Case Study on AirSky Airlines

Analyzing Causes of Flight Delays using Tableau

DATA VISUALIZATION TECHNIQUES CRN-15360-202301

Prepared by
Angdembe, Paruhang
Baskaran, Vignesh
Dabas, Rahika
Dimaculangan, Roan Kathrina
Ip, Chak Leung Vincent
Suresh, Priyanka Yadhav
Muthusamy, Karthikeyan
December 13, 2023

1. Introduction

Overview

AirSky is newly established airline in the American aviation industry, connecting people and destinations with a commitment to innovation, and unparalleled service. Established in 2023 in North America, AirSky has evolved into a cornerstone of air travel, operating a vast fleet of state-of-the-art aircraft.

Despite maintaining a high standard of service, the airline has been experiencing delays in its flight schedules. To address this issue, the airline's data analytics team decided to conduct a thorough analysis of the causes of flight delays using Tableau.

Objective

The primary objective of this case study is to identify and analyze the key factors contributing to flight delays in AirSky Airlines. Through Tableau's data visualization capabilities, the team aims to gain insights that will enable the airline to implement targeted solutions and improve overall operational efficiency. Secondly, to provide solutions and/or recommendations on the directly impacted business processes of AirSky Airlines such as the following:

Identified Business Processes to Improve

Business Operation	Objective
Flight Operations	To reduce flight delays to optimize passenger experience with on-time performance.
Revenue Growth Strategy	To implement advanced predictive analytics models to anticipate potential delays based on historical data, weather forecasts, and other relevant variables.
	To develop proactive strategies for preemptive delay management, such as adjusting schedules, optimizing resources, and enhancing operational efficiency.
	To enhance real-time communication with passengers during delays, providing transparent information and managing expectations
	To implement effective service recovery strategies, such as compensation packages, loyalty program incentives, and personalized offers to retain and satisfy affected customers.

2. Data Source

Data Collection

Below are the datasets used to create hyper files using Tableau Prep Builder.

- **Airline Lat lng data.csv:** This dataset contains the latitude and longitude values of where the airport is situated with airport code as the main attribute.
- **Airline_Delay_Cause.csv:** Describes the delay of carrier, arrivals, NAS, security and weather in each airport.
- **Airline_data_passenger_dest us.csv:** Contains the details of people travelling to US with passenger Id and from which state or country they are travelling.
- **Airline_data_passenger_orig us.csv:** Contains data about people travelling from US to different state or nation. The data is described using passengerid and airlineid.
- **AverageFare_Annual_Merged.csv:** Contains the details of fare about the airplane the passenger is travelling. Also contains the state, the city the airport belongs to.
- **Us_disaster_declaration.csv:** Describes the type of disaster that has taken place in which place and in which year and month.

Data Preparation

Below are the data processing performed to the datasets to keep important fields, handle missing values, standardize formats, and ensure consistency across all datasets to create the final hyper files suitable for analysis in Tableau.

