

# Tableau Project Report: Delayed Flight Analysis

Shreya Todmal, MTech GI  
Indian Institute of Technology, Kanpur

---

## Problem Statement

The primary objective of this analysis is to investigate the causes and patterns of flight delays across different airlines, airports, and time periods using Tableau. By understanding these patterns, the analysis aims to identify key factors contributing to delays, which can help the aviation industry make informed decisions to minimize future delays.

## Data Overview

The dataset includes information on flight delays across various airlines and airports. Each record in the dataset represents flight delay details for a specific month and year, categorized by different causes of delay. With 13,700 records, the dataset is substantial enough to conduct a detailed and comprehensive analysis of flight delays.

Key Columns:

- Year: The year the data was recorded.
- Month: The month the data was recorded.
- Carrier: The code representing the airline.
- Airport: The airport code where the flight was scheduled.
- Arrivals: Total number of flights arriving.
- Arrivals on Time: Total number of flights that arrived on time.
- Carrier Delay: Total number of minutes delayed due to the airline.
- Weather Delay: Total number of minutes delayed due to weather conditions.
- NAS Delay: Total number of minutes delayed due to National Airspace System (NAS) issues.
- Security Delay: Total number of minutes delayed due to security reasons.
- Late Aircraft Delay: Total number of minutes delayed due to late-arriving aircraft.
- Total Delay: The sum of all delay causes for the period.

## Methodology

### Step 1: Data cleaning and preparation

Ensure the dataset is clean and free from missing or erroneous data. Convert necessary columns to appropriate data types for date and time. Split the column for airport into city, state, airport name.

airline_delay_causes.csv	airline_delay_causes.csv	airline_delay_causes.csv
Airport Name	Arrival Flights	Delays
Cleveland, OH: Cleveland-Hopkins International	230	
Denver, CO: Denver International	258	
Dallas/Fort Worth, TX: Dallas/Fort Worth International	608	
Detroit, MI: Detroit Metro Wayne County	890	
Newark, NJ: Newark Liberty International	293	
Fort Lauderdale, FL: Fort Lauderdale-Hollywood International	1,483	
Niagara Falls, NY: Niagara Falls International	30	
Houston, TX: George Bush Intercontinental/Houston	496	
Las Vegas, NV: McCarran International	1,054	
Los Angeles, CA: Los Angeles International	682	

Calculation	Calculation	Calculation
City	State	Airport Name
Cleveland	OH	Cleveland-Hopkins International
Denver	CO	Denver International
Dallas/Fort Worth	TX	Dallas/Fort Worth International
Detroit	MI	Detroit Metro Wayne County
Newark	NJ	Newark Liberty International
Fort Lauderdale	FL	Fort Lauderdale-Hollywood International
Niagara Falls	NY	Niagara Falls International
Houston	TX	George Bush Intercontinental/Houston
Las Vegas	NV	McCarran International
Los Angeles	CA	Los Angeles International

Figure 1 Split the airport column with three more column

Tables

- Airline
- Airport
- Airport\_Name
- City
- month
- State
- Year
- Measure Names
- Arr Cancelled
- Arr Diverted
- Arrival Flights
- Carrier Ct
- Delayed Flights
- Late Aircraft Ct
- Nas Ct
- Security Ct
- Weather Ct
- airline\_delay\_causes
- Measure Values

Filters

Marks

Automatic

Color Size Text

Create Calculated Field...

Create Parameter...

Create Folder (use group by folder)

Group by Folder

Group by Data Source Table

Sort by Name

Sort by Data Source Order

Hide All Unused Fields

Show Hidden Fields

Expand All

Collapse All

Sheet

Date

DATE (DATEPARSE ("YYYY-MM", STR ([Year]) + '-' + STR ([month])) ) )

The calculation is valid.

Apply OK

Figure 2 Create a Date parse as Year is misclassified in dimension

## Step 2: Exploratory Data Analysis

Using Tableau to create visualizations that explore the distribution of delays across different variables. Identify patterns and correlations between delay causes and other factors like time of year, airline, and airport.

