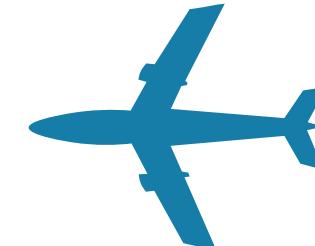


Data Analytics Portfolio:

# Flight Delay Cause Analysis

Output: Dashboard Visualization with Tableau

***Presented by Raon Spielberg Berek***  
Junior Data Analyst



**Tools used:**



Microsoft Excel





## About me:

Hi, I'm **Raon Spielberg Berek**, a final-year Information Systems student at Harapan Bangsa Institute of Technology and a **Junior Data Analyst**. I possess strong **public speaking** and **analytical skills**, honed through experience coordinating numerous projects and events, as well as working on various data analysis projects using **Excel, SQL, Python, Tableau, PowerBI, and Looker**.

## Employment Objective:

A career in **Information Systems**, specifically in the areas of **Data Analysis**, leveraging proficiency in data analysis and visualization tools (**SQL, Python, Tableau, etc.**), and **analytical thinking** and **communication skills**, to translate complex data into **actionable insights** and guide clients through their **transformation journey**.



Kalideres, West Jakarta



raonbereknew@gmail.com



Raon Spielberg Berek

# Experience and Certifications



**Data Mining Lab Assistant Lecturer**  
- Institut Teknologi Harapan Bangsa, Bandung City

Feb 2025 - Jun 2025



**Data Scientist Intern (Project-Based)**  
- PT IDX Consulting, Online

Mar 2025 - Mar 2025



**Machine Learning Distinction Graduate**  
- Google Bangkit Academy 2024 Batch 1, Online

Feb 2024 - Jun 2024



**Learn Data Analysis with Python**  
- Dicoding Indonesia, Online

Mar 2024 - Mar 2024

# Agenda

**1** About the Project: Business Understanding

**4** Modeling and Evaluation (Tableau Dashboard)

**2** Data Understanding

**5** Strategic Recommendations to Reduce Delays

**3** Data Preparation

**6** Deployment (Tableau Public)

# Business Understanding



Image Source: NewsBytes

## Background:

- The aviation industry relies heavily on **time efficiency** to maintain smooth operations.
- Flight delays represent a recurring issue that leads to **significant financial losses** for airlines.
- Delays **negatively impact customer satisfaction** and **trust**. They also disrupt airport efficiency and overall operational performance.

## Dataset provided:

- **Name:** Airline On-Time Statistics and Delay Causes.
- **Source:** Bureau of Transportation Statistics (BTS), USA.
- **Time Range:** May 2023 – May 2025.
- Dataset covers flight schedules, **delays**, and their contributing factors.

# Business Understanding - Problem Statements

Based on the background, management needs, and the dataset provided, the following **problem statements** will be addressed in this project.

1. What are the **different types of flight delays** and how many delays fall into each category?
2. What are the **top-ranking causes of flight delays** based on their frequency of occurrence?
3. How do **carrier-related delays rank** in terms of total delay counts?
4. How has **the average percentage of carrier-related delays changed over time?**
5. How is the **total delay distributed** across U.S. states and cities?



Image Source: istockphoto

# Business Understanding - Project Objectives

To address the identified problem statements, the following **project objectives** are defined. An **interactive dashboard** will be developed with **filters** for month, year, airport name, state, and city, featuring the following visualizations:

1. **Scorecards** displaying the total number of delays and the count of each delay type.
2. **Bar chart** showing the ranking of delay causes by frequency of occurrence.
3. **Bar chart** ranking airlines based on the number of carrier-related delays.
4. **Line chart** illustrating the trend of the average carrier delay percentage over time.
5. **Map chart** highlighting flight delay hotspots across U.S. states and cities.