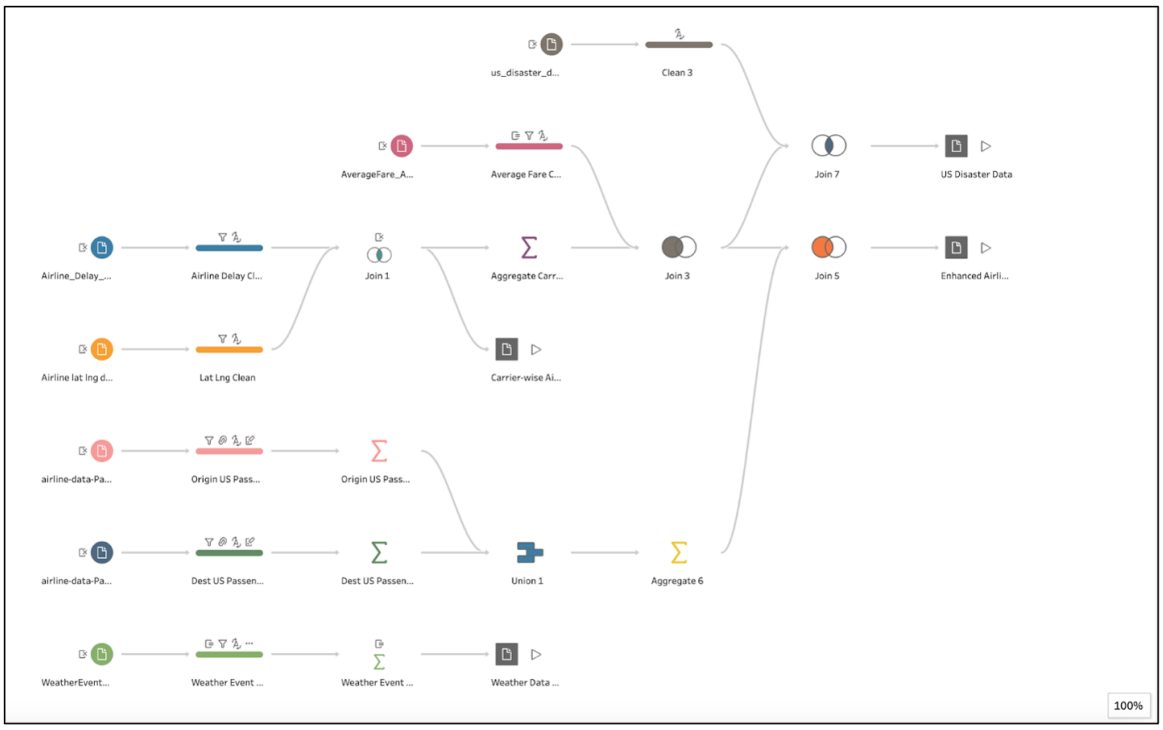


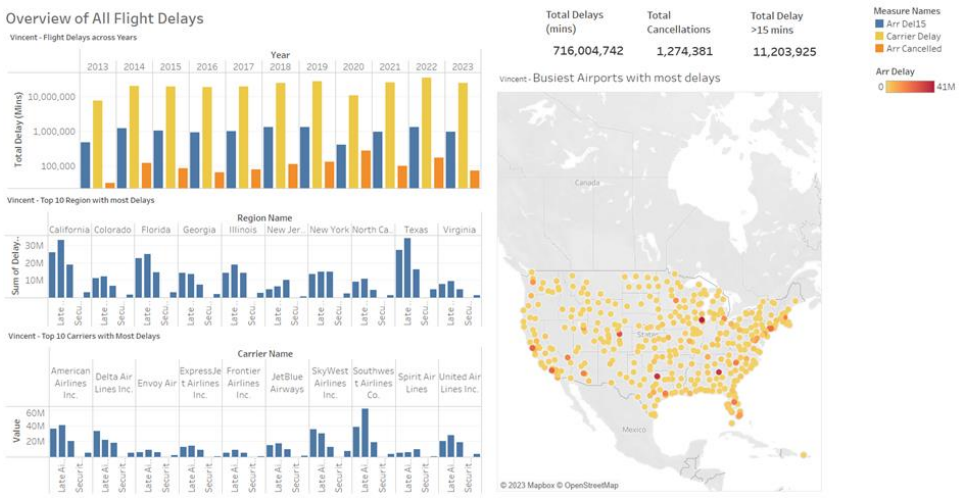
Tableau Data Processing Concept Used	Description
Data Connection and Cleaning	Connected multiple data sources, including Latitude-Longitude data, airline delay cause data, people traveling from/to US data, and US disaster declaration data.
	Identified and addressed missing values, outliers, and potential issues in the datasets.
	Cleaned and replaced data as needed to ensure quality and consistency.
Airline Delay Cause Visualization	Considered the airline delay causes dataset as the main source.
	Derived visualizations for delay categories such as carrier, weather, security, NAS (National Airspace System), and arrivals.
	Used the cleansed airline delay cause data, filtered by airport, to acquire specific delay causes for airlines.
Hyper File Creation - US Disaster Data	Cleaned and processed the US disaster declaration data.
	Performed an inner join with the cleansed airline delay cause data to create a hyper file named "US Disaster Data."
Hyper File Aggregation and Joining	Considered Airport Latitude and Longitude data.
	Performed an inner join with the US Disaster Data to create a hyper file.

	Combined both data sources using an inner join, aggregating to find carrier details for passengers.
US Disaster Declaration Data and Average Fare Data	Explored US_disaster_declaration data with details like declaration type, incident type, designated area, and dates.
	Considered Average Fare data with details on inflation-adjusted fare, airport details, and city names.
	Joined data based on airport codes to create a hyper file.
Visualization and Analysis	Demonstrated cleaning, aggregation, and joining processes to obtain four hyper files: US Disaster Data, Enhanced Airline Data, Carrier Data, and Weather Delay Data.
	Utilized these hyper files for analysis and visualization, integrating them into a scorecard.
Enhanced Airline Hyper File	Merged data sources containing details of people departing from and arriving to the US (airline_data_origin_US and airline_data_dest_US).
	Both datasets included state, city, and postal code, filtered based on state.
	Aggregated results and performed a union, followed by additional aggregation.
	Left joined the aggregate of airline causes with average fare and then left joined with US disaster data to generate the "Enhanced Airline" hyper file.

3. Tableau Analysis

Overview of Delays

This comprehensive dashboard provides a detailed overview of flight delays spanning the years 2013 to 2023. It offers insights into the distribution of total delays based on various criteria, including types of delays, regional breakdowns, carrier-specific data, and highlights the busiest airports experiencing the highest delay frequencies. The dashboard further presents key performance indicators such as total delay minutes, overall cancellations, and delays exceeding 15 minutes.



Top Left: Total Flight Delays Across Years

This visual representation illustrates the cumulative flight delays over the years, providing readers with a comprehensive overview of trends and patterns within the entire time span. The chart serves as a valuable tool for identifying any discernible fluctuations or recurring themes in the realm of flight delays.

Middle Left: Top 10 Regions with the Highest Delays

Delve into the geographical nuances of air travel disruptions with this chart showcasing the top 10 regions experiencing the highest delays. Beyond a mere ranking, the chart breaks down delays into sub-categories by type, unraveling the causes behind these delays. This insightful presentation aims to offer readers a nuanced understanding of regions with bustling air traffic and areas grappling with persistent delay challenges.

Bottom Left: Top 10 Carriers with the Most Delays

Navigate through the performance landscape of airlines as this chart unveils the top 10 carriers bearing the brunt of delays. Similar to the regional breakdown, the chart categorizes delays by type, offering readers a detailed insight into the factors influencing each airline's punctuality. This feature provides a comparative analysis, enabling readers to discern whether certain carriers are consistently outperforming others or facing significant challenges.

Right: Geographical Map of Busiest Airports with Significant Delays in the US

Embark on a geographic exploration with this map pinpointing the busiest airports in the United States grappling with substantial delays. By visually correlating locations with delay intensity, this map enhances comprehension of regional delay patterns, aiding readers in identifying potential geographical factors influencing delays.

Other Features:

In addition to the visual representations, key figures are highlighted to provide a quick snapshot of the overall impact of flight delays. Discover crucial metrics such as Total Delay

Minutes, offering a quantifiable measure of cumulative delay duration. Stay informed about Overall Cancellations, providing insights into the extent of disrupted travel plans. Furthermore, monitor Total Delays Exceeding 15 Minutes, offering a detailed perspective on delays with notable consequences for passengers and operational efficiency. This multifaceted dashboard ensures a comprehensive understanding of the complex landscape of flight delays in the aviation industry.

Finding the Airports with the greatest number of delays

In the vast expanse of air travel, understanding the intricacies of delays is paramount. Our journey begins with an exploration into the dynamic of air travel delays unravelling patterns, discovering correlations, and seeking insights that shape the passenger experience.

Passenger Count vs. Total Arrival Delay

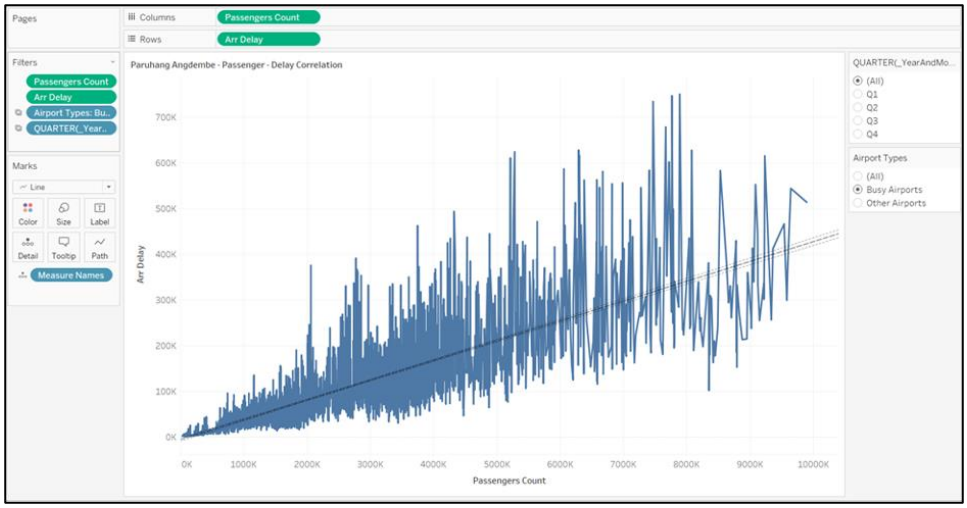


Figure 1 Correlation between Passenger Count and Total Arrival Delays

As we gaze upon Figure 1, a striking revelation emerges—a positive correlation between passenger counts and total arrival delays. As passenger counts soar, we observe a corresponding increase in total arrival delays. This intriguing correlation prompts us to delve deeper into the dynamics of air travel.