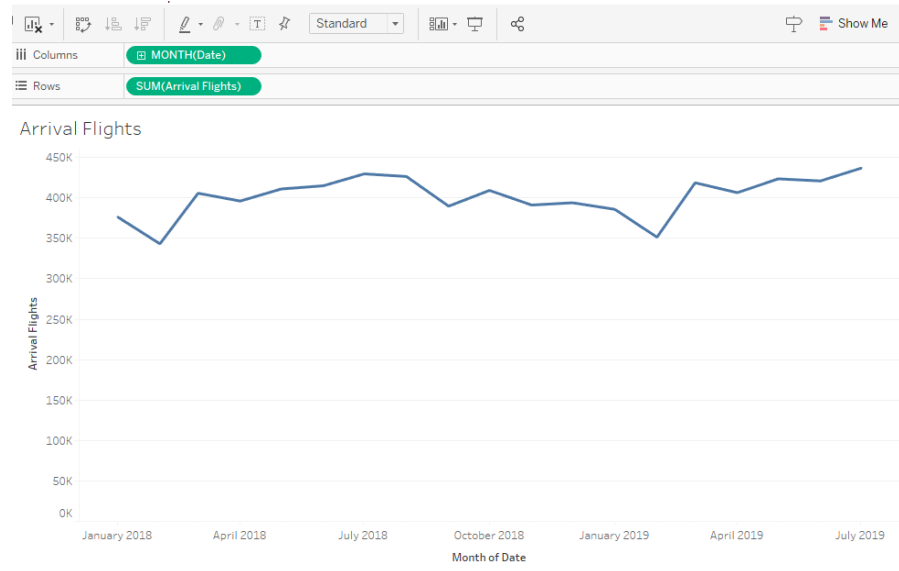


Figure 3 Flights arrived monthly throughout the year

The data is from January 2018 to July 2019 only, the above graph gives a trend of arrived flights.

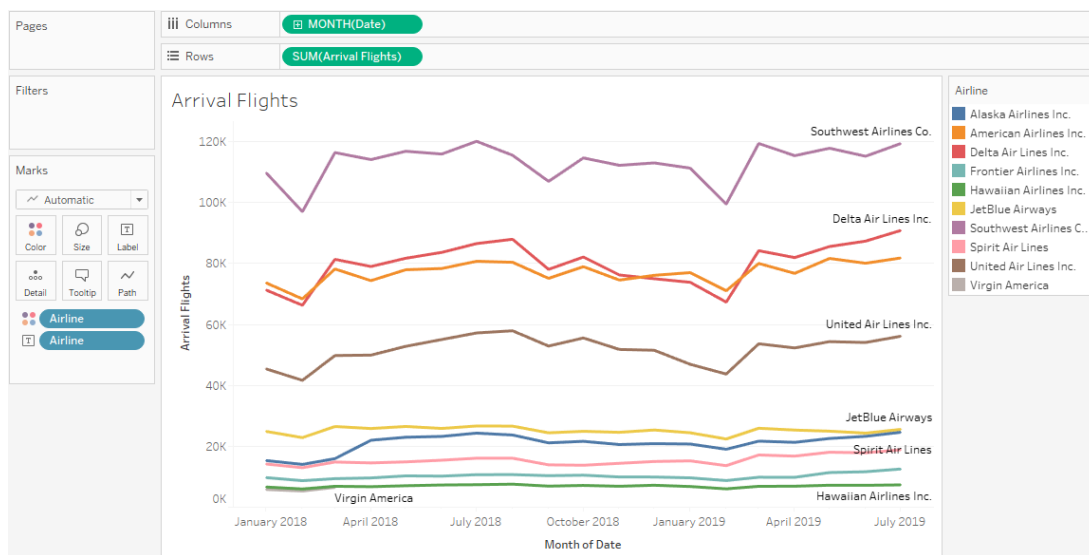


Figure 4 Arrival flight per airline

Here, the most number of flights arrived are of Southwest, Delta and American Airlines.

