific patterns across different seasons.

1. Monthly Cancellation Trends

Process:
An interactive bar chart was used to examine the number of cancellations across each month.



Insight:

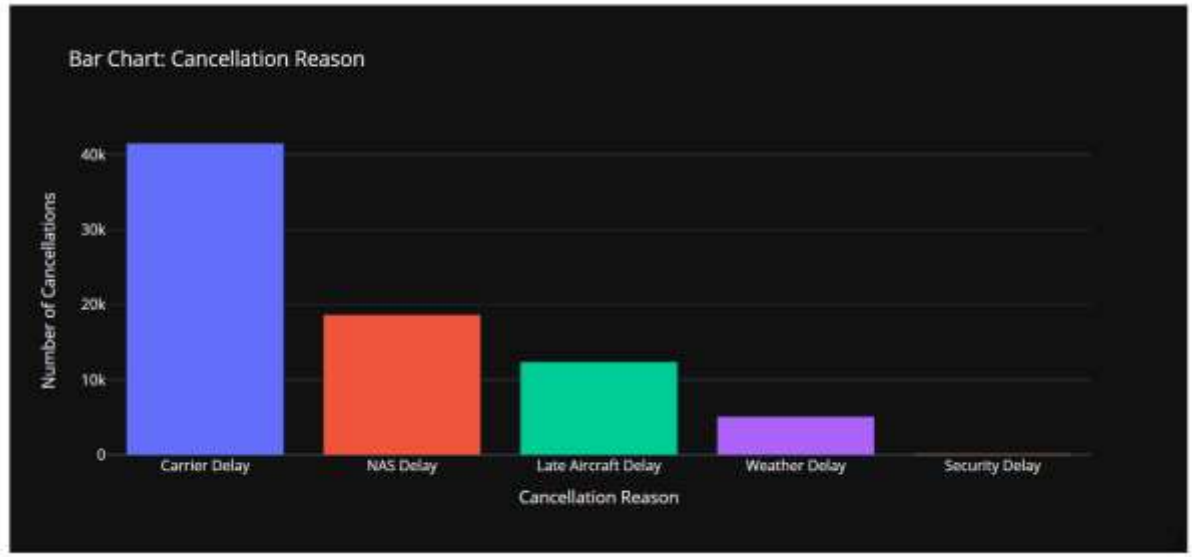
- **January** recorded the highest cancellations (over **20,000**), followed by **December** (~16,000) this can be due to snowfall in many areas.
- **September** showed the lowest number, with cancellations just above zero.
- Cancellations were concentrated in **winter months (December, January, February, November)**, while **spring and late-year months** had the fewest.

Measures:

- Investigate root causes for high cancellations during winter (e.g., weather conditions, post-holiday travel stress, or operational strain).
- Launch **retention and reliability campaigns** during high-risk months.
- Collect and analyze **customer feedback** to identify recurring issues.

2. Cancellation Reasons

Process:
A bar chart visualized the number of cancellations by reason to determine the most frequent causes.



Insight:

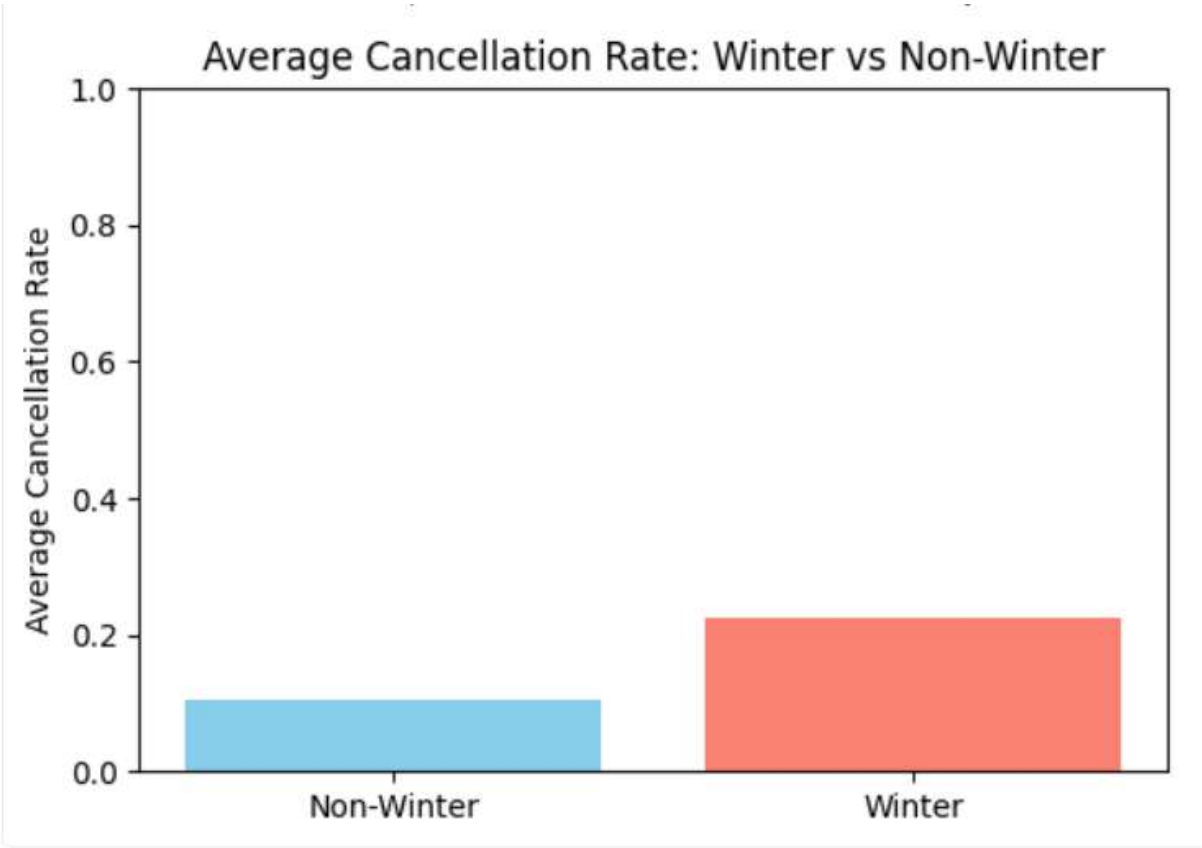
- **Carrier Delay** was the **top reason**, accounting for over **40,000 cancellations**.
- **NAS Delays** followed (~18,000–20,000), often linked to air traffic or system issues.
- **Late Aircraft Delays** contributed around **12,000**, while **Weather** and **Security** delays had minimal impact.

Measures:

- Improve **carrier-level efficiency** through better maintenance, crew, and scheduling.
- Collaborate with **air traffic authorities** to minimize NAS-related delays.
- Strengthen **weather monitoring systems** and implement **operational contingencies**.

3. Average Cancellation Rate: Winter vs Non-Winter

Process:
A comparative bar chart was used to measure average cancellation rates between winter and non-winter seasons.



Insight:

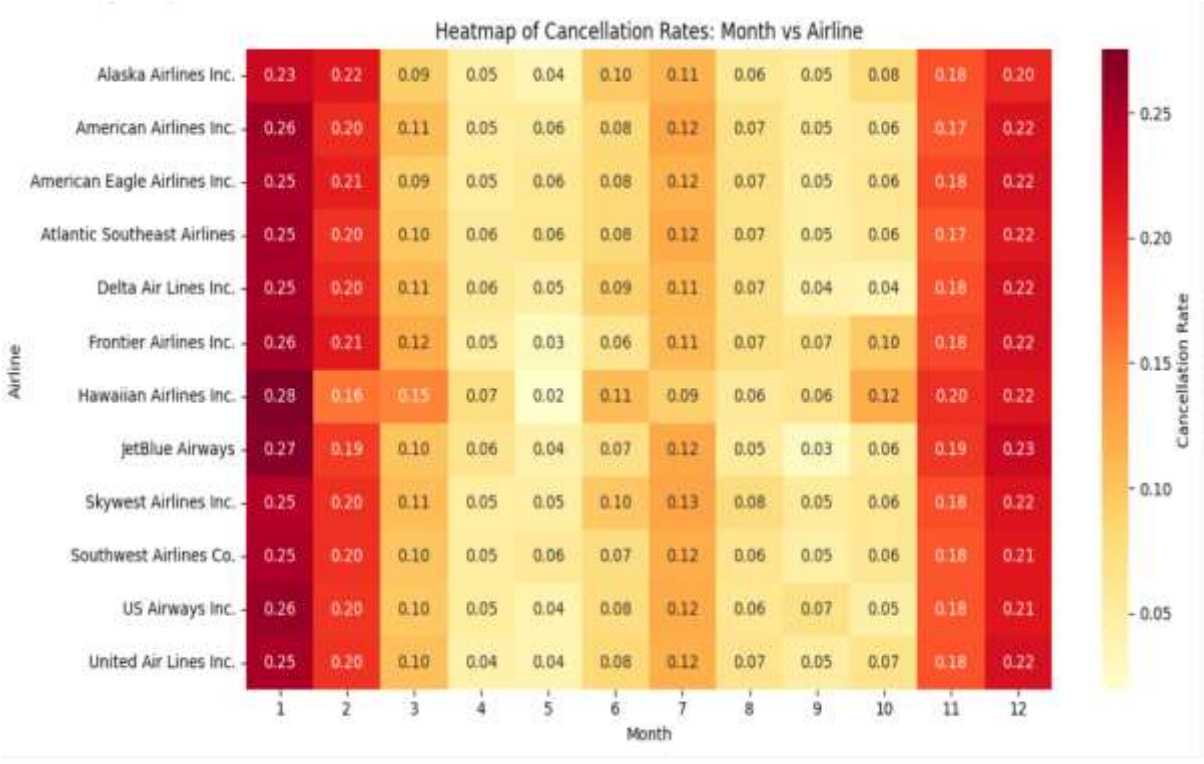
- **Winter** showed a **higher average cancellation rate (≈0.23)** than **non-winter (≈0.10)**.
- The difference indicates that weather disruptions and seasonal travel surges strongly affect operations.

