# **MIS41040 Assignment 2025**

**Team 34** **Team Members:**

* Guozhi Wang – 24202446
* Yixin Luo – 24301543
* Miguel Olaya – 24204642

**Link to Public Power BI Dashboard:[[1]](#footnote-1)**<https://app.powerbi.com/links/hAoDmzLZS1?ctid=420ec589-a866-4ad0-9a57-e6049e0d3bc0&pbi_source=linkShare>

## **1. Executive Summary**

The Decision Support System aims to optimize airline operations by analyzing flight delay patterns and their root causes. Leveraging data from sources like the U.S. Census Bureau and aviation entities, the system integrates real-time and historical data to provide actionable insights. The analysis identifies aircraft-related issues as the leading cause of delays, contributing to 40.03% of disruptions, closely followed by carrier-specific factors at 36.43%. National Air System congestion accounts for 18.09% of delays, underscoring challenges tied to high-traffic hubs. While weather-related delays represent a smaller proportion, their impact becomes more pronounced when evaluated through seasonal and temporal trends, revealing recurring patterns during adverse conditions.

The average delay duration is 15 minutes for both arrivals and departures. Seasonal variations and temporal disruptions further influence delays, with older aircraft likely contributing to maintenance-related delays. NAS delays correlate with high-traffic airports, while short-haul flights, with an average duration of 140 minutes, may experience distinct delay patterns compared to long-haul routes.

## **2. Information Requirements for Decision-Making**

### **2.1 Context of the Problem**

Flight delays represent a persistent and systemic issue within the airline industry, affecting both operational efficiency and customer satisfaction. According to the dataset provided in this project, sourced from Kaggle’s comprehensive compilation of U.S. civil flight data for 2023, delays can arise from multiple factors such as staffing shortages, weather disruptions, air traffic congestion, and aircraft maintenance checks. These delays are defined as the discrepancy between the scheduled and actual departure or arrival times.

The consequences of delays are not isolated to individual flights—they often result in cascading disruptions. For example, a late-arriving aircraft on a prior leg can cause delays across several subsequent flights, especially in tightly scheduled hub-and-spoke systems. This domino effect strains crew schedules, causes missed connections for passengers, and forces rescheduling of aircraft maintenance and gate slots.

From an operational standpoint, delays are costly. Airlines may need to compensate passengers, pay for additional staffing hours, and rearrange logistics at short notice. From the customer perspective, delays erode satisfaction, lead to missed appointments, and in some cases, contribute to lost business. For airports and policymakers, systemic delays reflect broader infrastructural or policy-level inefficiencies that demand coordinated solutions.

Understanding delay patterns—both in terms of causes and impacts—is therefore essential. A data-driven Decision Support System (DSS) can assist stakeholders in pinpointing problems and optimizing decision-making, whether it's airport runway scheduling, airline crew deployment, or national transportation planning.

### **2.2 Stakeholders and Their Needs**

To design an effective DSS, we must first identify the key stakeholders and align our analytical framework with their distinct information needs:

#### **Airline Operations Managers**

These individuals are concerned with internal efficiency and the minimization of operational disruptions. Their key informational requirements may include:

* Identification of peak delay periods (by hour/day/month).
* Detection of airports with recurrent delay issues.
* Correlation between staffing issues (e.g., crew availability or shift overlaps) and delay spikes.
* Aircraft turnaround times, especially in congested airports.

#### **Airport Authorities**

Airport authorities are tasked with optimizing runway capacity, air traffic flow, and passenger experience within the physical constraints of airport infrastructure. They may require:

* Analysis of delays due to weather events (rain, snow, wind, etc.).
* Impact of airport congestion (e.g., limited gates or taxiways).
* Frequency and timing of runway closures or maintenance periods.
* Correlation between flight volume and delay density across time.
* Seasonal trends in delays that may necessitate temporary expansions or mitigation plans.

#### **Policy Makers and Regulators**

At the national or state level, regulators aim to monitor system-wide trends and make long-term investments in infrastructure or legislation. Their needs may include:

* Aggregated delay statistics across states, regions, or airlines.
* Trend analysis over months or seasons to assess the impact of policy changes or infrastructure upgrades.
* Delay categorization by responsibility source (e.g., carrier, NAS, weather) to allocate accountability.
* Benchmarks across airports to identify underperforming hubs.
* Recommendations for federal funding allocation based on delay patterns.

#### **Passengers and Customer Service Teams**

Passengers and frontline service teams are directly affected by delays and must manage expectations and experience. Key information required may include:

* Expected delay durations for specific routes or airlines.
* Probability of delays during certain seasons or at airports.
* Real-time updates and historical reliability metrics by carrier.
* Factors influencing missed connections, especially on multi-leg journeys.
* Communication tools to convey reasons for delays and estimated resolution.