Image Source: Canva

# Data Understanding

The image shows two screenshots. The top screenshot is from the Bureau of Transportation Statistics website, specifically the 'Airline On-Time Statistics and Delay Causes' page. It includes a search bar, navigation links for Topics and Geography, Statistical Products and Data, National Transportation Library, and News. Below the header is a note about the data being for January 2018. A form allows users to select a carrier, an airport, and dates. The main content is a table titled 'On-Time Arrival Performance' for May 2023, listing columns such as year, month, carrier, carrier\_name, airport, airport\_name, arr\_flights, arr\_del15, carrier\_ct, weather\_ct, and nas\_ct. A note explains the arr\_del15 column as the difference between actual and scheduled arrival times. The bottom screenshot shows a data analysis interface with a preview of a CSV file named 'Airline\_Delay\_Cause\_2023-2025'. The preview table has columns labeled 'lat', 'year', 'month', 'carrier', 'carrier\_name', 'airport', 'airp', 'alb', 'atl', 'atw', 'aus', 'avl', 'bgr', 'bhm', 'bis', 'bmi', and 'rna'. The interface also shows a 'Queries & Connections' panel indicating 47,097 rows loaded.

- The dataset contains **21 columns** related to airline on-time statistics and delay causes.
- Most columns record the **number of delays by type** and **the duration of delays** (in minutes).
- The column **arr\_del15** is an **arrival delay indicator**:
  - **≥ 15 minutes → Delay**
  - **< 15 minutes → On-Time**
- Total records: **47,097 rows** (from May 2023 – May 2025).
- 57 rows contain **missing values**.
- Data types:
  - **Text** → 4 columns
  - **Whole Number** → 12 columns
  - **Decimal** → 5 columns

# Data Preparation

	A	B	C	D	
1	year	month	date	carrier	carrier_name
2	2025	5	May-25	9E	Endeavor Air Inc.
3	2025	5	May-25	9E	Endeavor Air Inc.
4	2025	5	May-25	9E	Endeavor Air Inc.
5	2025	5	May-25	9E	Endeavor Air Inc.
6	2025	5	May-25	9E	Endeavor Air Inc.
7	2025	5	May-25	9E	Endeavor Air Inc.
8	2025	5	May-25	9E	Endeavor Air Inc.
9	2025	5	May-25	9E	Endeavor Air Inc.
10	2025	city	state	airport_name	
11	2025	Allentown/Bethlehem/Easton	PA	Lehigh Valley International	
		Albany	GA	Southwest Georgia Regional	
		Nantucket	MA	Nantucket Memorial	
		Alexandria	LA	Alexandria International	
		Augusta	GA	Augusta Regional at Bush Field	
		Albany	NY	Albany International	
		Atlanta	GA	Hartsfield-Jackson Atlanta International	
		Appleton	WI	Appleton International	
		Austin	TX	Austin - Bergstrom International	
		Asheville	NC	Asheville Regional	
		Kalamazoo	MI	Kalamazoo/Battle Creek International	
		Binghamton	NY	Greater Binghamton/Edwin A. Link Field	
		Bangor	ME	Bangor International	
		Birmingham	AL	Birmingham-Shuttlesworth International	
		Bloomington/Normal	IL	Central IL Regional Airport at Bloomington	
		Nashville	TN	Nashville International	
		Boston	MA	Logan International	
		Brunswick	GA	Brunswick Golden Isles	
		Burlington	VT	Burlington International	
		Buffalo	NY	Buffalo Niagara International	
		Baltimore	MD	Baltimore/Washington International Thurgood Marshall	

- Dropped 9 irrelevant columns:
  - arr\_flights, arr\_cancelled, arr\_diverted
  - Delay duration columns (in minutes): arr\_delay, carrier\_delay, weather\_delay, nas\_delay, security\_delay, late\_aircraft\_delay
- Removed rows with **missing values** (57 rows).
- Created a new column **date** (Month-Year format).
- Extracted **city** and **state** from **airport\_name** into separate columns, and **trimmed extra spaces**.