Measures:

- Enhance **winter preparedness**, including de-icing and weather response systems.
- Adjust **flight schedules and staffing** for flexibility during peak disruption months.
- Provide **proactive communication** to passengers during winter months.

4. Heatmap of Cancellation Rates: Month vs Airline

Process:
A heatmap was used to analyze cancellation rates by airline across each month.



Insight:

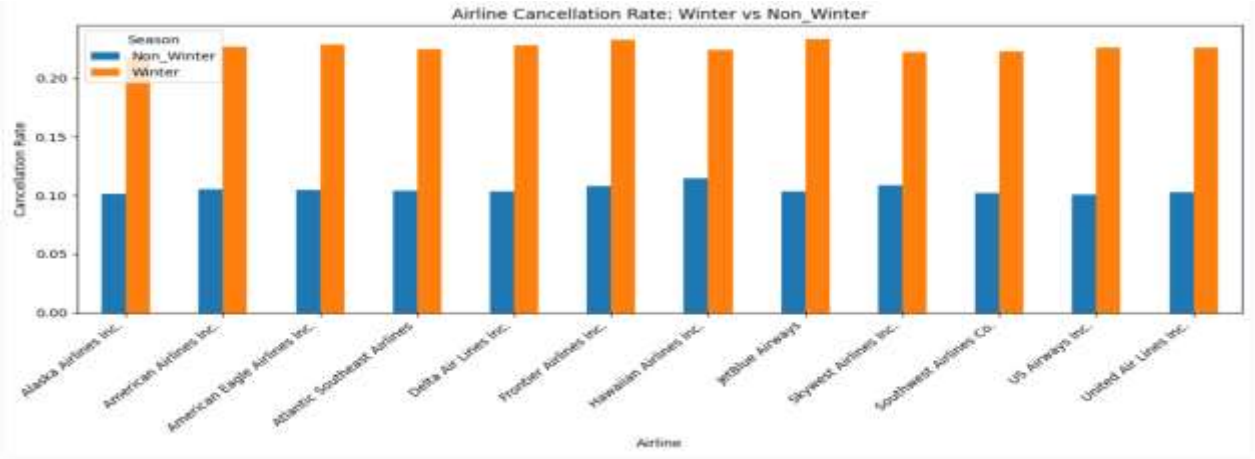
- **January and February** consistently had the **highest cancellation rates** across all airlines (0.23–0.28).
- **Mid-year months (April–September)** showed the **lowest rates**, below 0.1.
- Airlines such as **Hawaiian Airlines** and **JetBlue Airways** had slightly higher winter peaks than others.
- Rates increased again during **November and December**, showing clear **seasonal repetition**.

Measures:

- Improve **winter operation planning** across all airlines.
- Conduct **airline-specific performance analysis** to address internal issues.
- Ensure **preventive maintenance and resource allocation** before high-risk months.

5. Airline Cancellation Rate: Winter vs Non-Winter

Process:
A grouped bar chart compared cancellation rates for each airline during winter and non-winter seasons.



Insight:

- All airlines experienced **higher cancellation rates during winter**, averaging **0.20–0.23**, versus **0.10** in non-winter months.
- The pattern was consistent across all major carriers, confirming that **seasonal weather** is a key disruption factor.