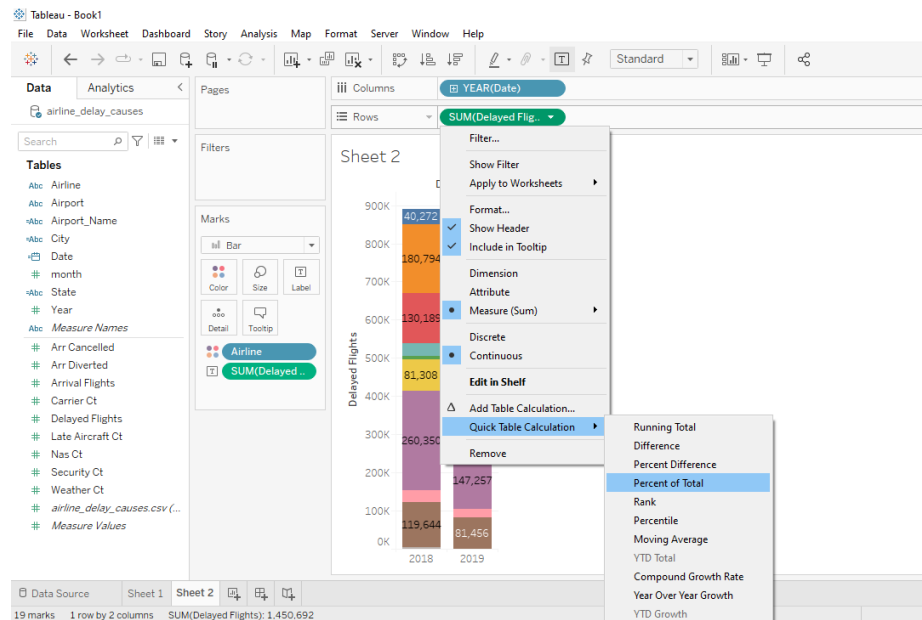


Figure 5 Converting the data representation to total percentage

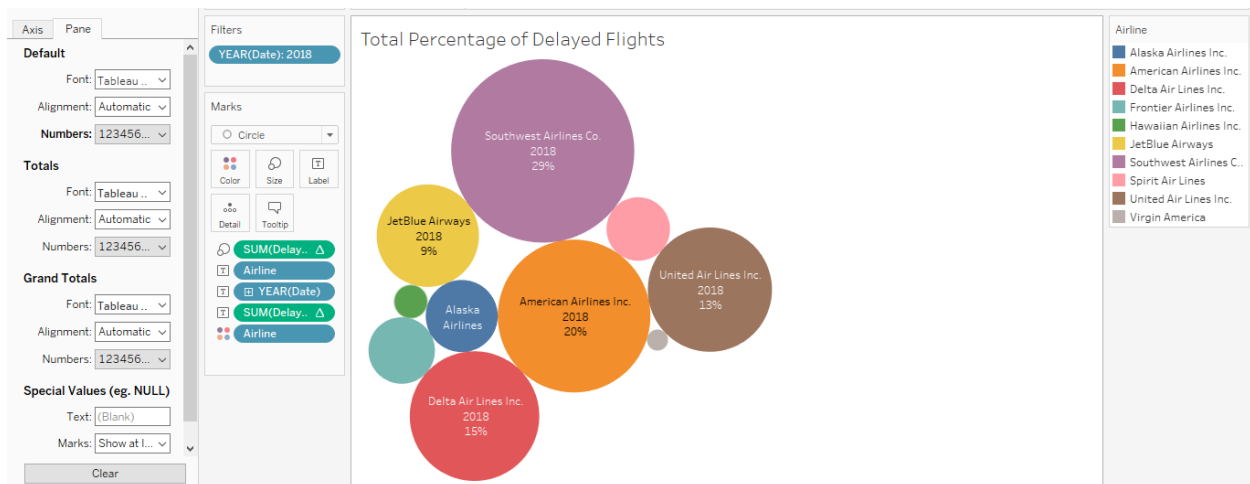


Figure 6 Total percentage of Delayed Flights per Airline

Here, the highest percentage of Delayed flights for the year 2018 is of Southwest Airline followed by American Airline and Delta Airline.