# Modeling & Evaluation - Tableau Worksheets

- **Filters:**

A screenshot of a Tableau interface showing four filter dropdowns. The first dropdown is set to 'May 2025'. The subsequent three dropdowns are all set to '(All)', indicating no specific filter has been applied to those dimensions.

- **Viz 1: Scorecards**

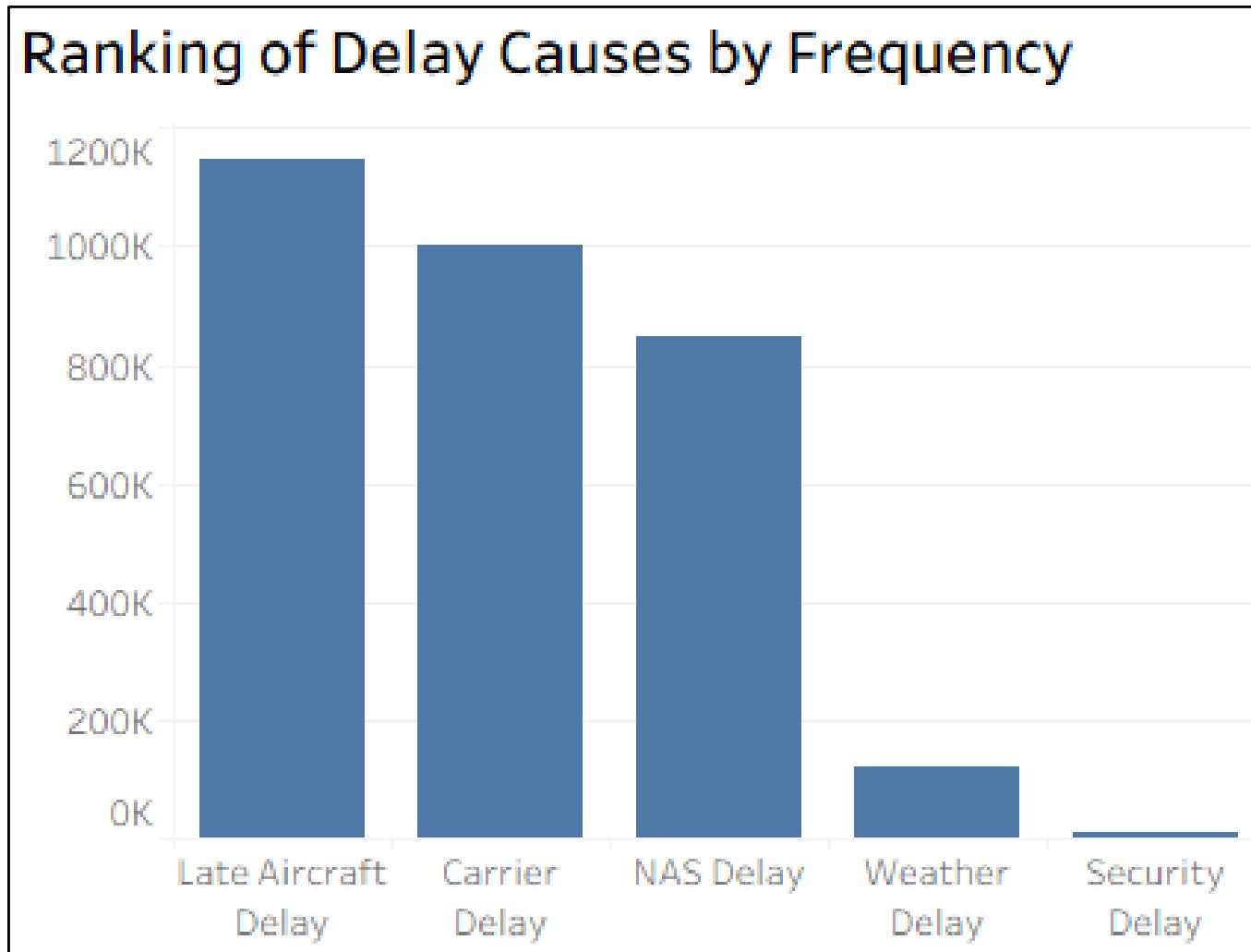
Delay Total	Carrier Delay	Late Aircraft Delay	NAS Delay	Security Delay	Weather Delay
3,128,184	1,003,392	1,149,371	849,733	8,237	117,447