Insights:

Higher passenger counts may intensify operational complexities, contributing to delays in airport processes, boarding, and air traffic management. This prompts you to investigate further:

- Are delays occurring during the boarding process due to the higher volume of passengers?
- Is there a strain on airport resources, leading to congestion and longer waiting times?

Types of Delays and Its Significance

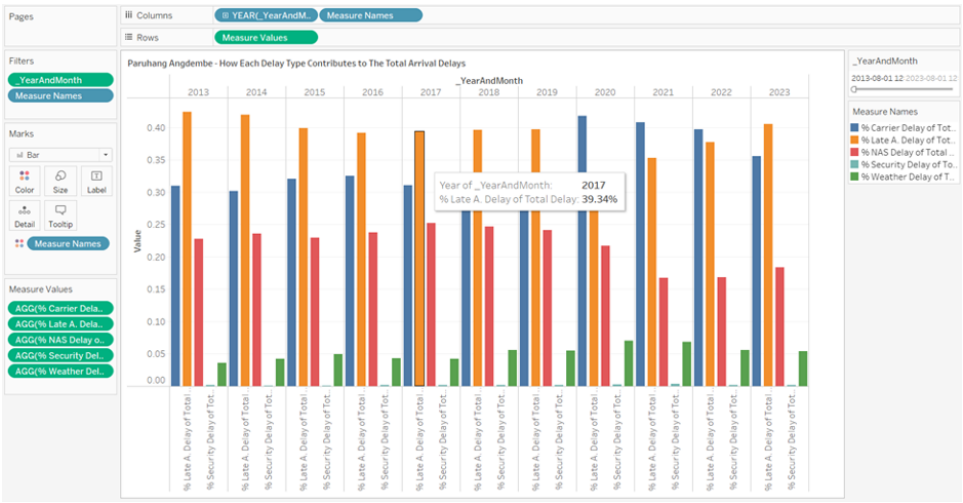


Figure 2 Percentage Contribution of Each Delay Type to Total Arrival Delay

Now, let's dissect the contributing factors behind these delays. Figure 2 reveals show cases the percentage contribution of each delay type to the total arrival delay.

Over the year we could see that *Carrier delays, Weather-induced disruptions, NAS-related challenges, Security Issues, and late aircraft arrivals*—each has a unique role in shaping the total arrival delay landscape.

Insights:

By examining the percentage contribution of each delay type, ***Carrier Delay and Late Aircraft Arrival Delay*** are the top 2 factors that consistently dominate the scene thus demanding a detailed further examination.

For our airline, identifying the major contributors to delays is pivotal:

- Are there specific carriers experiencing operational challenges?
- Is there a need for better scheduling to minimize late aircraft arrivals?

Addressing these questions becomes essential for optimizing overall performance.

Connection between Graphs:

Linking these insights, we contemplate the interplay between passenger count and delay types. Do higher passenger volumes amplify certain delay categories? Does the increase in passenger count intensify challenges for certain carriers or contribute to the late arrival of connecting flights?

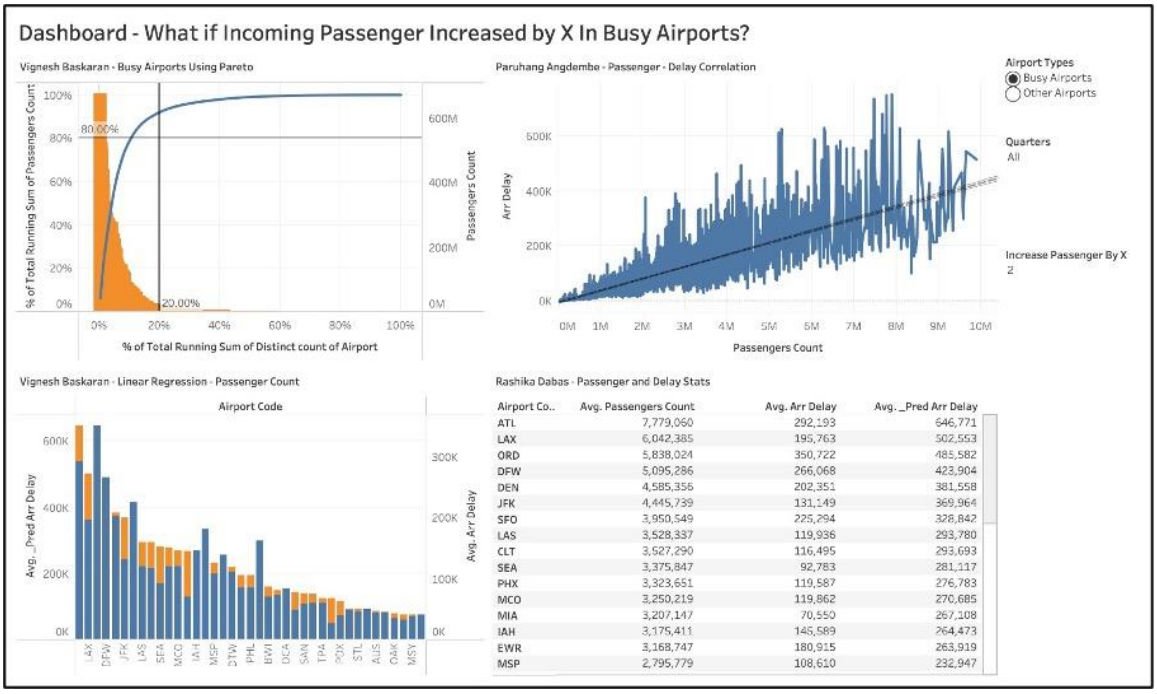
Sub-conclusion

In conclusion, understanding the intricate relationship between passenger count, delay types, and operational challenges, we are better equipped to implement targeted strategies. These strategies may include refining boarding processes, collaborating closely

with carriers to address operational issues, or implementing strategies to mitigate the impact of weather-related delays during peak seasons. Ensuring a seamless travel experience for passengers while maintaining operational excellence.

