Project: Data Visualization

By Layla Alsaabna

Introduction:

Getting to our destination with the minimum delay is one of the most important goals when it comes to traveling.

In this project I will create visualizations to reveal insights from the flight delays data set. I will create data visualizations that tell a story and highlight patterns in the flight delays data set.

About Dataset:

Dataset can be found here.

The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics tracks the on-time performance of domestic flights operated by large air carriers. Summary information on the number of on-time, delayed, canceled, and diverted flights is published in DOT's monthly Air Travel Consumer Report and in this dataset of 2015 flight delays and cancellations.

The flight delay and cancellation data was collected and published by the DOT's Bureau of Transportation Statistics.

The dataset has 3 files, flights.csv, airlines.csv, and airports.csv.

Flights data:

Column Name	Column Type	Description
Year	Number	The year of the flight (e.g., 2023)
Month	Number	The month of the flight (1–12)
Day	Number	The day of the flight (1–31)
Day_Of_Week	Number	Day of the week the flight occurred (1 = Monday, 7 = Sunday)
AIRLINE	String	The name or code of the airline operating the flight (e.g., "Delta", "DL")
FLIGHT_NUMBER	Number	The flight number assigned by the airline (e.g., 123)
TAIL_NUMBER	String	The unique registration number of the aircraft (e.g., N12345).
ORIGIN_AIRPORT	String	The IATA code of the departure airport (e.g., JFK, LAX)
DESTINATION_AIRPORT	Number	The IATA code of the arrival airport (e.g., SFO, ATL)

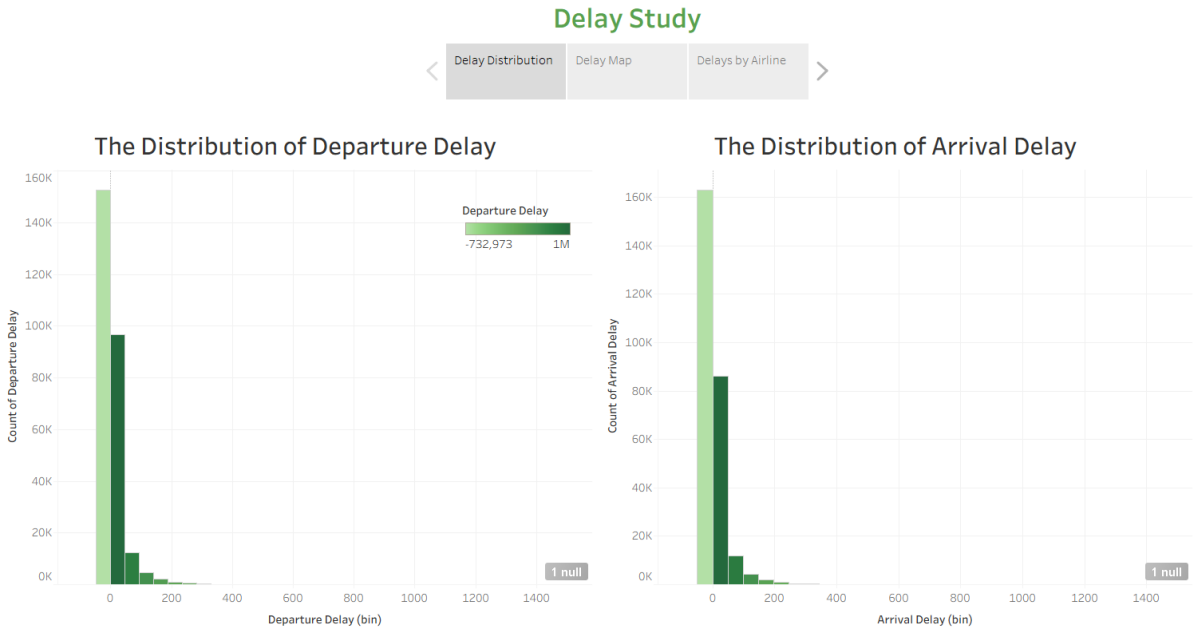
SCHEDULED_DEPARTURE	String	Scheduled local departure time in HHMM format (e.g., 1330 = 1:30 PM)
DEPARTURE_TIME	Number	Actual local departure time in HHMM format
DEPARTURE_DELAY	Number	Delay in departure in minutes (positive = late, negative = early)
TAXI_OUT	Number	Time in minutes between departure from the gate and takeoff (wheels off)
WHEELS_OFF	Number	Local time when the aircraft leaves the ground (takeoff), in HHMM format
SCHEDULED_TIME	Number	Scheduled duration of the flight in minutes (as planned)
ELAPSED_TIME	Number	Actual total time from departure to arrival (including taxi, flight, delays) in minutes
AIR_TIME	Number	Time from wheels off to wheels on (actual airborne time) in minutes
DISTANCE	Number	Distance between origin and destination airports in miles
WHEELS_ON	Number	Local time when the aircraft lands (touchdown), in HHMM format
TAXI_IN	Number	Time in minutes between landing (wheels on) and arriving at the gate
SCHEDULED_ARRIVAL	Number	Scheduled local arrival time in HHMM format
ARRIVAL_TIME	Number	Actual local arrival time at the destination gate, in HHMM format
ARRIVAL_DELAY	Number	Delay in arrival in minutes (positive = late, negative = early)
DIVERTED	Number	Binary indicator: 1 if the flight was diverted, 0 if not
CANCELLED	Number	Binary indicator: 1 if the flight was cancelled, 0 if not
CANCELLATION_REASON	String	Reason for cancellation (e.g., A = Airline, B = Weather, C = National Air System, D = Security)
AIR_SYSTEM_DELAY	Number	Delay caused by air traffic control, or other air system issues (in minutes)
SECURITY_DELAY	Number	Delay due to security-related issues at the airport (in minutes)
AIRLINE_DELAY	Number	Delay attributable to the airline (e.g., maintenance, crew) (in minutes)
LATE_AIRCRAFT_DELAY	Number	Delay due to the previous flight of the same aircraft being late (in minutes)
WEATHER_DELAY	Number	Delay caused by weather conditions (in minutes)