## **3. Data Handling and Preparation**

### **3.1 Data Sources Used**

* US\_flights\_2023.csv – This is the main dataset containing detailed records for domestic U.S. flights in 2023. It includes variables such as flight dates, airline codes, aircraft tail numbers, departure and arrival airport codes and cities, delay types, and aircraft specifications.
* weather\_meteo\_by\_airport.csv – This file provides essential weather variables (temperature, precipitation, snow, wind, and pressure) on a daily basis for major U.S. airports. It adds contextual insight into how environmental conditions may affect flight operations.
* airports\_geolocation.csv – Used to map airport IATA codes to their corresponding cities and geographical coordinates. This mapping is critical for spatial analysis and integrating external location-based datasets
* cancelled\_Diverted\_2023.csv – A dataset capturing canceled and diverted flights for the same period. It was used primarily as a data quality filter to ensure that such records were not included in the main analysis pipeline.
* maj us flight - january 2024.csv – This serves as a validation subset. It contains data from January 2024, used to test model performance and generalize patterns beyond the training dataset.

### **3.2 Data Cleaning Process**

* Tools use: The primary tool for data preparation was Python (Pandas, NumPy), with supplementary checks and previews done in Power BI during the modeling stage.
* Summary of missing data: where, how much, patterns: Initial exploration revealed missing values primarily in weather-related variables such as prcp, snow, and wspd. Some aircraft metadata fields (e.g., manufacturer or model) also contained nulls, but these were deemed non-critical for the delay-focused analysis. For the weather dataset, the missing rate ranged from 3–8% per variable depending on the airport and month. In the flights dataset, certain delay breakdown fields (Delay\_Carrier, Delay\_Weather, etc.) were missing for on-time flights, which is logical given their conditional nature.
* Summary of data quality issues:
  + **Outliers**: Some records showed extreme values, such as arrival delays greater than 24 hours, which are highly improbable and likely data entry errors.
  + **Misclassifications**: Some values in the DepTime\_label field (e.g., "Evening") were categorical and non-numeric, requiring transformation for time-based analysis.
* How issues were addressed:
  + **Remove**: All rows with arrival delays above 1440 minutes (24 hours) were removed to maintain realistic bounds.
  + **Fix**: Missing weather data were selectively imputed using forward-fill and airport-based grouping to preserve temporal coherence. Weather fields like wspd and prcp received minimal imputation to support correlation analysis.
  + **Use with Caution**: Incomplete records involving aircraft metadata were retained but flagged with documentation, as they did not impact core delay modeling.

### **3.3 Data Integration**

* The airports\_geolocation.csv file was merged using the Dep\_Airport and Arr\_Airport codes. This allowed for a location-based analysis and supported airport clustering.
* Weather data was joined with flights on FlightDate and airport code (Dep\_Airport → airport\_id), ensuring that each flight record could be associated with environmental conditions at the point of departure.
* Delay fields (Dep\_Delay, Arr\_Delay) were filtered and transformed to create meaningful derived fields. Flights with arrival delays greater than 15 minutes were flagged as “Delayed,” while others were labeled “On-Time.” This threshold aligns with FAA’s standard delay definition.

### **3.4 Feature Engineering**

* Several new features were created to support downstream visualizations and forecasting models:
  + Delay Status Flag: A binary label (Delayed vs. On-Time) was engineered using Arr\_Delay. This flag forms the basis of correlation analysis and categorical breakdowns.
  + Delay Reason Grouping: Delay types such as carrier delay, NAS delay, weather delay, and security delay were consolidated into grouped tags to allow stacked and aggregated analyses.
  + Time-Based Variables: Month was extracted using pd.to\_datetime(FlightDate).dt.month\_name() to support seasonal trend visualizations.
  + Day\_Of\_Week was preserved and formatted to analyse weekday/weekend patterns.
  + DepTime\_label (e.g., Morning, Afternoon, Night) was used to create an Hour variable for temporal heatmaps after transforming the categorical values into numeric representations (e.g., Morning = 8, Afternoon = 14, Night = 21).

## **4. DSS System Description**

### **4.1 System Overview**

* Platform: Power BI
* User Interface: Multiple dashboards with filters and tooltips
* Navigation: Clear tabs for different perspectives (airline, airport, time, weather)

### **4.2 Dashboard Pages**

* **Overview:** Key metrics and charts (total delayed flights, avg delay time, etc.)

The home page serves as the primary interface, offering a consolidated view of airline operational health through core metrics and visualizations. It reports 6.7 million total flights with an average duration of 140 minutes, reflecting the scale and typical route lengths of operations. The 15 airlines analyzed operate fleets averaging 13 years in age, highlighting potential maintenance demands linked to aging aircraft. Disruptions are quantified through 88,000 cancelled flights and 17,000 diversions, while consistent 15-minute average arrival and departure delays underscore systemic inefficiencies in scheduling or turnaround processes.

Delay Causation prioritizes internal operational challenges, with LastAircraft issues leading at 40% of delays, closely followed by Carrier inefficiencies at 36%. NAS congestion accounts for 18%, while Weather and Security delays contribute smaller shares at 5% and 0.2%, respectively. This hierarchy directs stakeholders to focus on maintenance and process improvements over external factors.

The Delay Flight Map visualizes geographic delay patterns, correlating causes with regions, for example, frequent NAS delays near major hubs like Atlanta or weather-related disruptions in storm-prone areas such as the Midwest. Interactive filters for airlines, dates, or aircraft models enable users to isolate trends, transforming broad data into actionable insights.