Measures:

- Implement **comprehensive winter-readiness strategies**, including staff training and equipment checks.
- Optimize **flight scheduling** to include buffer time during high-risk months.
- Use **predictive analytics** to anticipate and mitigate cancellation risks.

Conclusion

The analysis clearly demonstrates that **seasonal factors—especially during winter months—have a major impact on flight cancellations.** January and December consistently show the highest cancellation rates, driven primarily by **carrier-related issues** and **weather conditions**.

To reduce future cancellations, airlines should focus on:

- **Improving operational efficiency** through maintenance and scheduling enhancements.
- **Strengthening winter preparedness** and coordination with air traffic control.
- **Leveraging data analytics** to predict and proactively respond to disruption risks.

By addressing these areas, airlines can improve reliability, enhance customer satisfaction, and ensure smoother operations throughout the year.

Milestone 4: Report and Presentation

Week 7: Visual Report or Dashboard Preparation

- Combine plots into a coherent storyline
- Use markdown, presentation slides, or Streamlit
- Ensure plots include labels, titles, legends, and axis clarity

Introduction

This report presents a comprehensive analysis of airline operations and performance using Power BI dashboards.
The visualizations collectively narrate the story of how airlines perform in terms of flight volume, delays, cancellations, and route patterns across different months and airports.

The goal of this visual report is to combine multiple analytical perspectives into a coherent storyline that uncovers insights on:

- The scale of flight operations,
- The causes and patterns of delays,
- Airline performance comparisons,
- Airport and route-level analysis, and
- Trends in delays and cancellations over time.

Through this interconnected dashboard design, we can track the journey from a broad dataset overview to specific airline insights and finally conclude with key findings and recommendations.

Storyline Across Dashboards

1. Dataset Overview – Understanding the Scale of Operations

- The story begins with the Dataset Overview Dashboard, which introduces the overall flight dataset and provides key performance indicators (KPIs). Here, we see the magnitude of the data with 483.37K total flights, covering an average distance of 753.45 miles per flight.
- The total recorded delay amounts to nearly 29 million minutes, showing how significant delay management is within the aviation industry. Despite these delays, the percentage of cancelled flights remains extremely low (0.09%), reflecting strong operational reliability.
- Additionally, the pie chart reveals that Southwest Airlines holds the largest share of total flights, followed by American Airlines and United Airlines. This dashboard sets the foundation for the entire report — providing a macro-level understanding before diving deeper into performance and delay specifics.

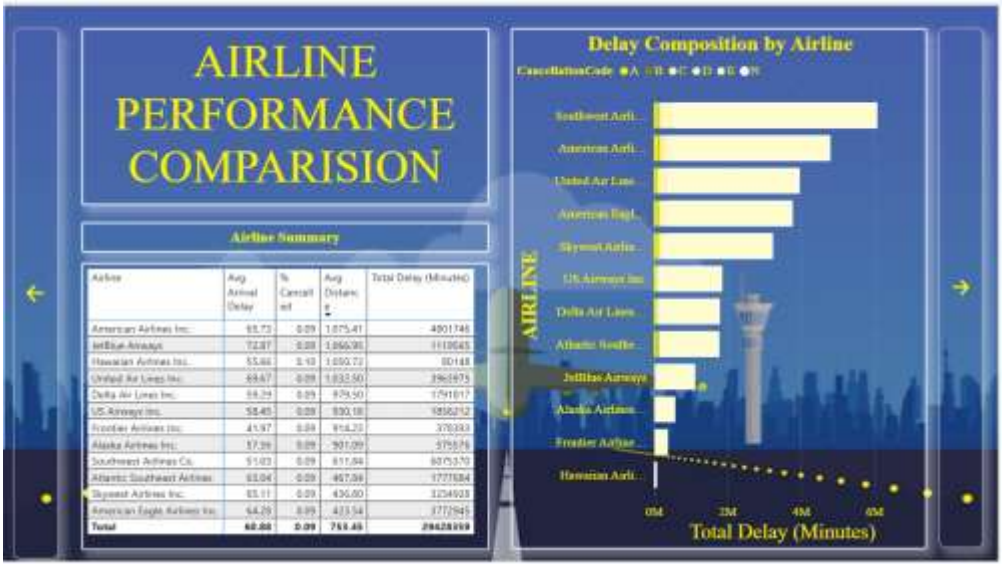
2. Flight Summary – Exploring Delay Components