Question 1:

Which airlines or airports have the worst delays?

Link: Story.

Before answering this question, let's look at the distribution of the delays, the first page of our story.



Summary

The chart above revealed that most flights actually depart early (negative delay minutes), followed by flights that depart exactly on time (0 minutes). Positive delays, while less frequent, are the ones that impact passengers the most. Therefore, the following visualizations focus on positive delays only.

Design

The Histogram chart type was chosen because it clearly shows the frequency and spread of delay times across a large dataset, and it allows for easy comparison between departure and arrival delays when plotted together, and Gradient Color Scheme Was chosen because it reflects a natural progression from early → on time → delayed.

Answering the question-Which airports have the worst delays? - may save you some time when choosing an airport (if applicable), this question can lead to another question: What hours have the worst delays and should be avoided? If you choose the right time to travel, it would be amazing!

To ensure that early departures (negative delay values) did not reduce the total delay minutes, I created two calculated fields DepDelPos, and ArrDelPos. These fields only count positive values, effectively isolating real delays from early departures. This provided a clearer picture of delay burdens across states, cities, and airlines

Departure Delay Map:

The link to the visual can be found here, and it's the second page in the story.

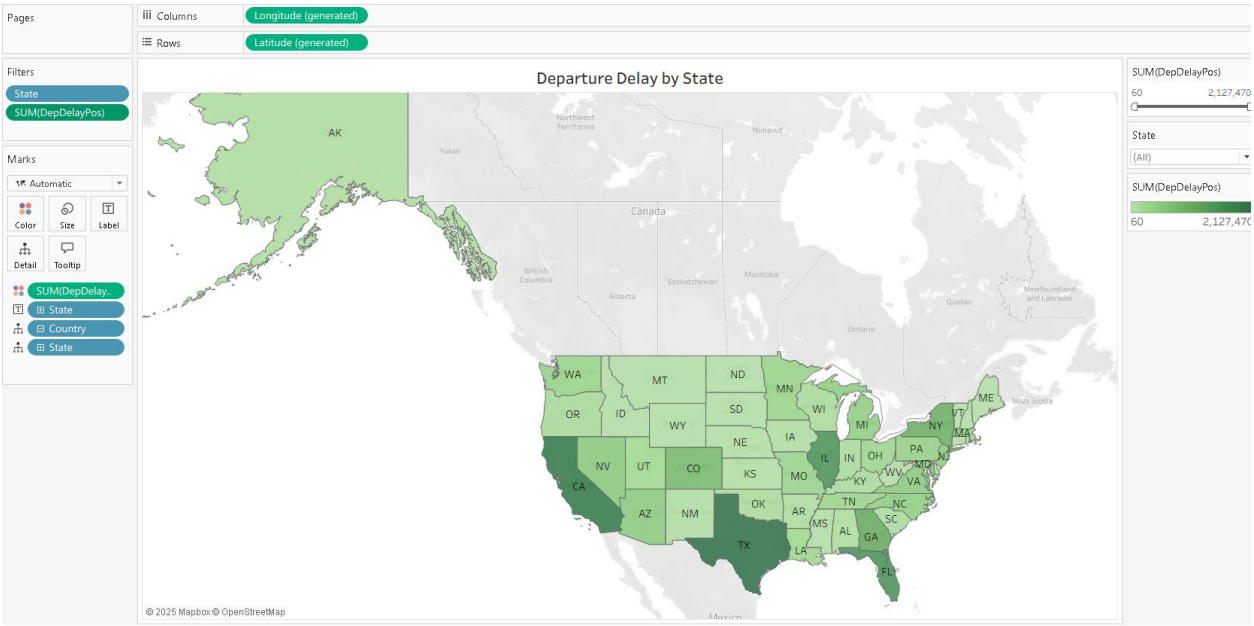
Summary:

Departure delays are critical because they often lead directly to arrival delays. To simplify the analysis, this visualization focuses only on departure delays, aggregated at the state and city level. The map highlights which regions suffer the most delays, with Texas and California standing out as the worst. Using the tooltip, we can also explore airline-level contributions to the delays. For example, Southwest Airlines in California and American Airlines in Texas are major contributors to delay minutes.

Design:

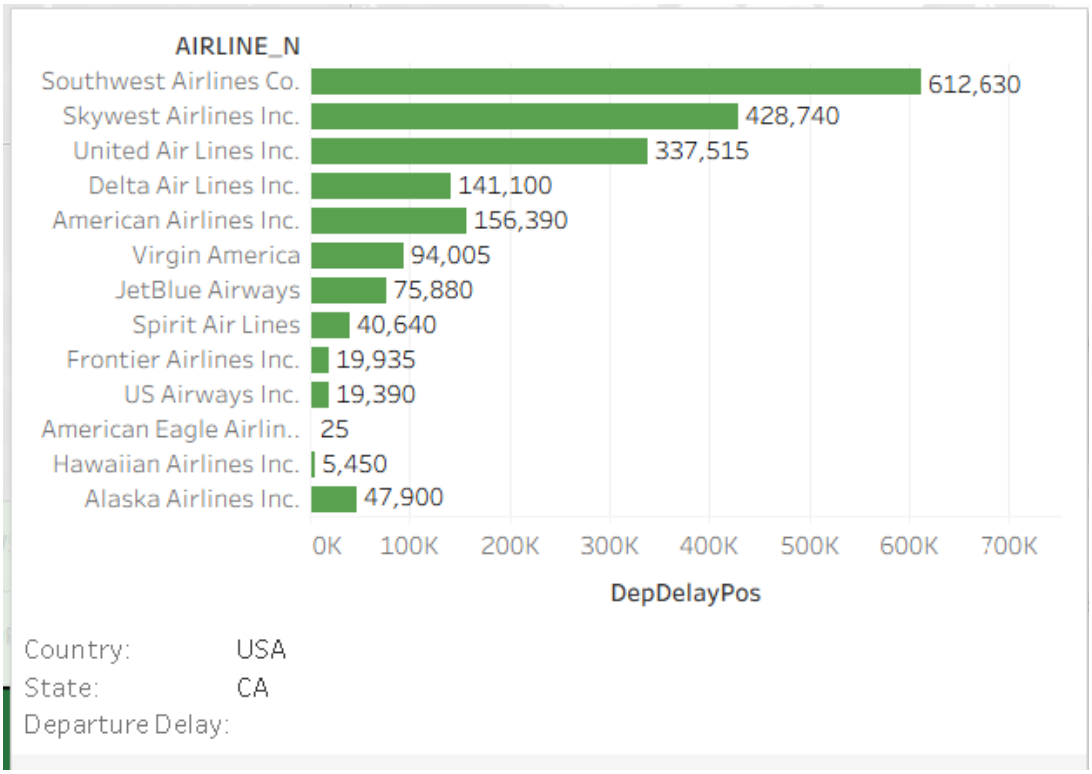
I chose a filled map to represent departure delays because the geographic spread of delays is easier to interpret visually when linked to states and cities. Each state is labeled for clarity, and I used tooltips to show airline-specific delays within that region, displayed in a bar chart format for quick comparison. Filters were added for state, city, and airline, so users can narrow down the data interactively. This design makes it simple to start broad at the geographic level and then drill down into specific airlines or locations contributing to delays.