As the gateway to deeper analysis, the home page anchors strategic decision-making by spotlighting priority areas. Top navigation tabs such as Seasonal Analysis or Airline Performance allow seamless transitions to detailed investigations, ensuring a cohesive workflow from trends to targeted solutions.

* **Delay Cause:** Key metrics and charts (total delayed flights, avg delay time, etc.)

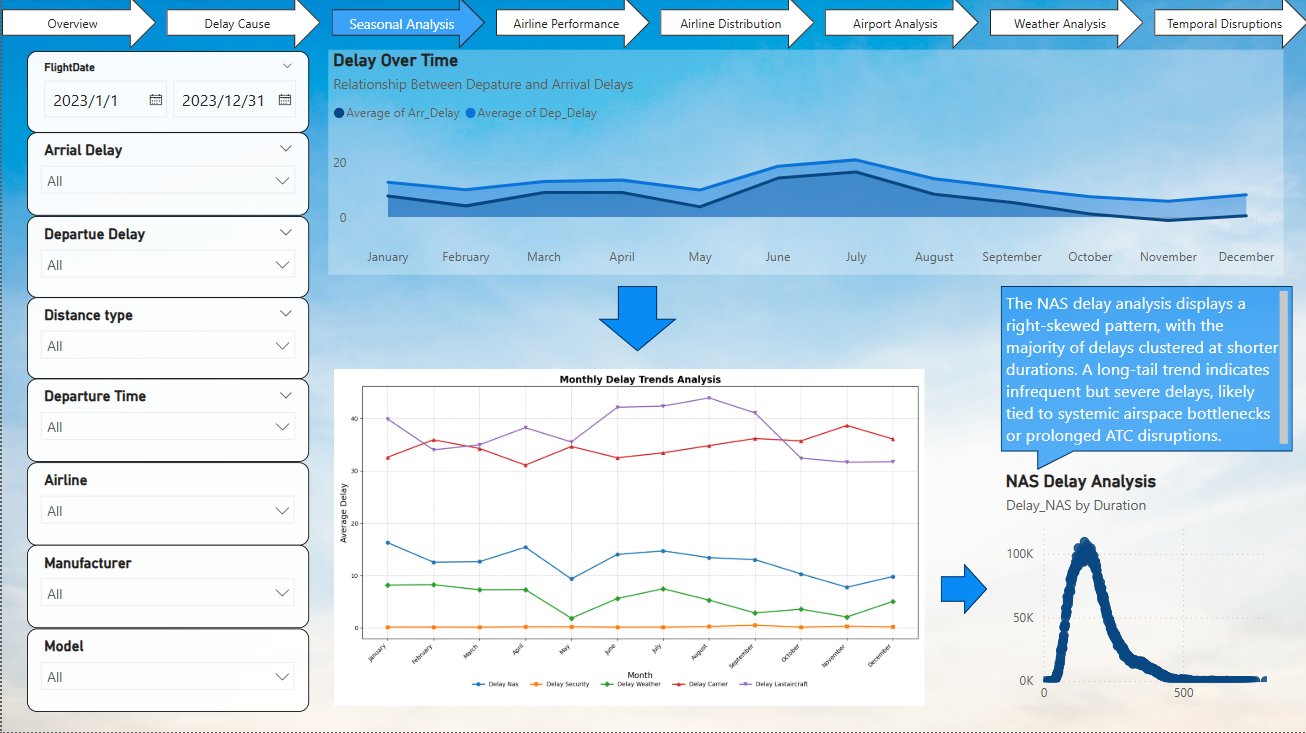
The correlation heatmap illustrates the relationships between the five primary causes of flight delays—Carrier, Air System, Security, Weather, and Aircraft Issues. While these factors initially appear largely independent, the integration of interactive filters enables users to uncover context-specific interactions. For example, selecting a particular airline or timeframe may reveal that Weather Delays, though generally isolated, correlate more strongly with Aircraft Issues during peak travel seasons, suggesting adverse weather exacerbates maintenance challenges. Similarly, narrowing the analysis to older aircraft models could show heightened dependencies between Air System Delays and Carrier-related inefficiencies, indicating systemic bottlenecks in high-traffic scenarios. By dynamically adjusting parameters such as FlightDate, Aircraft Age, and Flight Duration, stakeholders can isolate critical patterns, transforming broad independence into actionable insights for targeted operational improvements.

* **Seasonal Analysis:** Key metrics and charts (total delayed flights, avg delay time, etc.)

The Relationship Between Departure and Arrival Delays graph underscores a critical operational challenge: departure delays frequently cascade into arrival delays. This correlation highlights the importance of prioritizing on-time departures, optimizing turnaround times, and incorporating buffer periods to minimize downstream disruptions.

The Monthly Delay Trends Analysis chart reveals seasonal patterns, with delays peaking during summer from June to August. These spikes likely reflect weather disruptions and heightened travel demand. Managers can use this to preemptively allocate staff, adjust maintenance schedules, or secure backup resources during high-risk periods.

The NAS Delay Analysis chart depicts a right-skewed distribution, indicating that most National Air System delays are under 30 minutes but occasionally escalate into severe, multi-hour bottlenecks. Managers can implement tiered responses such as routine protocols for common delays and contingency plans for prolonged disruptions to maintain operational fluidity.