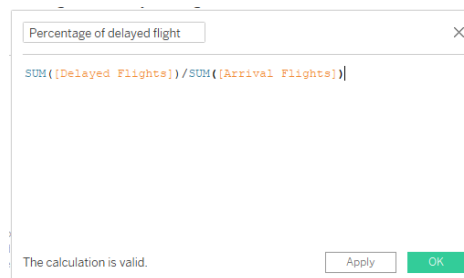


Figure 7 Creating a calculated field for percentage of delayed flight

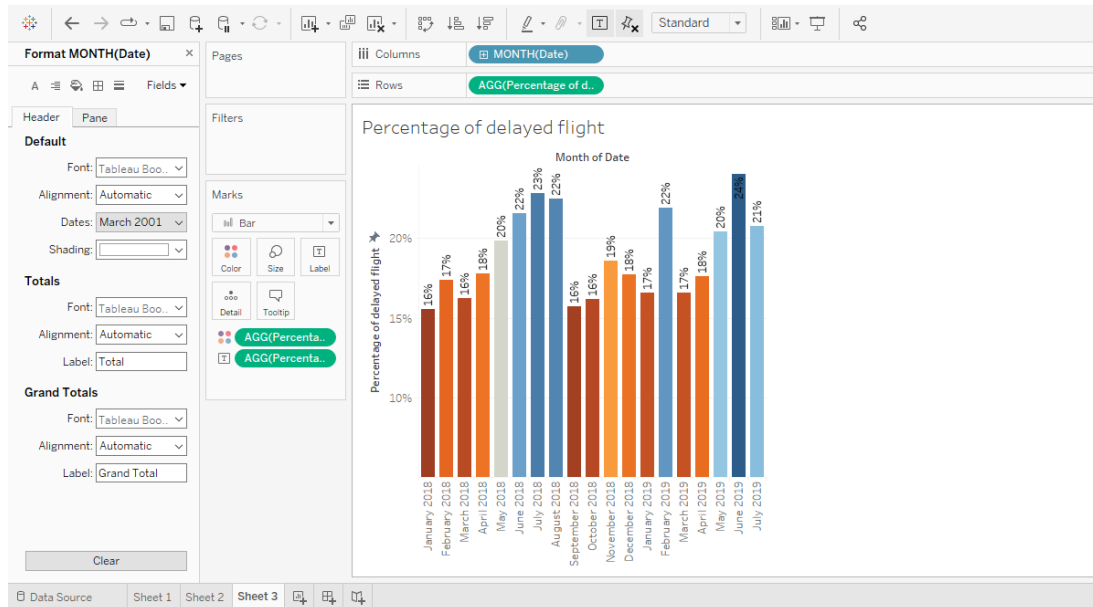


Figure 8 Percentage of delayed flights

The maximum percentage of delayed flights based on arrival flights is less than 25%. The maximum delay is caused in month of June and July in both the years.

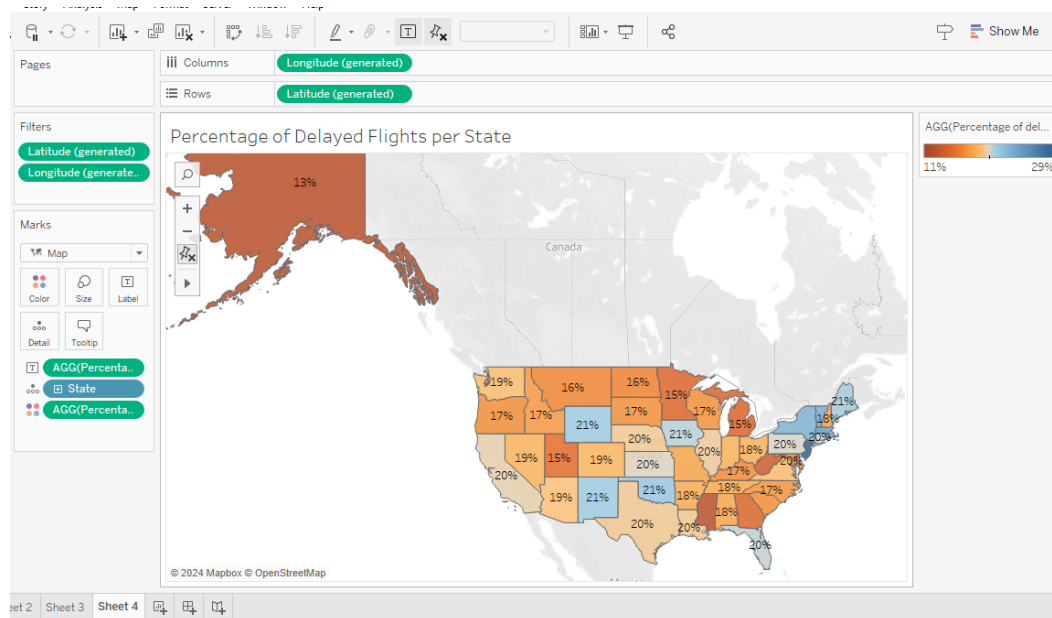


Figure 9 Percentage of Delayed flight per state

### Step 3: Trend and Cause Analysis

Analyze trends in delays over time to identify any seasonal patterns or long-term improvements/declines. Break down the delays by cause to see which factors are contributing the most to flight delays.