- The Flight Summary Dashboard examines the types of delays that contribute to total delay minutes.
  It breaks down delays into five major categories: Carrier, NAS (Air Traffic), Late Aircraft, Weather, and Security delays.
- Among these, Late Aircraft Delay (26.68 mins) and Carrier Delay (17.40 mins) are the most significant contributors, while Weather (3.15 mins) and Security (0.08 mins) have minimal impact.
- This indicates that operational and scheduling inefficiencies within airlines — rather than external factors — are the leading causes of flight delays.
- The dashboard also presents overall statistics such as total delay (29M mins), average distance (753.45 miles), and most recent flight data (30-06-2019), tying back to the dataset scale introduced earlier.

3. Airline Performance Comparison (Part 1) – Delay Distribution Across Airlines

- In this dashboard, we shift focus from overall causes to airline-level performance.
  Here, total delay minutes are compared among major airlines, along with their average arrival and departure delays.
- Southwest Airlines records the highest total delay (5.5M minutes), followed by American Airlines (4.4M) and United Airlines (3.6M). However, this is partially due to their higher flight volumes.
- The bar charts also reveal that JetBlue Airways experiences the highest average arrival and departure delays, while Hawaiian Airlines maintains consistently low delay averages — making it one of the most punctual carriers.

4. Airline Performance Comparison (Part 2) – Summary Metrics



This dashboard provides a comprehensive numerical summary of airline performance. Each airline's average arrival delay, cancellation rate, average distance, and total delay minutes are displayed side by side for direct comparison.

From this summary:

- JetBlue Airways emerges as the airline with the highest average delay (72.87 mins).
- Frontier Airlines records the lowest delay (41.97 mins), highlighting its operational efficiency.
- Delta Air Lines performs strongly, with balanced delay and cancellation metrics.
- Hawaiian Airlines stands out for maintaining low delays and cancellations despite longer average distances.

5. Flight Volume, Airline & Airport Analysis – Identifying Key Hubs

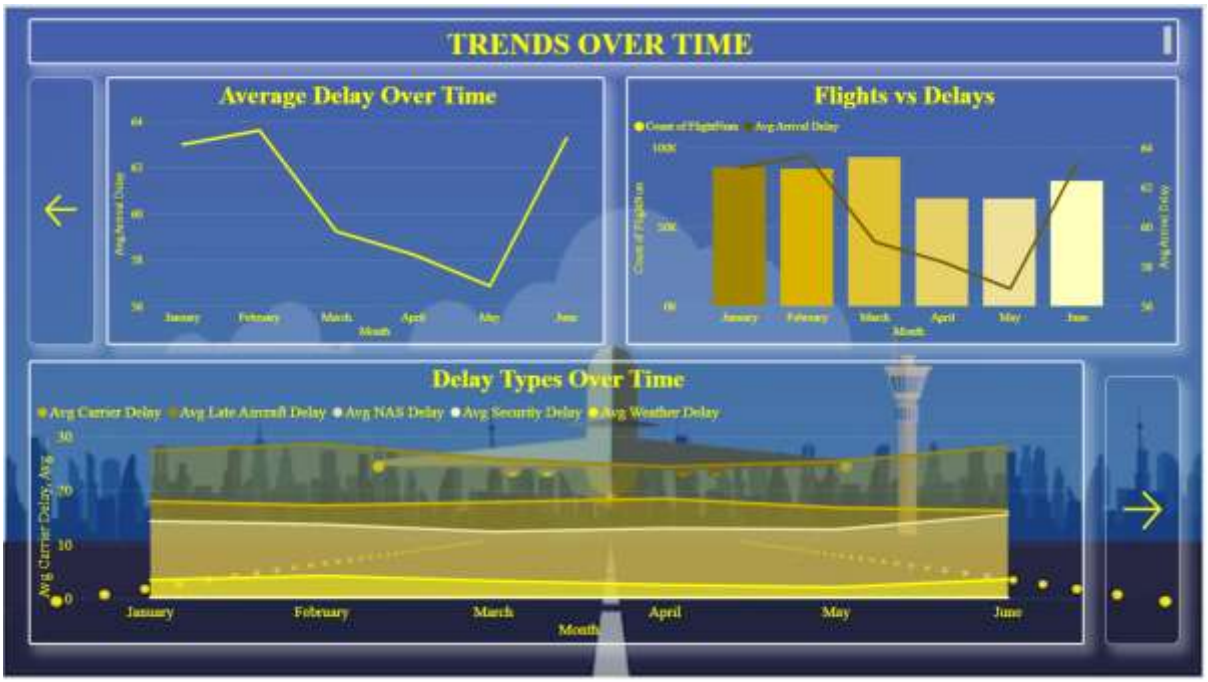
The Flight Volume & Airport Analysis Dashboard zooms out to examine flight patterns geographically.
The U.S. map visualization displays flight concentration by origin airport, clearly identifying major hubs like:

- ORD (Chicago O'Hare)
- DFW (Dallas–Fort Worth)
- ATL (Atlanta Hartsfield–Jackson)

These airports handle the highest flight volumes, suggesting they are critical to national air traffic flow.
A pie chart complements this by showing that American Airlines covers the greatest total distance (79M miles), followed closely by Southwest and United Airlines.

This dashboard connects airline operations to geography — highlighting how high-volume airports contribute to systemic congestion and delays.

6. Trends Over Time – Monthly Delay Patterns



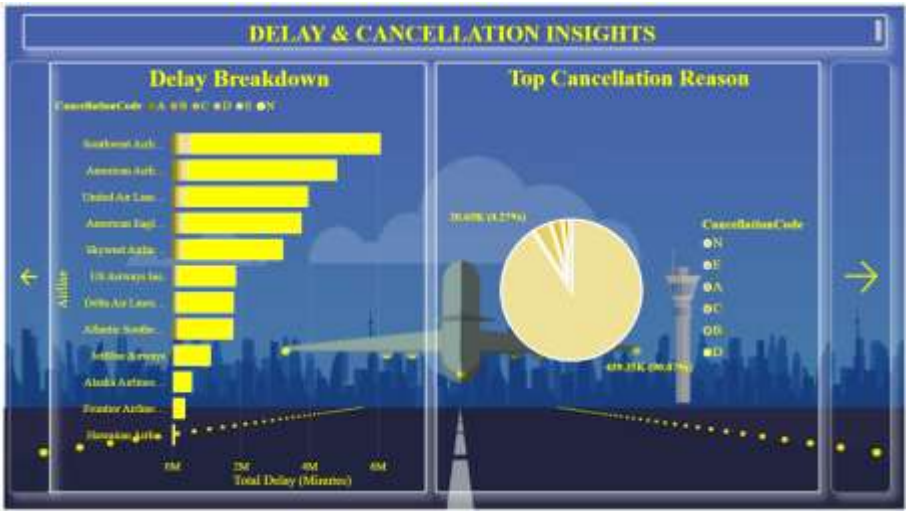
The Trends Over Time Dashboard visualizes changes in average delays from January to June.
The trend line shows:

- A peak in February,
- A decline through April and May, and
- A rise again in June.

This suggests that mid-year months experience increased operational and weather-related challenges.

The lower section shows delay types over time, where Late Aircraft and Carrier Delays remain consistently higher, while Weather and Security Delays fluctuate seasonally.

7. Delay & Cancellation Insights (Part 1) – Breakdown by Airline



- This dashboard compares total delay minutes by airline alongside the distribution of cancellation reasons.
- Southwest Airlines again leads in total delay minutes, followed by American Airlines and United Airlines.
  However, the pie chart on the right shows that the majority of flights (over 90%) are not cancelled (Code N).
  Among actual cancellations, Reason E appears most frequently, which is typically related to external factors such as weather or airspace restrictions.