* **Airline Performance:** Average delay stats by airline and route

The Departure Delay Analysis highlights persistent challenges faced by Southwest Airlines, where the last aircraft-related delays accounted for 40.03 percent of its total disruptions.

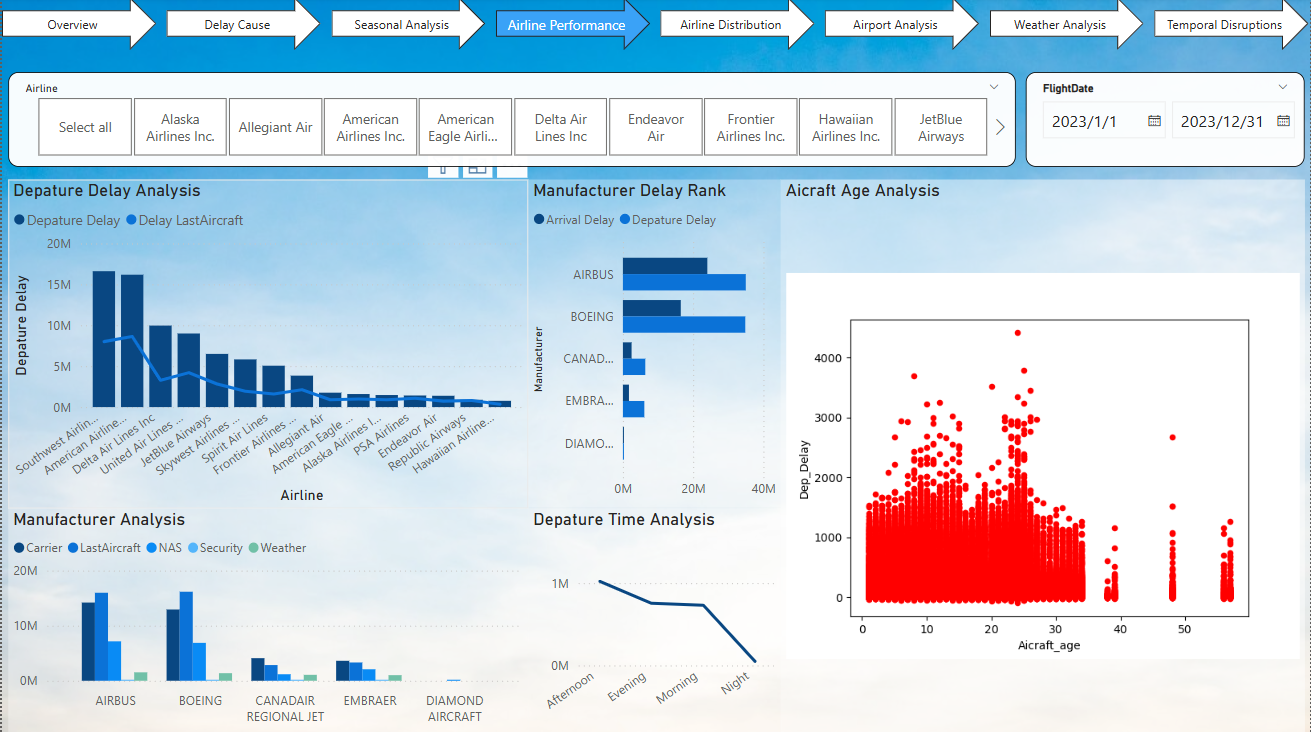
Airline Performance Metrics highlight disparities among carriers, with major airlines such as American Airlines Inc. and Delta Air Lines Inc. experiencing delays exceeding 800 million instances.

The Manufacturer Delay Rank ties delays to specific aircraft manufacturers, revealing that Boeing and Airbus models contribute significantly to disruptions, with 40 million and 20 million delay instances, respectively. Such trends may stem from fleet age or maintenance complexity, guiding decisions to prioritize upgrades or streamline maintenance protocols for high-risk models.

These insights enable managers to tailor strategies for Southwest Airlines, such as preallocating maintenance resources or adjusting schedules during peak periods, while also benchmarking against industry trends to address systemic challenges across larger fleets.