## Insights:

- All filters function as intended, **successfully filtering the data** across the entire dashboard.
- The first visualization (Viz 1) displays the **total count of delays for each of the five cause categories**: Carrier, Late Aircraft, NAS, Security, and Weather Delay.
- **Late Aircraft Delay** is the most frequent cause, accounting for **over 1.1 million incidents**, while **Security Delay** is the least frequent, with **fewer than 8,500 incidents**.

# Modeling & Evaluation - Tableau Worksheets

- Viz 2: Bar Chart

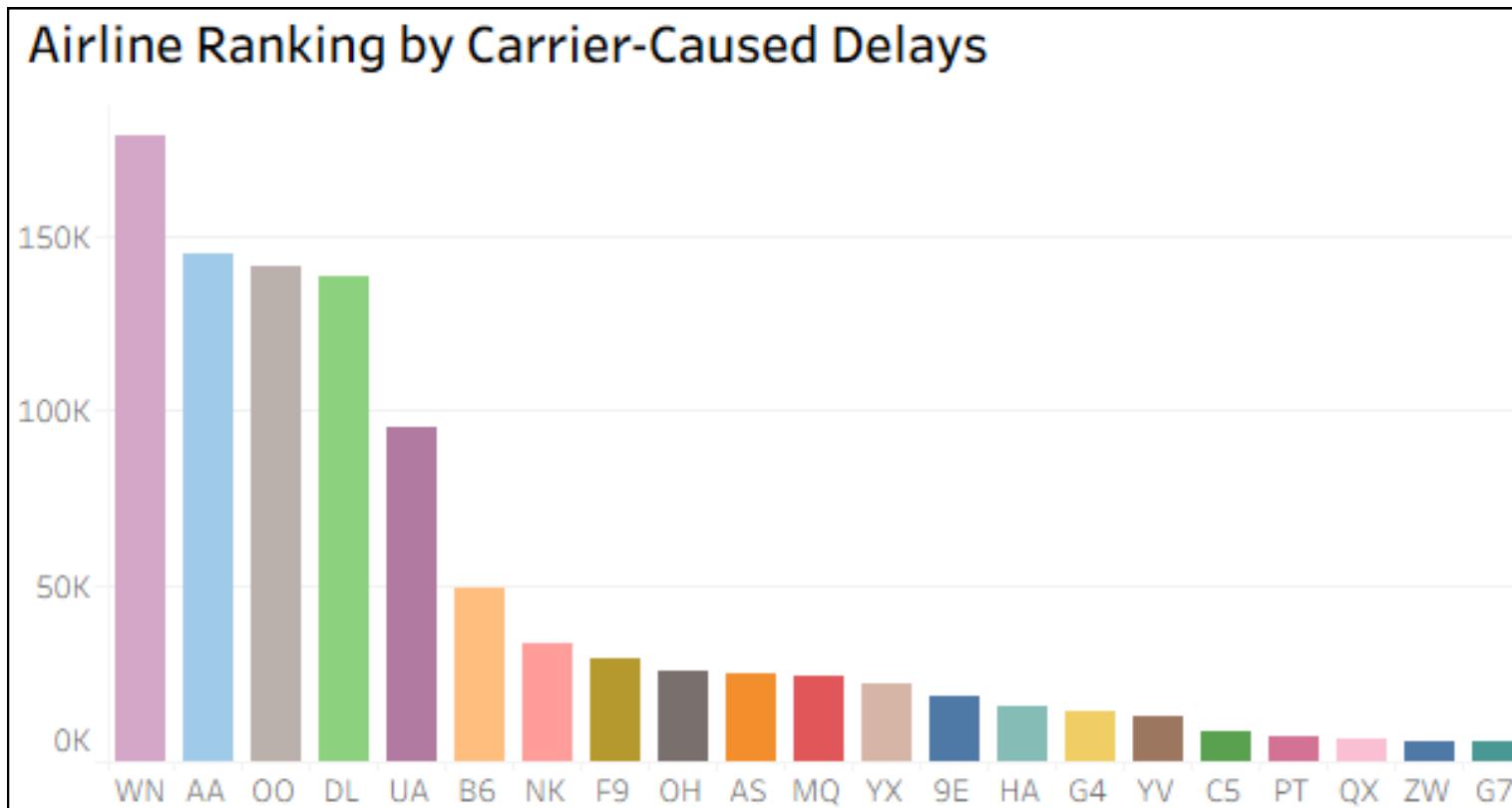


## Insights:

- **Top 3** Delay Causes: Late Aircraft Delay, Carrier Delay, and NAS Delay.
- Delays from late-arriving aircraft (Late Aircraft Delay) are the **primary driver**, indicating efficiency issues in connecting flight schedules.
- Two of the top three causes (Late Aircraft & Carrier Delay) are **directly within the airline's control**, accounting for the vast majority of all incidents.