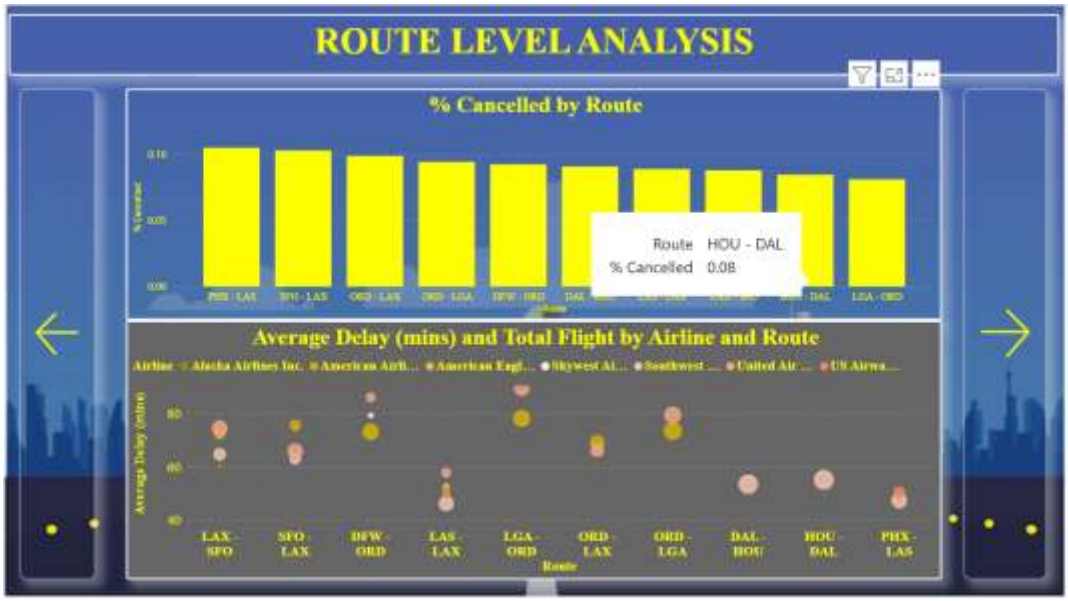
8. Delay & Cancellation Insights (Part 2) – Cancellation Summary

The second cancellation dashboard quantifies the total number of cancelled flights and identifies key patterns:

- Total Cancelled Flights: 44.12K
- Highest Cancellation Count: 21K
- Top Cancellation Reason: Code E
- Highest Cancellation Rate: Hawaiian Airlines (0.10%)

Despite these figures, the overall cancellation rate remains exceptionally low (0.09%), indicating strong service reliability across the industry.

9. Route Level Analysis – Understanding Performance by Flight Route



- The final dashboard ties together all insights into a concise summary of performance outcomes.

10. Summary & Key Findings – Drawing Conclusions

The final dashboard ties together all insights into a concise summary of performance outcomes.

Key Results:

- Best Airline: Frontier Airlines Inc. (lowest delay average of 41.97 mins)
- Worst Airline: JetBlue Airways (highest average delay of 72.87 mins)
- SkyWest Airlines operates the most flights but experiences moderate delays.
- Delta Air Lines stands out for strong on-time performance and balanced metrics.
- Weather delays peak during mid-year months.
- Overall cancellation rate: 0.09%, with a maximum of 0.10% for Hawaiian Airlines.

These insights collectively tell a clear story — while delays are inevitable, most airlines maintain high service continuity with minimal cancellations. Operational efficiency, rather than weather, remains the main differentiator in airline performance.

Conclusion

This connected series of Power BI dashboards narrates the complete performance journey of U.S. airlines — from the dataset's scale to granular insights on delay patterns, airport dynamics, and cancellations.

The storyline demonstrates how data visualization can transform raw numbers into actionable insights:

- Identifying top-performing airlines,
- Understanding delay causes and trends, and
- Recognizing patterns in airport congestion and weather effects.

Through this visual storytelling approach, the dashboards provide a comprehensive, data-driven understanding of airline performance — empowering decision-makers to improve efficiency, minimize delays, and enhance customer satisfaction.

Week 8: Documentation and Final Presentation

- Create final report (PDF or README)
- Build slide deck for presentation
- Record insights and visual walkthrough

Overview

Week 8 focused on compiling all project outcomes into clear, polished documentation and preparing the final presentation materials. This included creating a comprehensive final report, assembling a slide deck that summarizes the workflow and results, and recording key insights through a visual walkthrough of the analysis. The goal of this week was to consolidate the entire project into professional, shareable, and easily interpretable formats.

1. Final Report Creation (PDF / README)

Objectives

- Produce a complete, well-structured final report that covers the entire project lifecycle—from dataset loading to insights and business recommendations.
- Ensure that all methodology, code processes, cleaning steps, and analytical outputs are clearly documented.
- Provide a README-friendly version for GitHub or public submission.

Included in the Final Report

1. **Project Overview**
    - Purpose of the Airfly Insights project
    - Dataset description and origin
    - High-level objectives and KPIs
2. **Technical Workflow**
    - Week-by-week development stages
    - Data loading, preprocessing, cleaning, and feature engineering
    - Memory optimization and schema validation
3. **Methodology**
    - Detailed explanation of data cleaning logic
    - Justification for imputation strategies
    - Feature engineering approach and rationale
4. **Exploratory Analysis**

- Summary statistics
- Delay distributions
- Flight volume patterns
- Airport-level and route-level insights

5. **Visualizations**
    - Delay histograms
    - Airline on-time performance charts
    - Route efficiency maps
    - Airport delay comparisons
6. **Key Insights**
    - Operational inefficiencies
    - High-delay airports
    - Seasonal and weekly behavior patterns
    - Identified business opportunities
7. **Conclusion & Recommendations**
    - Proposed improvements for airlines and airports
    - Potential predictive modeling directions
    - Future scope for system automation
8. **Appendices**
    - Full column dictionary
    - Data transformations
    - Code snippets (optional)

The final report represents the complete narrative of the project from beginning to end.