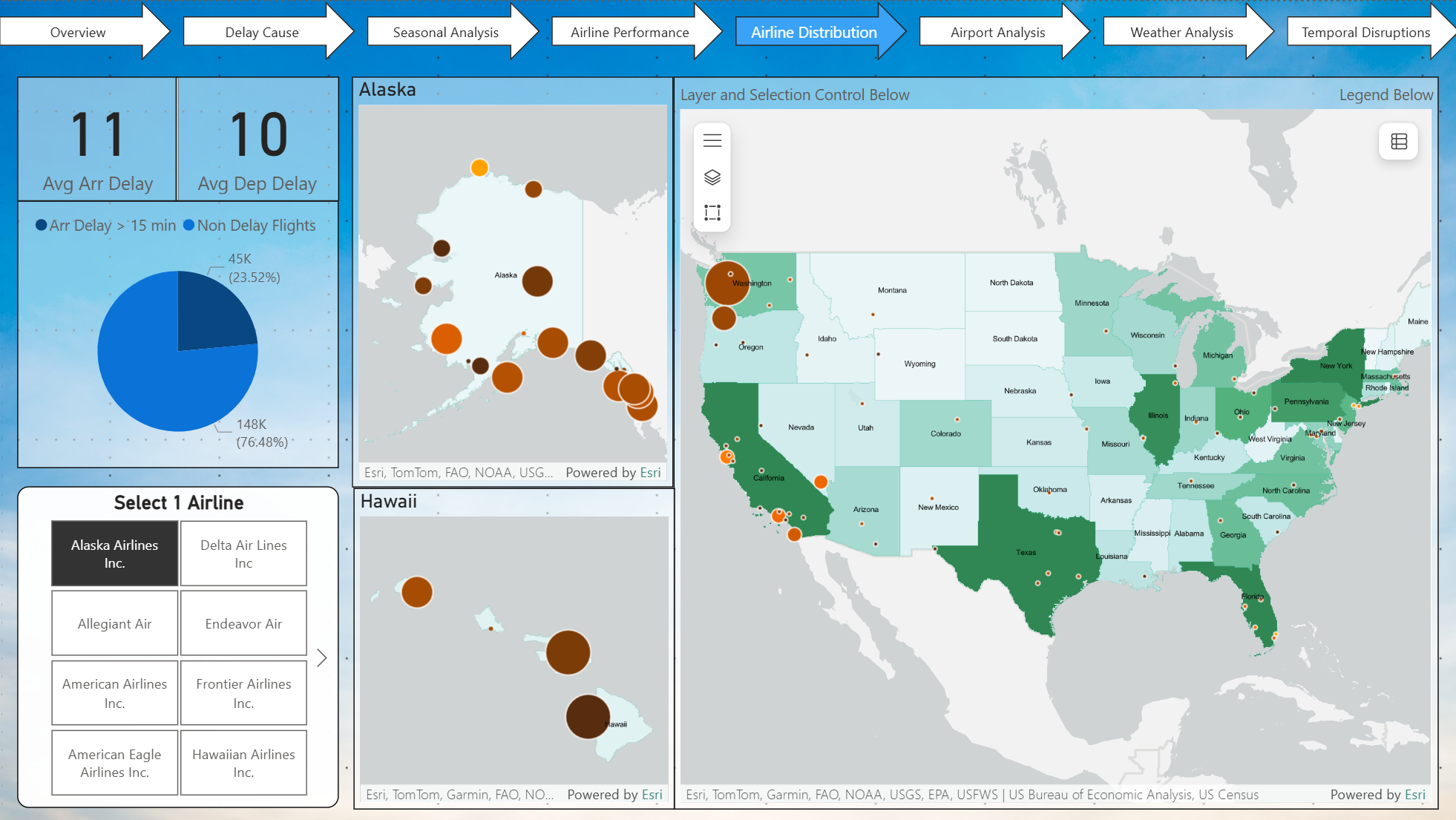
Departure Time Analysis identifies congestion peaks during mid-morning hours, particularly between 10:00 and 12:00, aligning with high passenger volumes. Managers can mitigate these bottlenecks by optimizing crew shifts, gate availability, or integrating buffer times into schedules.

By synthesizing these insights, the DSS empowers Operations Managers to address root causes, enhance fleet reliability, and align operations with demand fluctuations, fostering a more resilient and efficient aviation ecosystem.