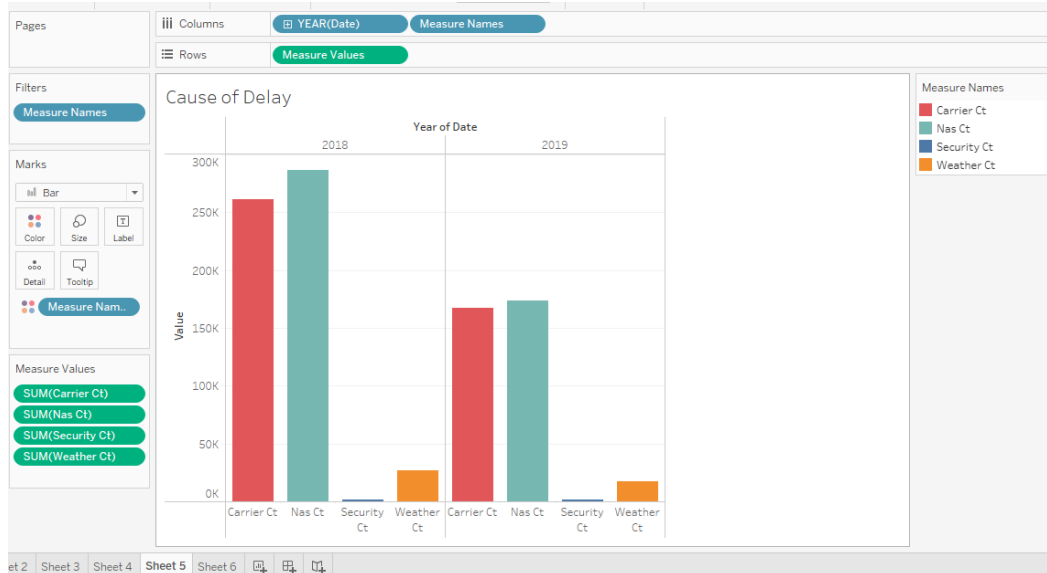


Figure 10 Causes of Delay

From the fig (10) it is clear that major delays are caused by National Airspace Issue and the second major delay is due to Carrier delay.

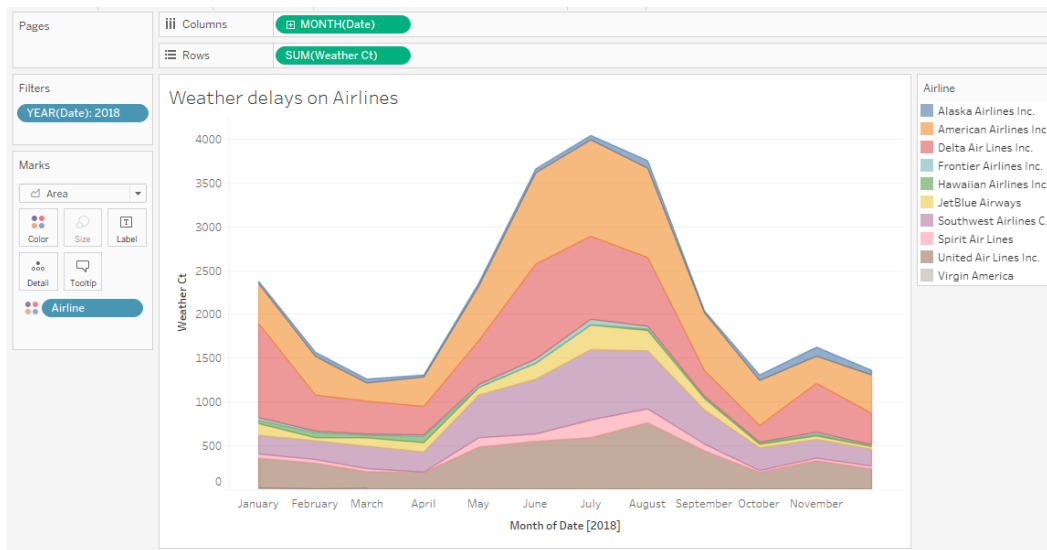


Figure 11 Weather Delay on Airline

Weather Delay is mostly in month of June, July and August. The major airlines affected are American and Delta airlines.

## Insights and Recommendation

### 1. Flights Show a Seasonal Trend

- Observation: The volume of flights peaks in July, while February has the lowest flight volume.