Analyzing the impact of increasing number of passengers by X in Busy Airport/s

The team performs an investigation on the impact of increasing the number of passengers in identified busy airports by using different visualization techniques such as (1) Data table for the detailed information and a granular view of key metrics, (2) Pareto Chart for visually highlighting the significant factors influencing outcomes, (3) Correlation Analysis to show understanding of relationships between different variables and a (4) Stack bar chart to show a dynamic representation of categorical data distribution



Top Left (Busy Airport Using Pareto)

During our investigation, we gathered the list of busiest airports from all the airports in the data source using Pareto Chart technique. We see that 20% of the busiest airports contribute 90% of the total passengers' count. Then, we created a set of high performing airports which we used as filters – busy airports (as the high performing airports) and other airports (as normal airports). This visualization is the foundation for all the other 3 visualizations we have created.

Top Right (Passenger - Delay Correlation Chart)

Using the correlation technique, we have found the relationship between passenger count and arrival delays using the R squared value. Since the R squared value is 0.6371 which is remarkably close to 1, we conclude that the number of arrival delays and the total passenger count are closely correlated to each other. The relationship can be summarized using this equation of the line:

$$\textbf{Arrival Delay} = 0.0433723*\text{Passengers Count} + -6610.8$$

Here, the P-value is less than 0.0001. The coefficients of the trend line can be described using the table below:

Term	Value	StdErr	t-value	p-value
Passengers Count	0.0433723	0.0005621	77.1647	< 0.0001
`1intercept	-6610.8	1796.95	-3.6789	0.0002379

After this, we have created a passenger multiplier parameter to modify the next chart.

Bottom Left (Linear Regression - Passenger Count Chart)

In this chart, we have used the Prediction method to determine the delays and compare them with the actual delay values. The blue bar represents the actual delays, and the orange bar indicates the predicted delays for some of the busy airports found at the start. We've created linear regression prediction using the correlation explained for the above chart and the created passenger multiplier parameter. So, ultimately this chart will predict the arrival delays if the passenger count is increased by X number. The predictions can be made for both busy and non-busy airports.

Bottom Right (Passenger and Delay Stats)

Passenger and Delay Stats text table shows all the numbers for the Average Passenger Count, Average Arrival Delays and Average of Predicted Arrival Delays for every busy airport found above. Data field values change significantly as the value for Increase Passenger By X filter is updated.

Carrier Delay

Datasource Used: Extract (Blended Datasource)

- Carrier delay analysis is crucial for airlines, impacting operational efficiency, customer satisfaction, financial performance, and strategic decision-making. It's a pivotal metric for evaluating airline performance and identifying areas for improvement within the aviation industry
- Analyzing trends or patterns in carrier delays allows for identifying areas of improvement within the carrier's operations. It can guide strategies to reduce delays and enhance overall service reliability.

Insights:

1. Carrier Delay Analysis – Regional Wise
2. Carrier delay Pareto analysis in relation to Carrier line
3. Carrier delay impact over Average fare and inflated average fare across regions

Carrier Delay Analysis – Regional Wise

Figure 3.2.1 illustrates the trendline showcasing the average carrier delay and the percentage contribution of carrier delays across various regions throughout the United States. This visualization is further segmented into yearly and quarterly periods, facilitating straightforward analysis and interpretation

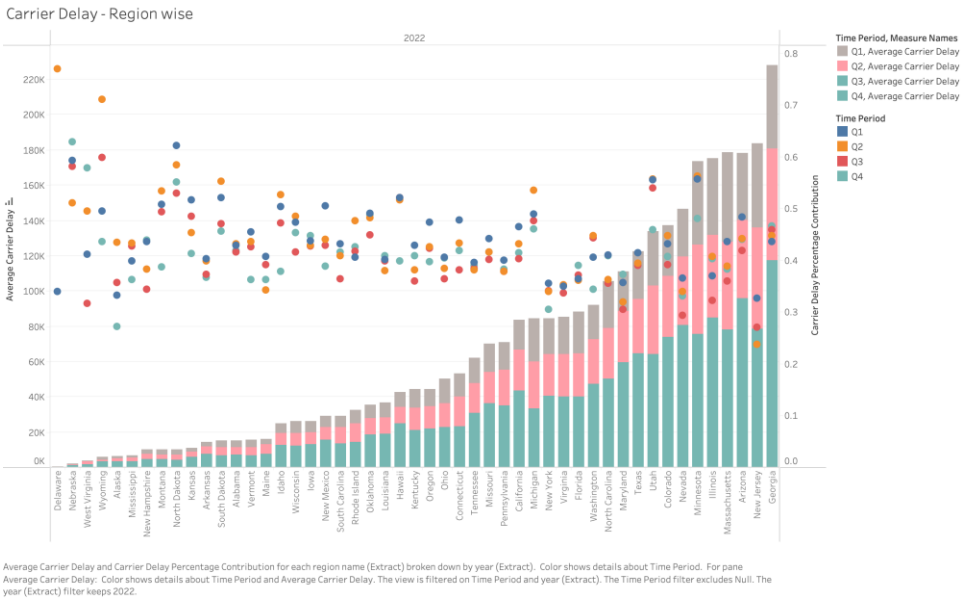


Fig 3.2.1 - Carrier Delay Contribution – Region Wise

The figure illustrates the relationship between the average duration of carrier delays measured in hours and the proportional contribution of carrier delays, presented as a percentage, across different regions. This analysis spans a four-quarter period within a year.

From this we can conclude that regions with carrier delay contribution (percentage) greater than 0.5 percentile should take measures to meet the airline regulatory conditions and cut down the arrival delay numbers.

Carrier Delay – Pareto Analysis:

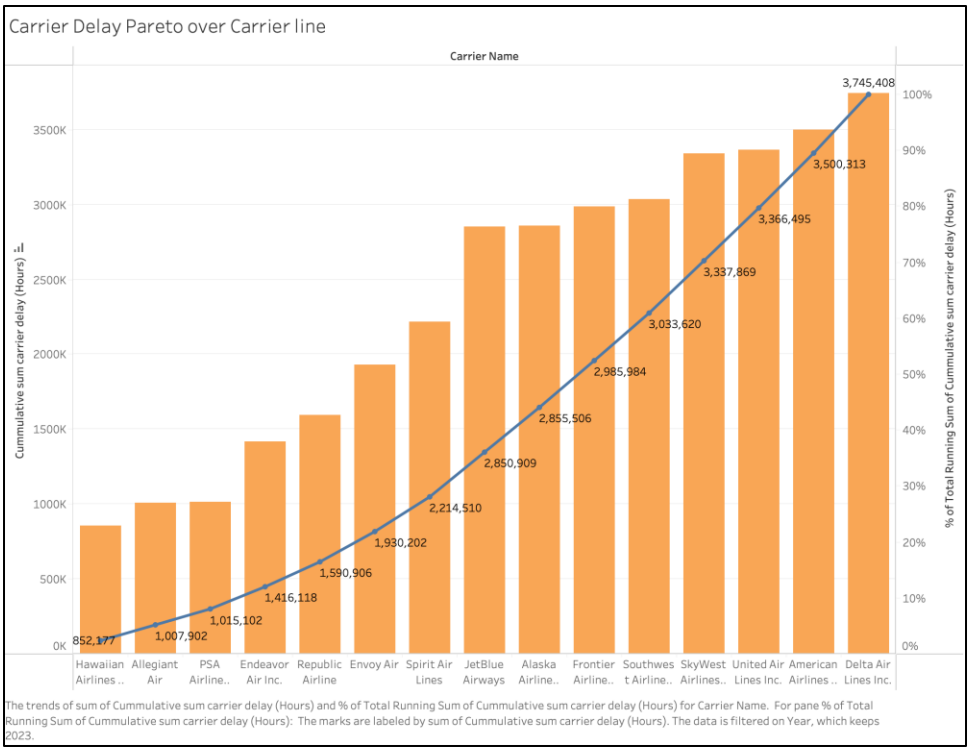


Fig 3.2.2 - Carrier Delat Pareto Analysis

Figure 3.2.2 showcases a Pareto analysis conducted using the 'Cumulative Sum of carrier delay,' a calculated metric that aggregates the total delay time (measured in hours) contributed by individual air carriers within the same year, and month.

The trends of sum of cumulative sum carrier delay (hours) and the percentage of Total running sum of cumulative sum carrier delay (hours) are plotted against the carrier's name available in the dataset. The data is filtered on Year, which keeps 2023

Carrier delay impact - Average fare and inflated average fare across regions

Fig 3.2.3 portrays the Carrier delay impact over average fare and average inflated fare across US regions

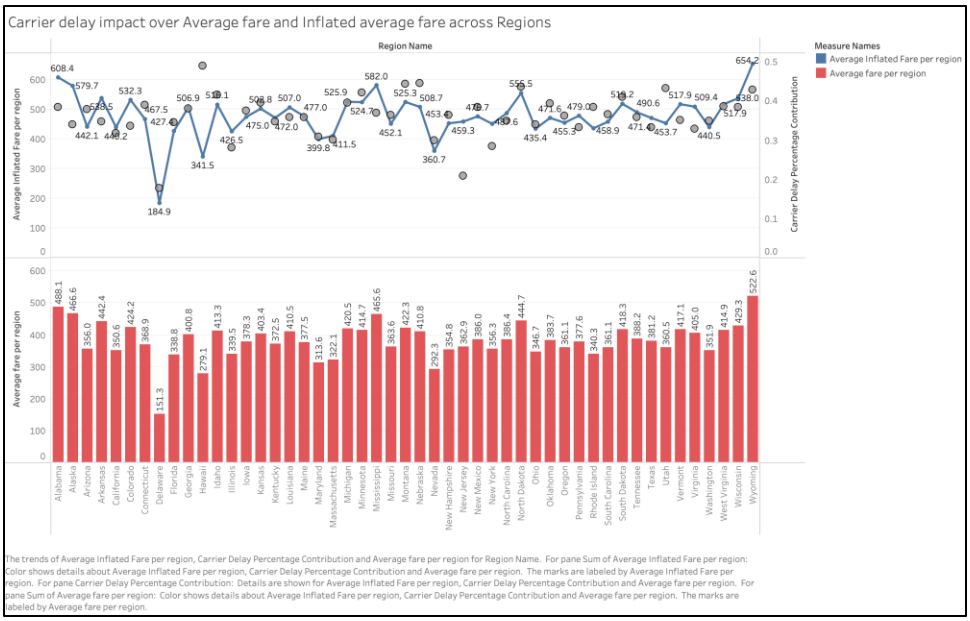


Fig 3.2.3 - Carrier delay impact average fare and inflated average fare

The observed plot strongly suggests a noticeable correlation between the calculated field, Carrier Delay Percentage Contribution, and the average inflated fare. The visual analysis indicates a close and evident relationship between these two variables.

Late Aircraft Delay

The dashboard examines the impact of late-arriving aircraft on subsequent flight delays. It's focused on understanding the flight delays concerning region, year, correlation with late aircraft count and month. The dashboard is structured as follows:

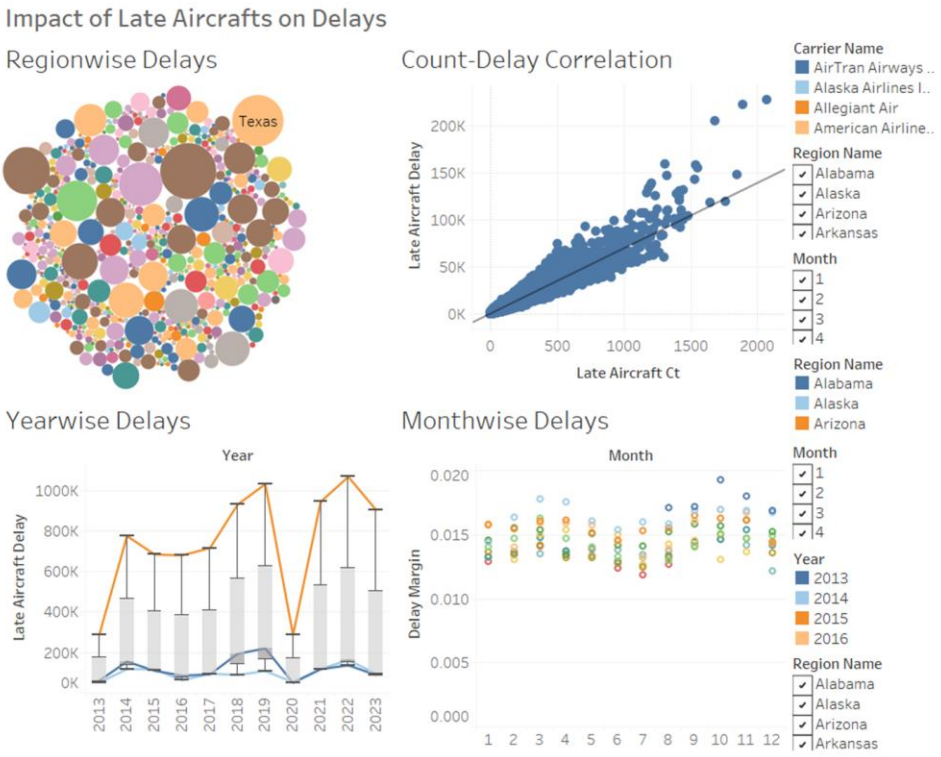


Figure 1: Overall Dashboard

It is divided into 4 visualizations:

1. **Top Left – Region-wise Delays**:
   This section looks at the number of delays in each region for different carriers. For each carrier, the color is different, the region is pickable using a filter and the size of the bubble corresponds to the number of flight delays.