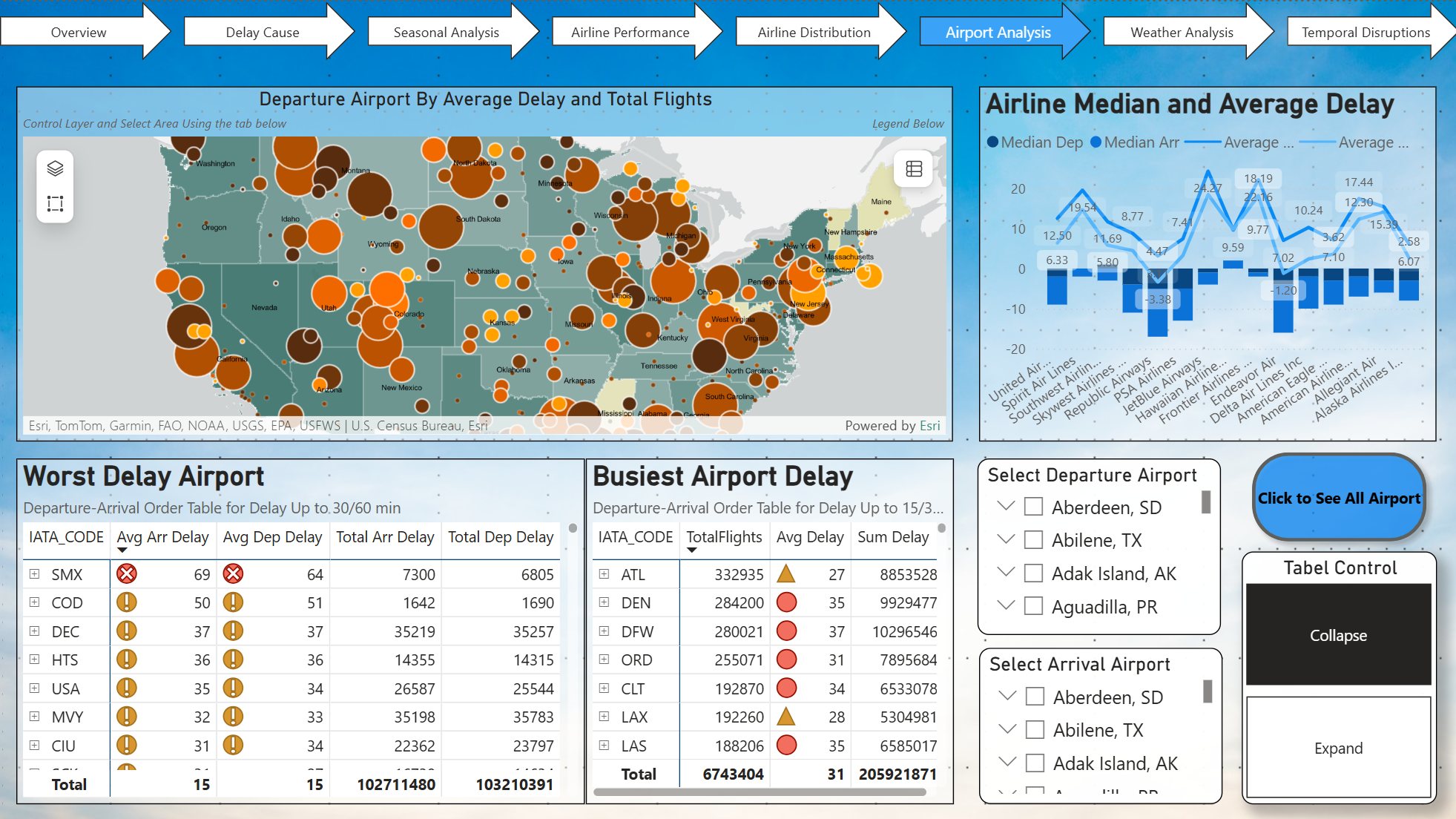
* **Airline Distribution:** Arrival airport distribution in power bi arcgis with delay and flight information, in addition to state GDP visualization(See appendices).

On the Airline Distribution page, the geographical visualization of airline flight distribution offers users a broader spatial understanding of carrier operations. Overlaying this with state GDP data, displayed using a color ramp, allows for exploratory analysis of potential correlations between airline coverage and regional economic activity. This feature supports strategic planning by offering insights into market presence and demand potential across states.



* **Airport Analysis:** Interactive map with additional population data (See appendices), bar chart and matrix table by delay and airport.

Within the Airport Analysis page, tracking average delay times serves as a key indicator of systemic inefficiencies at the airport level. The analysis highlights several airports with notably high average delays, pointing to potential operational bottlenecks. Santa Maria (SMX), for instance, records the highest average arrival and departure delays at 69 and 64 minutes respectively—significantly above national averages. This is followed by Cody (COD) and Decatur (DEC), each exhibiting average delays exceeding 35 minutes. Such figures suggest persistent delay issues that are disproportionate to overall flight volume and may stem from factors such as limited staffing, runway constraints, or suboptimal scheduling. In contrast, major hub airports like Dallas-Fort Worth (DFW), Denver (DEN), and Charlotte (CLT), while handling a substantially larger volume of traffic, maintain average delays within a moderate range of 27 to 37 minutes. These are still operationally significant due to the scale of activity. Additionally, the matrix can be expanded to reveal route-specific delays, with visual icons indicating delays exceeding 15, 30, or 60 minutes, aiding in quick identification of high-risk connections.



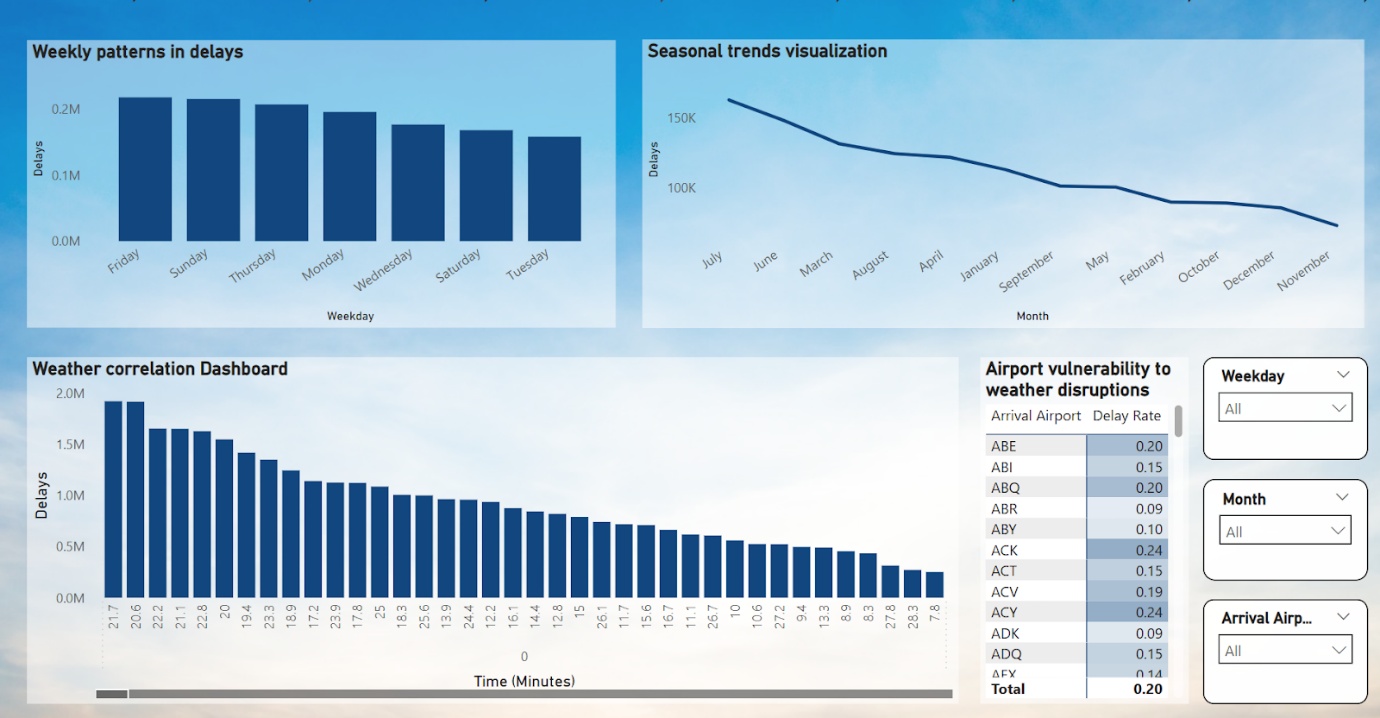
* **Weather Analysis:** Scatterplots showing delay vs. weather (e.g., precipitation, visibility)