Analysis



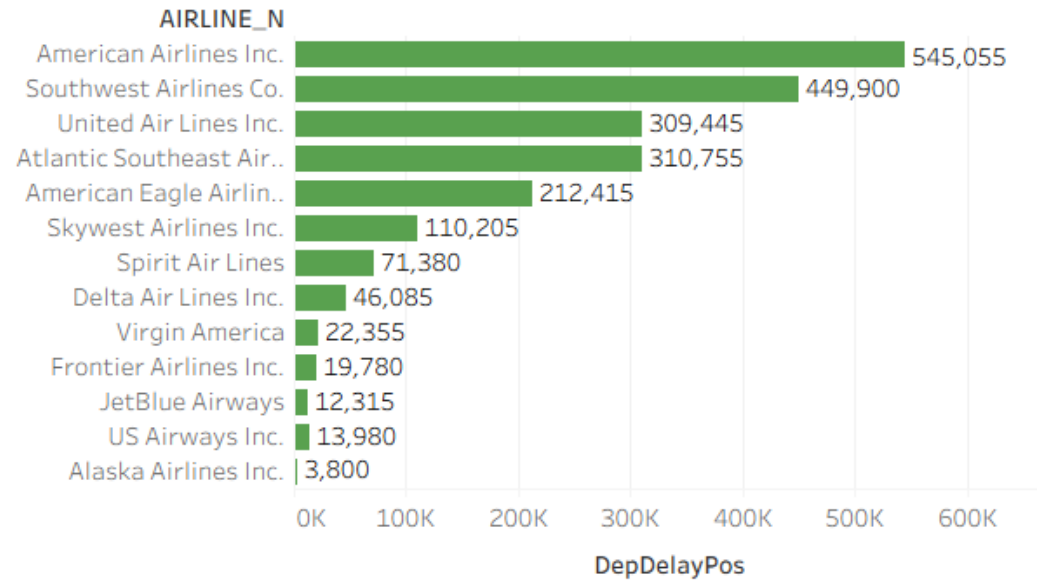
As we can see from the picture above the worst states are TX (Texas), and CA (California) are the worst states for having the longest departure delays, note that since a departure delay will mostly cause an arrival delay then for this visualization only the departure delay was studied.

If we hovered over California, we see the following airlines.



From the figure above Southwest Airline has the longest departure delay with over 600K delay minutes.