- By frequency, delays from external factors like Weather and Security are **far less common** than internal operational issues.
- To achieve the greatest impact, **prioritize improvements in aircraft turnaround efficiency and crew scheduling**.

# Modeling & Evaluation - Tableau Worksheets

- **Viz 3: Bar Chart**

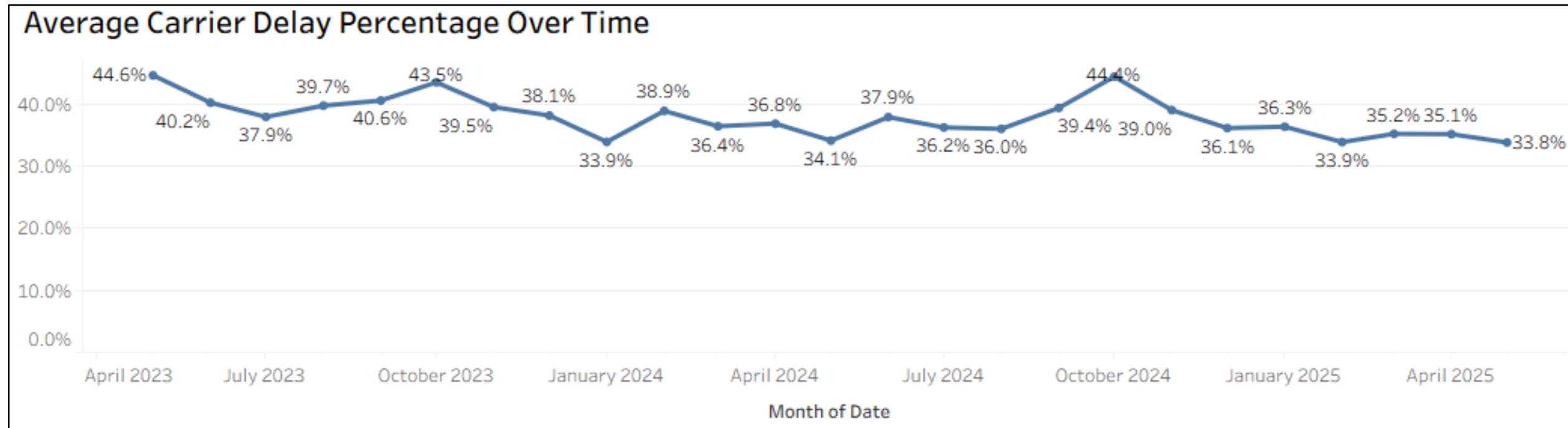


## Insights:

- A small group of airlines, led by **Southwest** (WN), **American** (AA), and **SkyWest** (OO), are responsible for the highest volume of self-inflicted delays.
- The top 5 airlines account for a disproportionately large share of all carrier-caused delays, indicating that industry-wide issues are **heavily concentrated within these key players**.
- A clear **performance gap exists** between the top-ranking airlines and the rest of the industry, highlighting a wide variation in operational efficiency and reliability.
- Focus operational improvement efforts on the **top 5 carriers** to address the **root causes** of their delays for the most significant industry-wide impact.