2. Slide Deck for Final Presentation

Purpose

To create a concise, visually engaging summary of all analytical findings, optimized for a live or recorded presentation.

Slide Deck Structure

1. **Title Slide**
    - Project name: *Airfly Insights*
    - Your name and affiliation

2. **Project Motivation**
    - Why flight delay analysis matters
    - Real-world value for airlines and airports
3. **Dataset Summary**
    - Size, time coverage, columns, engineering steps
4. **Data Cleaning & Preparation**
    - Null handling
    - Duplicate removal
    - Type conversions
5. **Feature Engineering**
    - Overview of all newly created features
    - Their contribution to modeling and insights
6. **Exploratory Analysis**
    - Distance metrics
    - Weekly flight volume
    - Taxi-in and taxi-out patterns
7. **Airport Insights**
    - Busiest airports
    - Airports with highest delays
8. **Route Insights**
    - Best and worst performing routes
    - Long-haul analysis
9. **Business Insights**
    - Operational efficiency observations
    - Delay contributors
    - Recommendations
10. **Conclusion & Next Steps**
    - Machine learning readiness
    - Future development
11. **Q&A Slide**

The slide deck allows the audience to understand the project in under 10 minutes.

3. Recorded Insights and Visual Walkthrough

Purpose

To produce a narrative explanation of the findings supported by visuals, demonstrations, and dataset exploration.

Walkthrough Components

- **Introduction:** Overview of the project and what the viewer will learn
- **Dataset Tour:** Visual display of columns, types, and scope
- **Cleaning Demonstration:** Before-and-after examples of nulls, duplicates, and conversions
- **Feature Engineering Explanation:** Why each engineered feature was created and how it affects analysis
- **Insight Narration:**
    - Delay trends
    - Airport and route performance
    - Key statistics narrated with supporting charts
- **Dashboard-style view:** (If applicable) Interactive visuals or plots shown while explaining outcomes
- **Conclusion:** Summary of actionable insights

This walkthrough ensures the audience can quickly understand the technical and analytical journey.

Conclusion

The **Airfly Insights** project provides a comprehensive, end-to-end analysis of U.S. domestic airline operations, transforming a large and complex dataset into meaningful insights through rigorous data preprocessing, exploratory visualizations, route-level evaluations, delay pattern studies, and interactive dashboard development. Over the course of multiple milestones and weekly tasks, the project evolved from basic dataset exploration to an advanced, insight-driven analytical narrative that highlights the operational realities and performance differences within the aviation industry.

The analysis shows that while most flights operate on schedule, **delays are heavily influenced by operational factors**, particularly *Late Aircraft Delays* and *Carrier Delays*, which consistently emerge as the largest contributors across airlines and airports. External factors like weather and security play smaller roles, although seasonal impacts—especially during winter months—significantly affect cancellations. Major airports such as **Chicago O'Hare (ORD)**, **LaGuardia (LGA)**, **Los Angeles (LAX)**, and **Atlanta (ATL)** are high-traffic hubs where delays accumulate rapidly, especially during early morning hours.

Route-level analysis reveals that some of the busiest and most economically important routes—such as ORD–LGA and LGA–ORD—also experience the longest delays, indicating a need for improved scheduling, ground operations, and aircraft rotation planning. Seasonal and cancellation analyses further highlight the vulnerabilities of airlines during harsh winter months and peak travel seasons.

The Power BI dashboards developed in the later phases consolidate these insights into a clear, visually compelling storyline, making the results accessible for decision-makers and operational teams. By integrating technical analysis with business-oriented visual summaries, the project demonstrates how data visualization can support more informed planning, resource allocation, and long-term operational improvements.

Overall, Airfly Insights successfully showcases the power of data analytics in aviation. The findings emphasize the importance of operational efficiency, proactive delay management, route optimization, seasonal readiness, and data-driven decision-making. The project not only serves as a strong analytical case study but also lays the foundation for future work such as predictive delay modeling, real-time monitoring dashboards, and enhanced airline performance strategies.