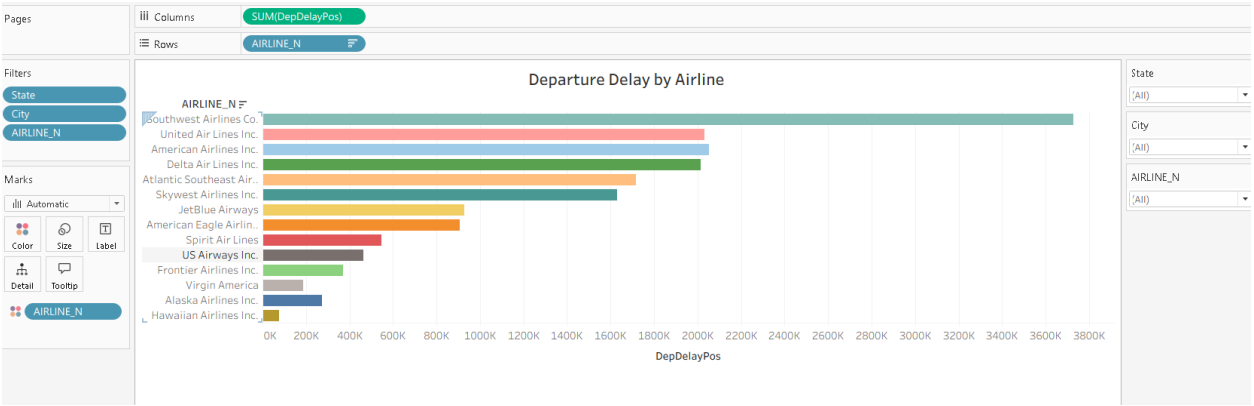
Texas state



For Texas state the worst airline is The American Airline with over 500K minutes of departure delay, so which airline is the worst? does this chart tells the whole story?. To answer these questions let's look at another visuals, starting with the bar chart of departure delay by airline, so we took the sum of this airline departure delay meaning it won't be spread across states, so we will look at the departure delay through airlines only.

Departure Delay by Airline:

The link to the visual is here, and it's the third page of the story.



Summary:

While the state-level map highlighted that California and Texas each have their own worst-performing airlines, this does not capture the full story. When airlines are considered across all states, Southwest Airlines emerges as the clear leader in departure delays, accumulating

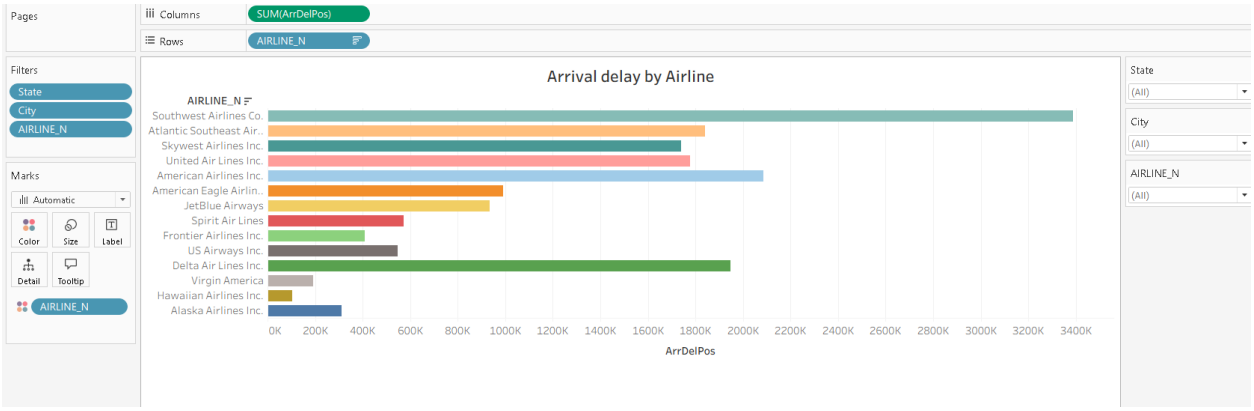
almost 3800K minutes. This broader perspective shows that Southwest consistently underperforms regardless of region.

Design:

I chose a horizontal bar chart to display total departure delays by airline because it allows easy comparison across categories. Each airline is assigned a distinct color to visually separate them. Filters for state, city, and airline were added so viewers can drill down into more specific contexts if needed. This design makes it simple to see both the overall trend (Southwest as the worst) and explore localized variations.

Arrival Delay by Airline:

The link to the visual can be found here, and it's also on the third page of the story.



Summary:

Just as with departure delays, Southwest Airlines also shows the highest total arrival delays, with almost 3400K minutes. This reinforces the conclusion that Southwest struggles significantly with both ends of flight schedules, impacting travelers the most across the dataset.

Design:

The design mirrors the departure delay bar chart to maintain consistency. A horizontal bar chart was used for direct airline-to-airline comparison, with distinct colors assigned to each airline. Filters for state, city, and airline were also applied to allow interactive exploration. Keeping the design consistent across the two charts makes it easy to compare departure and arrival delays side by side.

Summary for dashboard:

If you want to know the worst airline in delay we look at the bar chart to know the overall delay for each airline, if you however want to know the worst airline in each state to avoid it for certain states then we use the map.