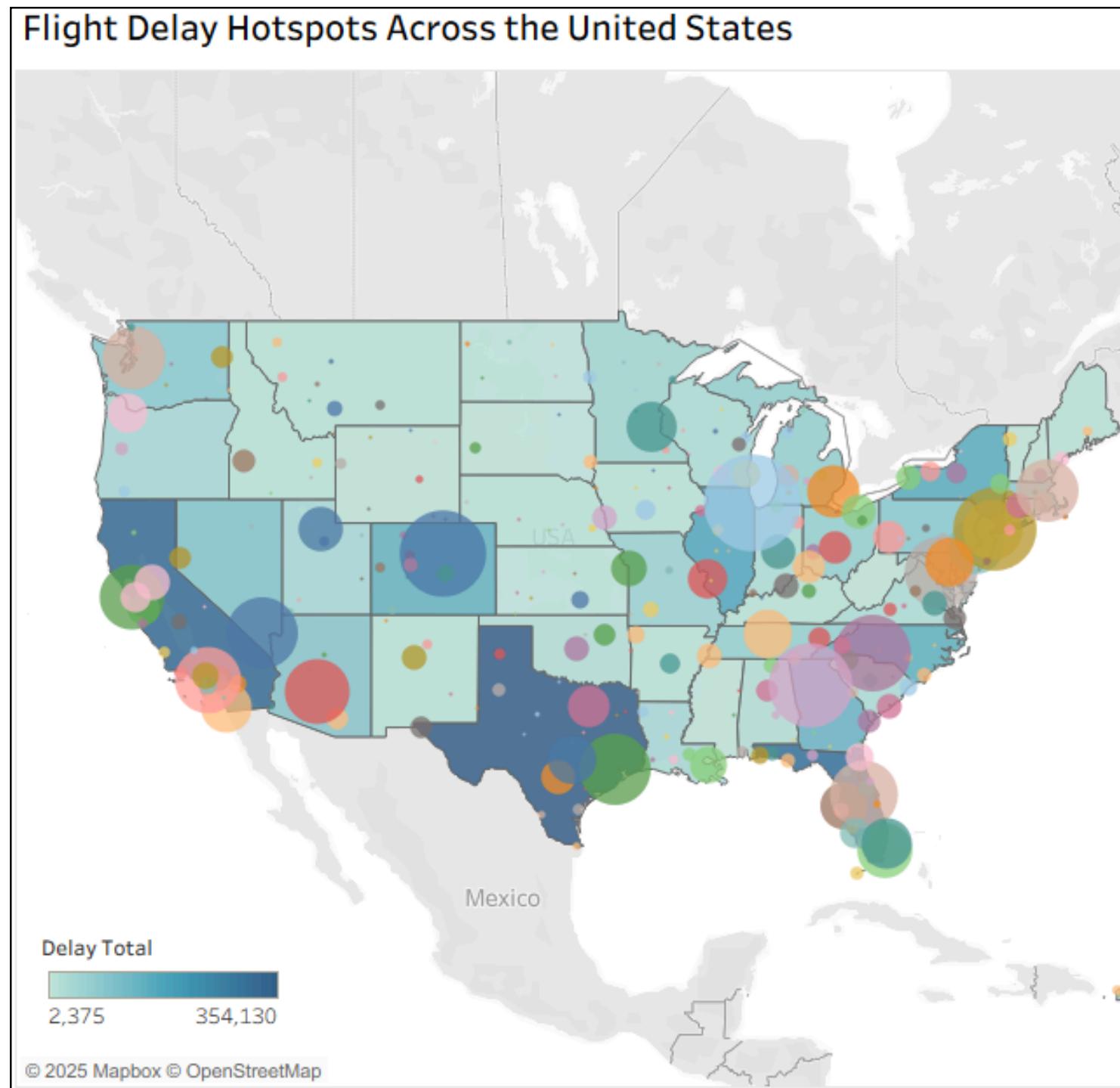
# Modeling & Evaluation - Tableau Worksheets