The Weather Analysis dashboard provides a focused look into how varying weather conditions influence flight delays across different U.S. airports. It includes interactive filters for Departure City, Arrival City, Departure Airport, and Arrival Airport, allowing users to tailor insights to specific routes or hubs. One of the key visuals is a line chart that maps monthly trends in weather variables such as precipitation, snowfall, and temperature shifts, helping identify peak disruption periods throughout the year. Each line represents a weather factor, enabling direct comparison across months. Additionally, the Airport Vulnerability Table ranks airports based on their sensitivity to weather-induced delays, offering operational teams a clear sense of which locations may require proactive mitigation strategies during adverse weather conditions.



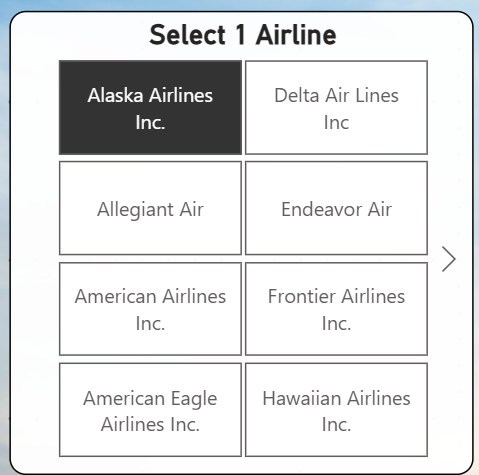
* **Temporal Disruptions:** Seasonal trends of delay frequency

The Temporal Disruptions dashboard digs deeper into the when of delays, revealing patterns that emerge across different timescales. Equipped with filters for Weekday, Month, and Arrival Airport, it enables users to dynamically explore trends in temporal disruptions. Key visuals include: A Weekly Delay Pattern heatmap highlighting which days are most prone to delays. A Weather Correlation Dashboard that visually ties flight delay rates to weather conditions over time. A Seasonal Trends Chart showing delay fluctuations across months, helping detect seasonal bottlenecks. And once again, an Airport Vulnerability Matrix, reinforcing insights on location-based weather sensitivity. Together, this dashboard paints a clear picture of how time and weather interact to influence flight disruptions, aiding both strategic planning and operational response.

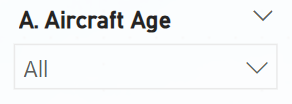


### **4.3 User Controls**

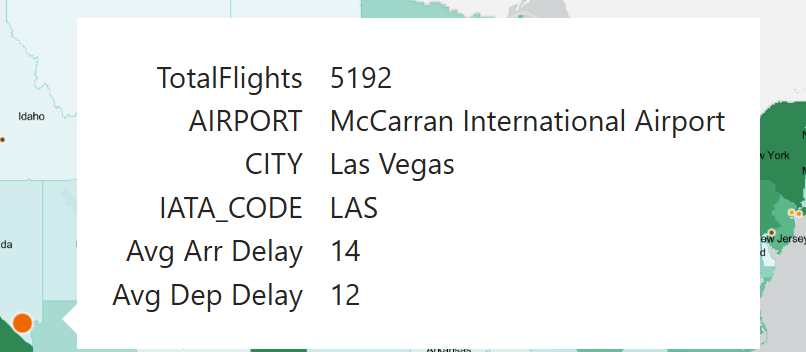
* Filters: Airport, airline, delay type, month/season, day of week, hour of day