Question 2:

What causes delays?

Link: Visualization.

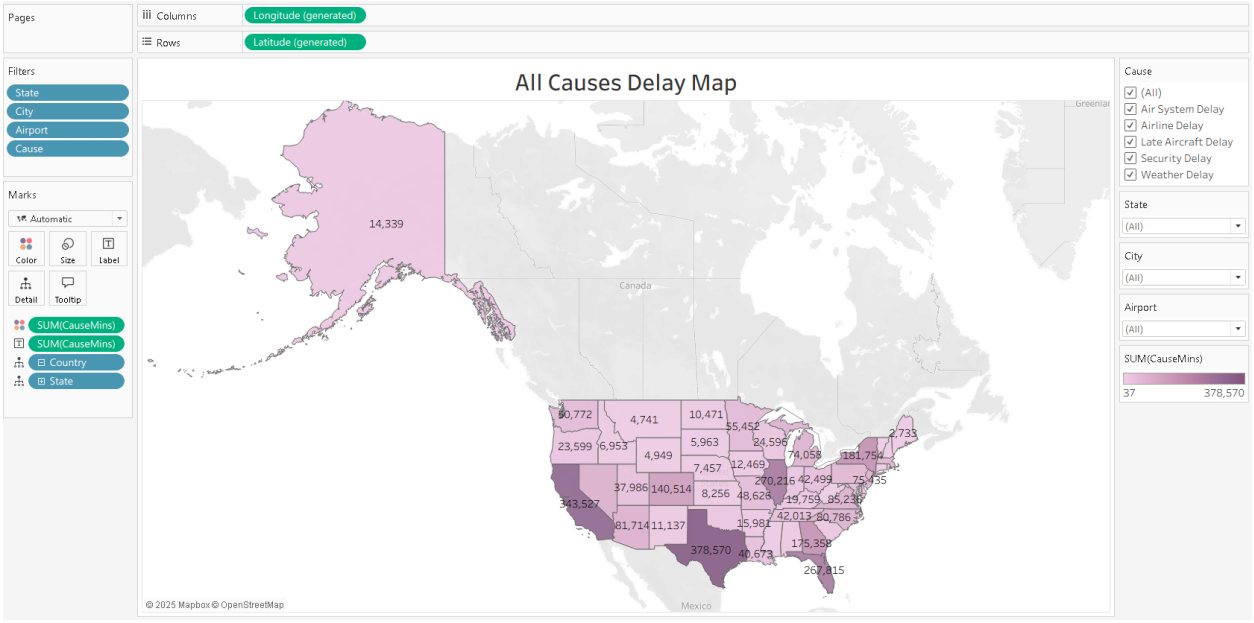
Summary

To explore what causes flight delays, I pivoted the dataset so each cause type (Weather, Airline, Air System, Security, and Late Aircraft) became its own category. This allowed me to map delay causes across states. The visualization shows that California and Texas again stand out with the highest overall delay minutes, but now we can see how much each cause contributes in different regions. For example, weather delays cluster heavily in certain states, while late aircraft and carrier delays dominate in others. This confirms that causes are more tied to airports and geography than to airlines alone.

Design

I first reshaped the dataset by pivoting the cause columns (Weather Delay, Airline Delay, Air System Delay, Security Delay, Late Aircraft Delay) into two fields: Cause (the category) and CauseMins (the measure). This long format made it possible to slice the data by cause.

Delay Causes Map



Summary

This map visualization highlights the geographic distribution of flight delay causes across U.S. states. Filters for *Cause, State, City, and Airport* allow for flexible exploration. When viewing all causes combined, Texas (378,570 minutes) and California (343,527 minutes) clearly emerge as the top two states for total delay time. However, breaking down by individual causes reveals more nuanced patterns:

- **Air System:** Texas (73,846) and California (65,473) dominate, followed by Florida (63,999) and Illinois (59,383).
- **Airline:** Texas (126,170) and California (101,915) remain the top two states.
- **Late Aircraft:** California (169,085) and Texas (150,353) are again the highest contributors.
- **Security:** Florida (794), California (793), and Texas (443) emerge as the top three.
- **Weather:** Illinois (28,766) and Texas (27,758) stand out here, with California not in the top two.