- **Viz 4: Line Chart**



# Modeling & Evaluation - Tableau Worksheets

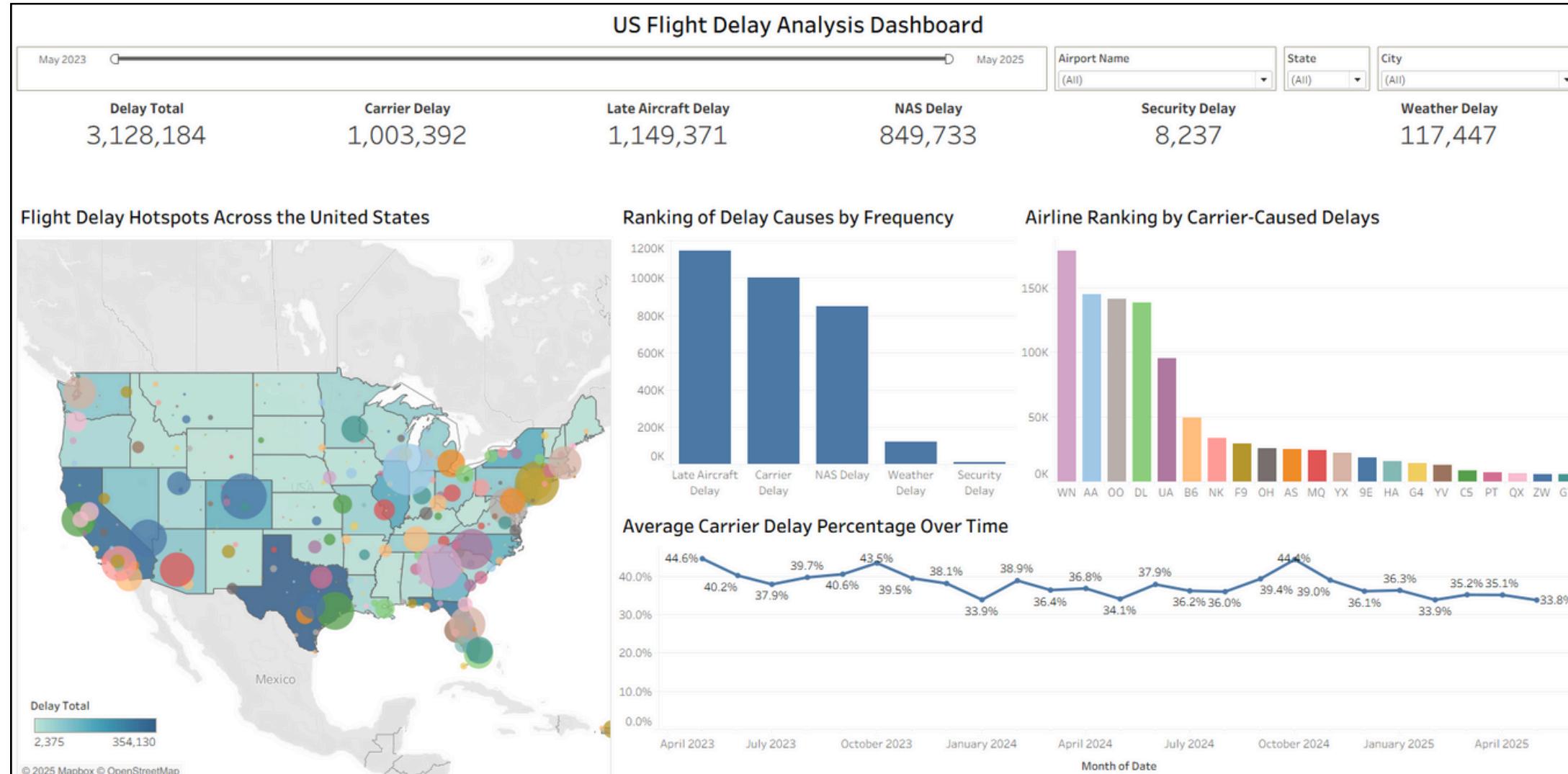
- **Viz 5: Map Chart**



## Insights:

- The total impact of flight delays is **not spread evenly across the country** but is heavily concentrated in a handful of **major hub airports** and **high-traffic regions**.
- Key delay hotspots are **clustered around the nation's busiest airports**, including those in the Northeast (NYC area), Midwest (Chicago), Southeast (Atlanta), and on the West Coast (California).
- The map reveals a pattern where delays are most prominent on the **East and West coasts**, with significant inland **hotspots** appearing exclusively at **major connecting hubs** like Denver and Dallas.
- For the greatest national impact, **focus delay-reduction** initiatives and **resources** on the **top 15 hotspot airports**, as improvements in these critical locations will **ripple positively throughout the entire system**.

# Modeling & Evaluation - Tableau Dashboard



## Insights:

- While total delays are massive, the **issue is not widespread**. The impact is heavily concentrated within a few key hub airports and driven by a small number of major airlines.
- The vast majority of delays stem from internal airline operations—the "**domino effect**" of late aircraft and carrier-specific issues—rather than external factors like weather or security.

- Airline operational **performance is not random**; it follows a clear seasonal cycle, consistently worsening in the **late Summer through Fall**.
- A targeted strategy focusing on improving operational efficiency at **key hotspot airports** during **predictable peak seasons** will yield the most significant reduction in delays nationwide.

# Conclusion

1