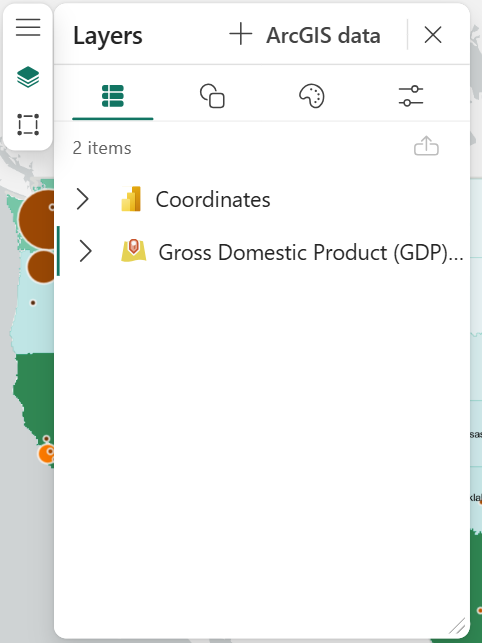
* Dropdowns and slicers to adjust time range, view metrics by group



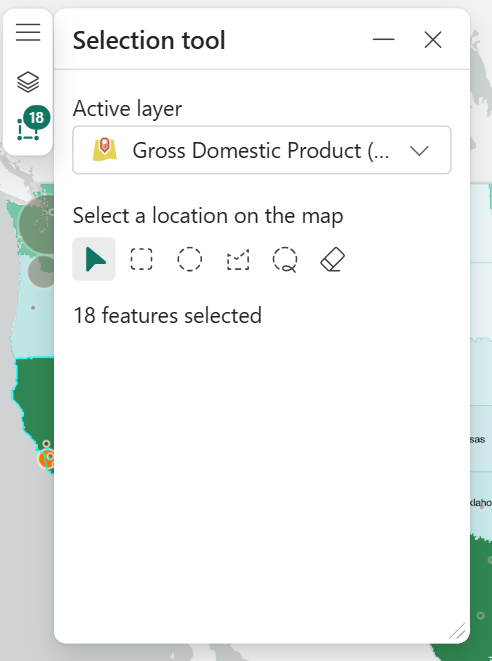
* Hover tooltips and on-screen help text for ease of navigation



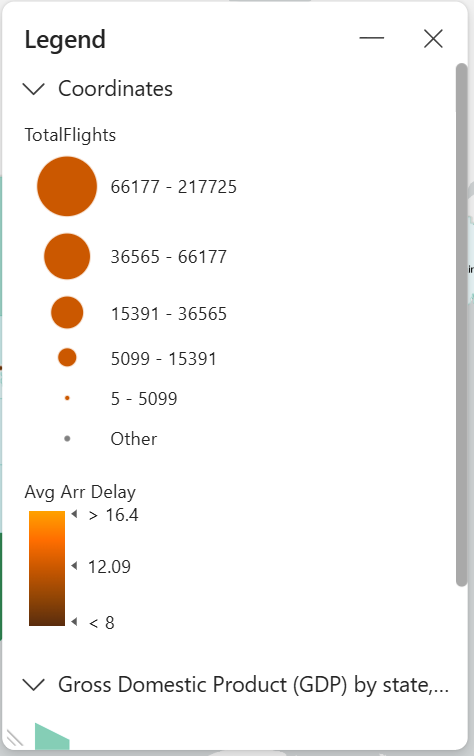
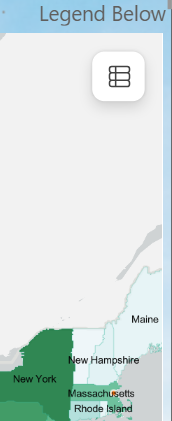
* Map Layer Control: edit the map layer data and visualization



* Map Selection Control: select an area or point in map



* Map Legend: see map legend information



### **5. Key Insights**

LastAircraft maintenance challenges and Carrier operational inefficiencies dominate flight delays, representing 40% and 36% of disruptions respectively. Southwest Airlines faces heightened risks from aging fleets, while major carriers like American Airlines and Delta Air Lines experience high delay volumes due to operational scale. Airport Santa Maria Public Airport (Capt G. Allan Hancock Field) consistently has longest average delays.

Seasonal trends reveal a 20% surge in delays during summer months, driven by weather disruptions and travel demand, particularly in regions like the Midwest. Mid-morning hours between 10:00 and 12:00 consistently show congestion-driven delays, linked to overlapping flight schedules and passenger volume peaks. Aircraft models from Boeing and Airbus disproportionately contribute to delays, emphasizing fleet modernization needs.

## **6. Appendices**

* Python/Prep scripts used for cleaning
  + Data Cleaning

A screenshot of a computer program

AI-generated content may be incorrect.

* + Dataset Merging

A screen shot of a computer

AI-generated content may be incorrect.

* References to public datasets
  + <https://services.arcgis.com/jIL9msH9OI208GCb/arcgis/rest/services/GDP_by_State/FeatureServer>
  + https://services.arcgis.com/P3ePLMYs2RVChkJx/arcgis/rest/services/USA\_Census\_2020\_DHC\_Total\_Population/FeatureServer

1. Due to the file size limit to upload to Power BI Galleries, we have grant access to this email [peter.keenan@ucd.ie](mailto:peter.keenan@ucd.ie) [↑](#footnote-ref-1)