- Explanation: The increased flight volume in July is likely due to the summer vacation season, where demand for air travel significantly rises as people travel for vacations, holidays, and family visits. Conversely, February's lower volume can be attributed to it being the shortest month of the year, as well as the post-holiday period when travel demand typically decreases.
- Implications: Airlines and airports can use this information to better prepare for the seasonal surge in passengers during the summer months by increasing staff, resources, and operational capacity to handle the higher demand. Conversely, during February, they can optimize operations to manage the lower flight volumes, potentially reducing operational costs.

## 2. National Airspace Security (NAS) is the Major Cause of Delays

- Observation: The majority of flight delays are attributed to National Airspace System (NAS) issues.
- Explanation: NAS delays occur due to a variety of factors, including air traffic control, airport operations, heavy traffic volume, and airspace limitations. The high percentage of delays caused by NAS suggests systemic challenges in managing air traffic and maintaining smooth operations across the national airspace.
- Implications: Addressing NAS delays requires a coordinated effort between airlines, airports, and regulatory bodies like the FAA. Improvements could involve upgrading air traffic control technology, optimizing flight schedules to reduce congestion, and enhancing communication between airports and airlines.

## 3. Southwest Airlines Experiences the Most Delays

- Observation: Southwest Airlines accounts for 29% of the total flight delays, the highest among all airlines, whereas Virgin America has the lowest delay rate, close to 0%.
- Explanation: The high delay rate for Southwest Airlines are due to various factors such as the volume of flights operated and the complexity of its network. Virgin America's low delay rate might be attributed to its smaller scale of operations.
- Implications: For Southwest Airlines, these findings highlight the need to review and improve operational practices to reduce delays. This could include better scheduling, enhanced aircraft turnaround processes, or more efficient resource allocation. On the other hand, Virgin America's practices could be studied and possibly replicated by other airlines to improve their on-time performance.

## 4. New Jersey State Has the Highest Number of Delays

- Observation: New Jersey reports the highest percentage of flight delays at 29%.
- Explanation: New Jersey's high delay rate is likely linked to the significant traffic at major airports like Newark Liberty International Airport (EWR), which is one of the busiest airports in the country and often faces congestion and weather-related challenges.
- Implications: Airports in New Jersey may need to invest in infrastructure improvements, better weather forecasting and management systems, and more efficient air traffic control

measures to mitigate delays. Coordination with airlines to optimize flight schedules and manage peak traffic periods could also help reduce the number of delays.

#### 5. Southwest Airlines Leads in Arrival Flights

- Observation: Southwest Airlines has the highest number of arrival flights, followed by Delta and American Airlines. Hawaiian Airlines and Virgin America have the fewest.
- Explanation: As one of the largest domestic carriers in the United States, Southwest Airlines' large number of arrival flights reflects its extensive network and frequent service. In contrast, Hawaiian Airlines and Virgin America's smaller networks result in fewer flights.
- Implications: The high volume of arrivals for Southwest Airlines necessitates efficient operations to manage the large number of flights and minimize delays. Airlines with fewer arrivals, like Hawaiian and Virgin America, may focus on niche markets or specialized services, allowing them to maintain high service quality and on-time performance.

#### 6. Weather Delays are High from June to September

- Observation: Weather-related delays peak between June and September.
- Explanation: This period coincides with the summer months, which often experience severe weather conditions such as thunderstorms, hurricanes, and high winds, especially in certain regions of the United States. These weather events can cause significant disruptions to flight schedules, leading to delays.
- Implications: Airlines and airports can prepare for the high likelihood of weather delays during these months by improving weather forecasting and communication with passengers, adjusting flight schedules to avoid severe weather periods, and enhancing operational flexibility to quickly respond to weather disruptions.

#### 7. Most Problematic Airports: ATL and LAX

- Observation: Atlanta's Hartsfield-Jackson International Airport (ATL) and Los Angeles International Airport (LAX) had the highest number of delays in 2018, with 73,951 and 56,981 delays, respectively.
- Explanation: Both ATL and LAX are among the busiest airports in the world, handling millions of passengers annually. The high number of delays at these airports can be attributed to heavy traffic, weather conditions, and complex operations. Additionally, their roles as major hubs mean that any delay can have a ripple effect, impacting connecting flights and other airports.
- Implications: To reduce delays, these airports might consider expanding infrastructure, improving air traffic management systems, and enhancing coordination with airlines to optimize flight schedules. Given their importance in the national and global aviation networks, reducing delays at these airports could have a significant positive impact on overall flight punctuality across the United States.