Design

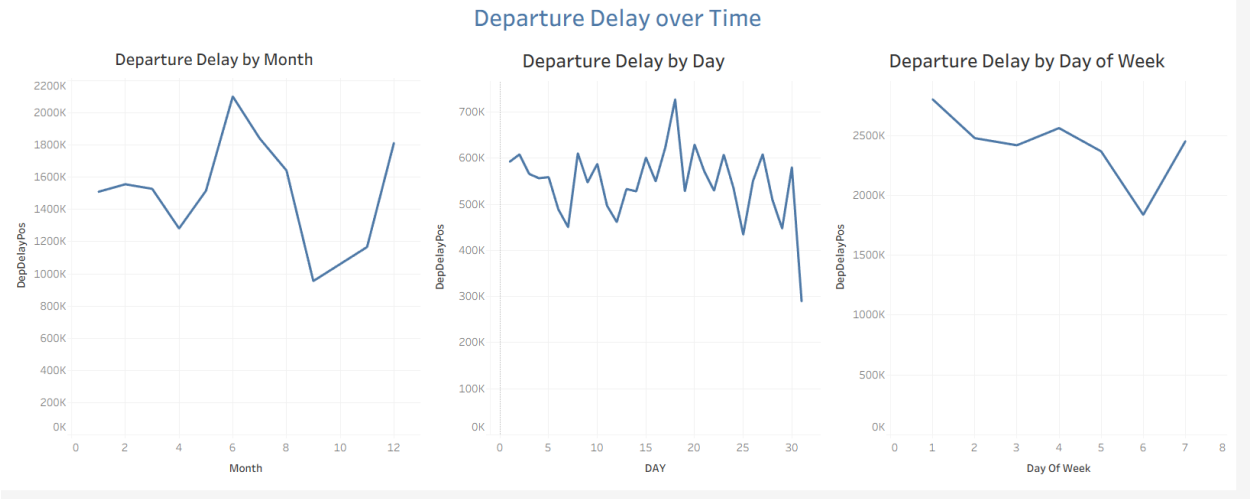
- **Color:** States are shaded by **SUM(CauseMins)**, allowing quick visual comparison of intensity.
- **Label:** States display **SUM(CauseMins)** directly on the map for exact values.
- **Filters:**
  - **Cause** → lets users isolate or compare individual causes.
  - **State, City, Airport** → provide drill-down from nationwide to local detail.
- **Interactivity:** Switching between cause types dynamically changes the distribution, surfacing different “hot spots” of delays depending on the chosen filter.

- **Tooltips:** Showed both the total delay minutes and the breakdown by cause for the hovered state, giving detail beyond the color intensity.

Question 3

Which times of day or days of the week have the worst departure or arrival delays?

Link: Dashboard.



Summary

The results showed strong temporal patterns in departure delays. Month 6 (June) recorded the highest total positive departure delays, with approximately 2 million minutes, suggesting peak summer travel significantly increases delays. On the daily level, the 18th of the month stood out as the worst day, possibly reflecting mid-month travel surges. At the weekly level, Day 1 (Monday) experienced the highest delays, which may indicate heavy business travel demand at the start of the week. Overall, these findings suggest that both seasonal peaks (summer), monthly cycles, and weekly travel rhythms contribute to increased departure delays.

Design

To analyze when delays are most severe, I used the calculated DepDelPos to isolate only actual delays (ignoring early departures). I then plotted the total DepDelPos as a line chart against three different time dimensions: Month, Day of Month, and Day of Week. This design allowed me to detect recurring temporal patterns and identify specific time periods associated with the worst departure delays.

References

- *Build a simple map*. (n.d.). Tableau. Retrieved September 9, 2025, from https://help.tableau.com/current/pro/desktop/en-us/maps_howto_simple.htm
- *Create views in Tooltips (Viz in Tooltip)*. (n.d.). Tableau. Retrieved September 8, 2025, from https://help.tableau.com/current/pro/desktop/en-us/viz_in_tooltip.htm
- *Tableau Community Forums*. (n.d.). Tableau. Retrieved September 9, 2025, from https://community.tableau.com/s/question/0D54T00000C6117SAB/sum-values-from-multiple-fields