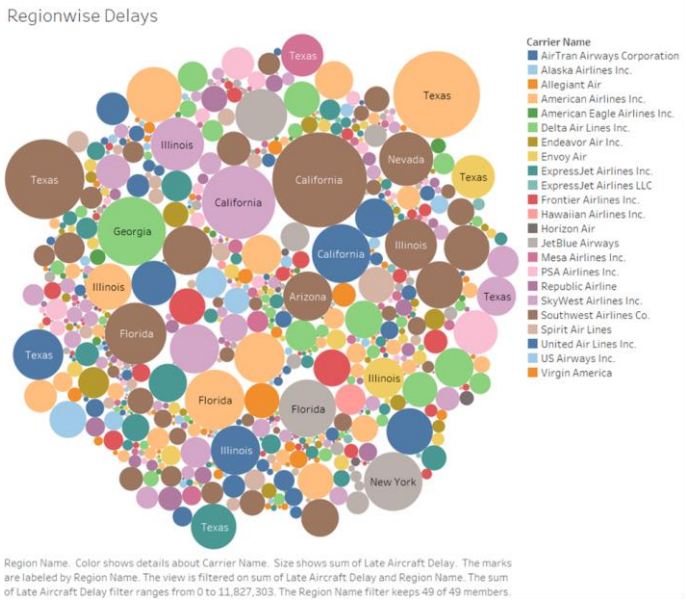


Figure 2: Region-wise Delays

2. **Bottom left – Year-wise Delays:**
   This section looks at the number of delays for each year. I have picked three regions: Alabama, Alaska and Arizona, and this can be edited in the worksheet. We can even filter to see the trend for some specific months. I have also added a box plot to conceptualize the statistics for each year.

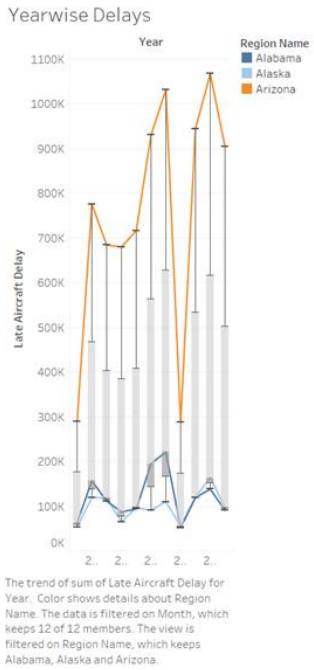


Figure 3: Year-wise Delays

3. **Top right – Count-Delay Correlation:**

This visualization showcased the relationship between the delays and the count of late aircraft. This depicts a clear positive correlation between the number of late arriving aircraft and the number of delays. So, when there are more late arriving aircraft, there are also more flight delays.

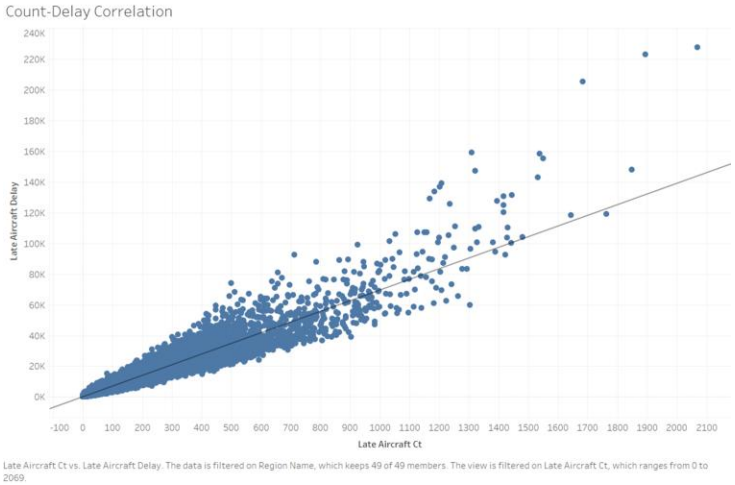


Figure 4: Count-Delay Correlation

4. **Bottom right – Month-wise Delays:**

This section looks at the delay margin (SUM([Late Aircraft Ct]) / SUM([Late Aircraft Delay]) in each month. It is also possible to filter different months and see the value for the corresponding year with the help of its colour.