## The Problem is Concentrated

- Flight delays are **not a widespread issue but a concentrated one**, with the largest impact originating from a handful of key hub airports and major airlines.

2

## The Root Cause is Operational

- The primary driver of delays is **overwhelmingly operational**, stemming from internal airline efficiency rather than external factors like weather or security.

3

## The Pattern is Predictable

- Delay **patterns are not random but cyclical**, following a consistent seasonal trend that consistently peaks in the **late summer and fall**, allowing for proactive measures.

# Strategic Recommendations to Reduce Delays

1

## Launch a Targeted Efficiency Program at Key Hubs

- **Why:** Our analysis shows that the vast majority of delays are operational and highly **concentrated** within a few major airports.
- **Action:** Conduct a deep-dive review of aircraft turnaround processes and crew scheduling at the **top 15 hotspot airports** to identify and resolve critical bottlenecks.

2

## Implement Proactive, Seasonal Resource Planning

- **Why:** Performance data reveals a **predictable seasonal cycle**, with operational delays consistently peaking during the late Summer and Fall months.
- **Action:** Develop an enhanced **resource allocation plan** for ground staff, technicians, and reserve crews to prepare for these identified high-risk periods.

# Deployment (Tableau Public)

 : [US Flight Delay Cause Analysis Dashboard](#)

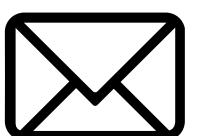


: [US Flight Delay Cause Analysis Dashboard by raonsb](#)

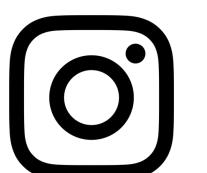
# Thank You

*If you have any questions, discussions, or potential collaborations,  
please feel free to contact me through the details below.*



 [raonbereknew@gmail.com](mailto:raonbereknew@gmail.com)

 [Raon Spielberg Berek](https://www.linkedin.com/in/Raon-Spielberg-Berek)

 [@raonsb](https://www.instagram.com/@raonsb)