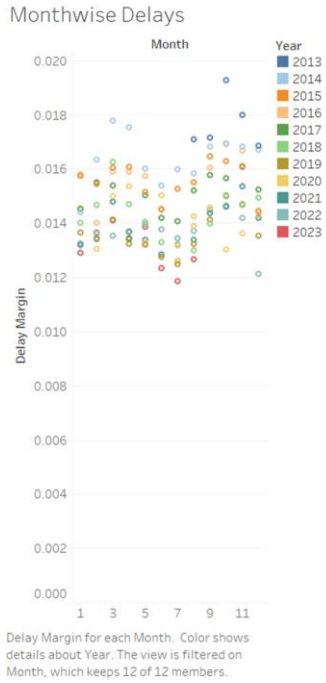


Figure 5: Month-wise Delays

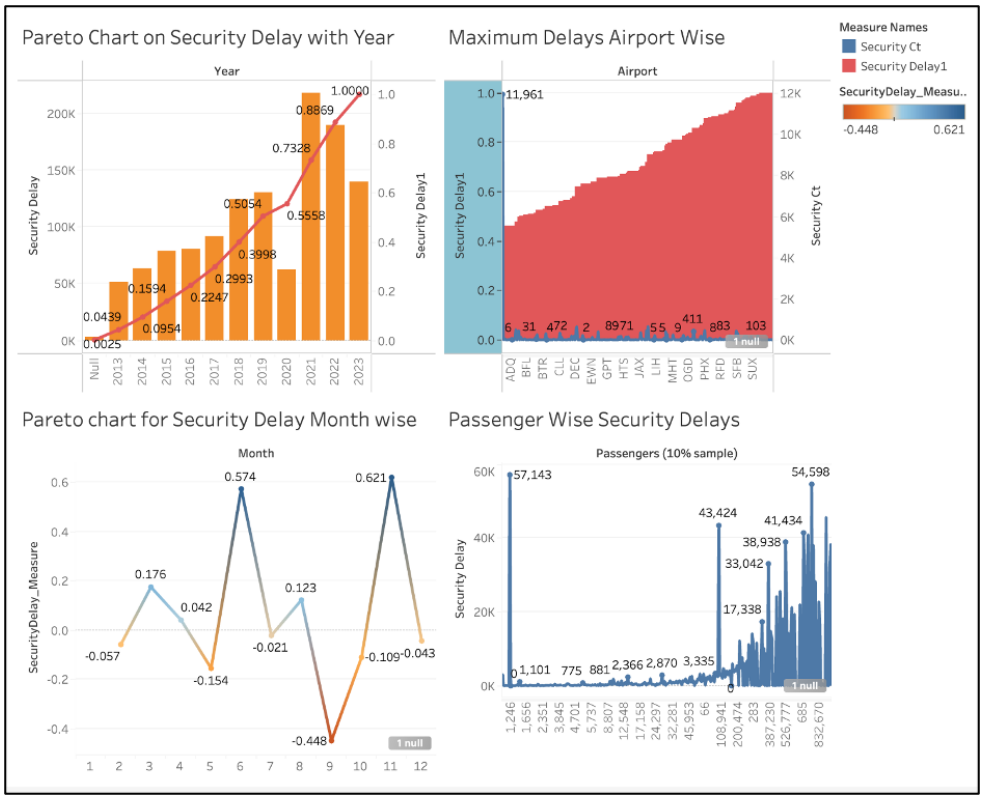
The findings of this dashboard are as follows:

- The regions with the most delays are Texas, Alabama, and Arizona.
- The year with the most delays is 2019.
- Late Aircraft Count and Delays are closely related.
- The month with the most delays is August.

To decrease such delays:

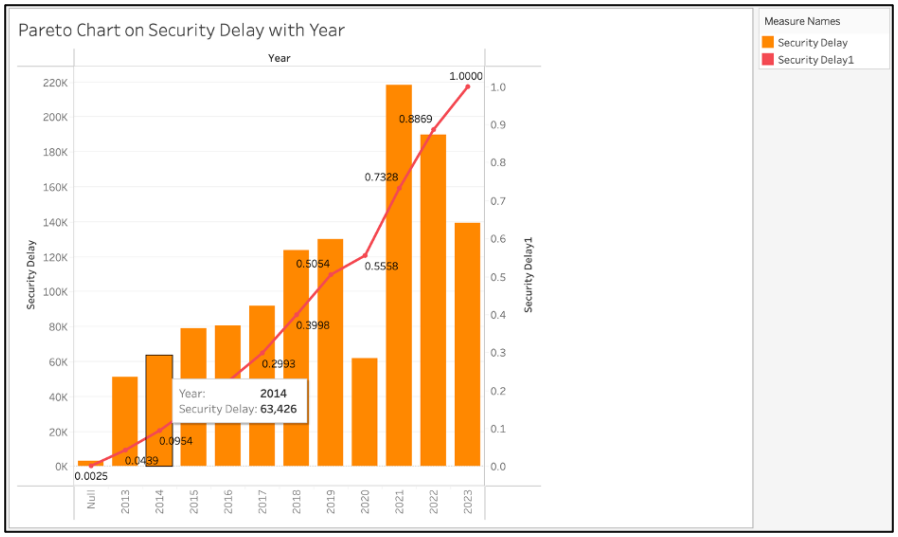
- Airlines should take sufficient steps to reduce the number of late-arriving aircraft by improving on-time performance, scheduling more buffer time between flights, and investing in new technology.
- Airports should take steps to improve the efficiency of their operations by adding more runways and improving traffic flow.

Security Delay



Airport security delays can happen for a variety of reasons, but they frequently stem from the strict security protocols put in place to guarantee passenger safety. The above dashboard consists of four visualizations that show security delays i.e. year wise, month wise, maximum delays in airports and passenger wise.

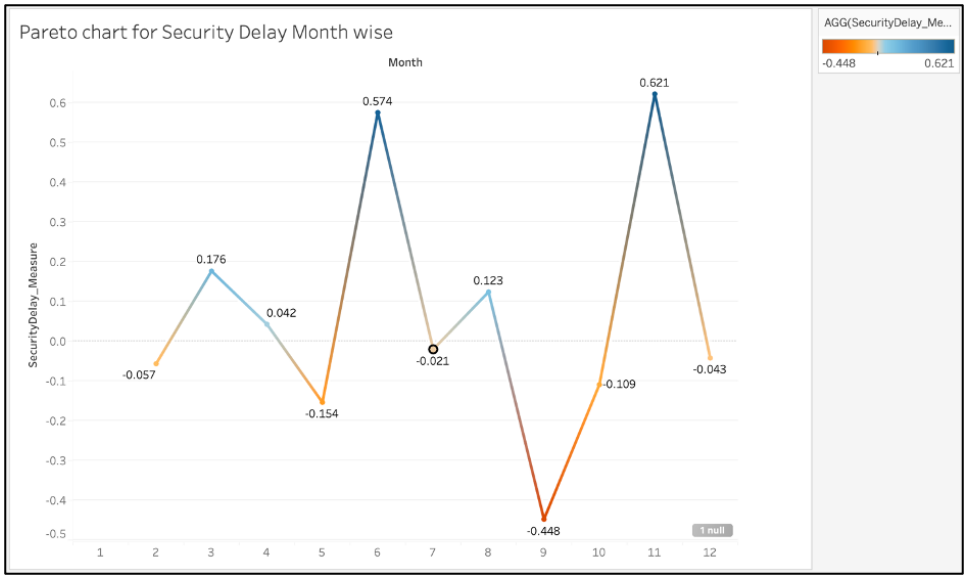
1. **<u>Top Left Chart (Pareto Chart on Security Delay with Year):</u>**



Visualizing the primary reasons for delays and determining the biggest contributors is necessary when creating a Pareto chart on security delays in an airline setting with years. The following instructions will help you create a Pareto chart for security delays over several years:

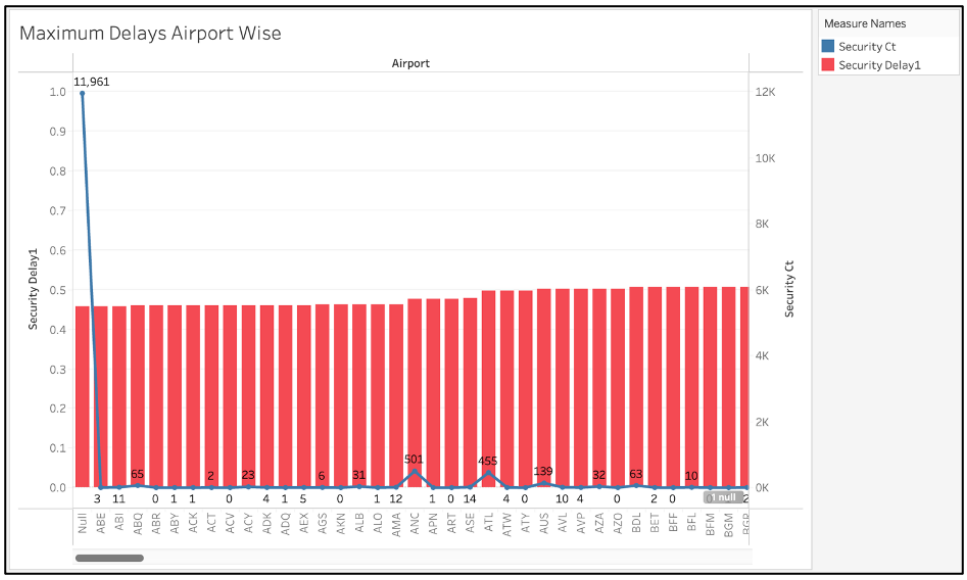
- Compile information on security delays every year, mentioning the reasons or elements that contribute to the delays. Make sure several reasons or causes are used to categorize the data.
- Using tableau tool, the derived steps:
    - a. X-Axis: Sort the causes by cumulative percentage and list them in descending order.
    - b. Plot the cumulative percentage for each reason on the Y-Axis (Bars).
    - d. Line Chart: To show the cumulative proportion of the total, overlay a line chart.
- To determine the most important reasons behind security delays each year, interpret the Pareto chart. Concentrate your efforts on resolving the issues that cause most delays.

2. Down Left Chart (Pareto Chart on Security Delay with Month):



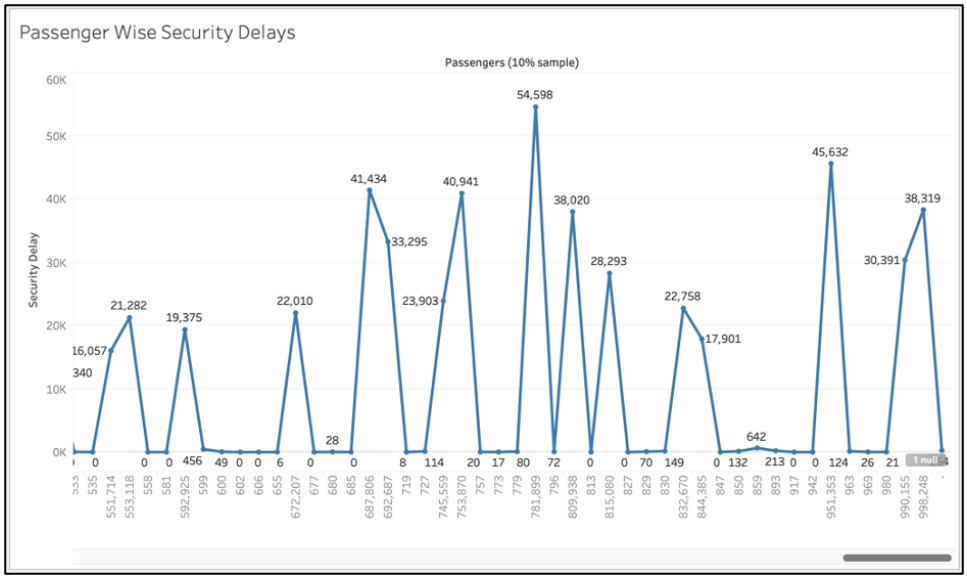
- In this visualization we have plotted month in x axis and security delays the calculated field in y axis
- We can see a line that shows the min and max values of security delays for each month.

3. Top Right Chart (Maximum Delays Airport wise)



- You would need access to data that includes the total number of delays at different airports in order to calculate the maximum delays at airports.
- Analysis:
    - Identify the main airports that cause delays.
    - Examine the possibility that operational problems, bad weather, or air traffic congestion are causing delays at these airports.
- We have considered airport data in the x axis and the security count in y axis with a line plot that contains sum of security delays.

4. <u>Down Right Chart (Passenger wise security Delays)</u>

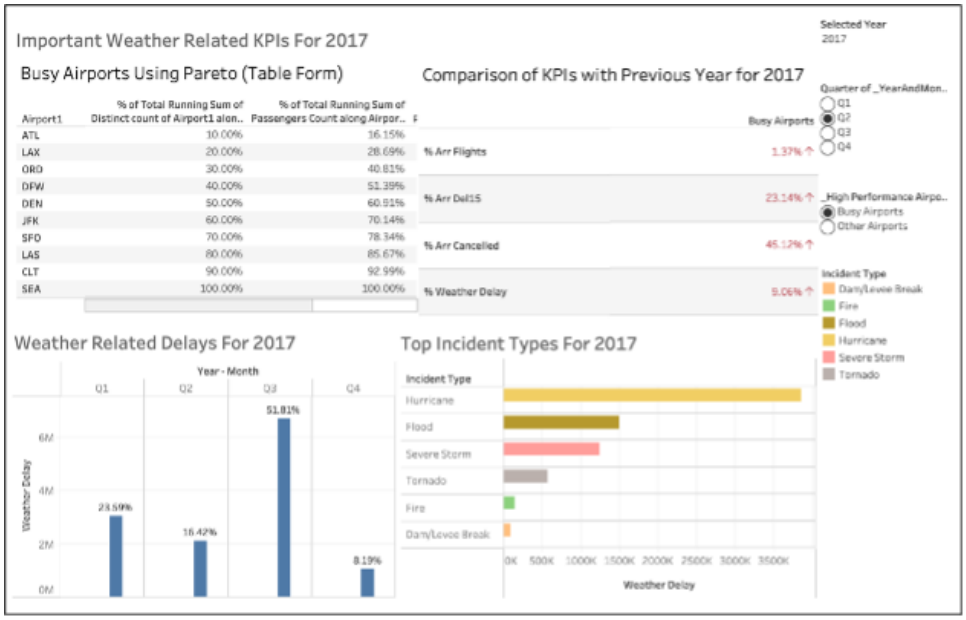


Passenger-level analysis of security delay data can shed light on the unique difficulties that each person has and assist in customizing security protocols or interventions to target specific problems.

We have taken Passengers up to 10% sample versus overall security delays. By doing this we can get a visualization that plots the data for passengers that experienced security delays.

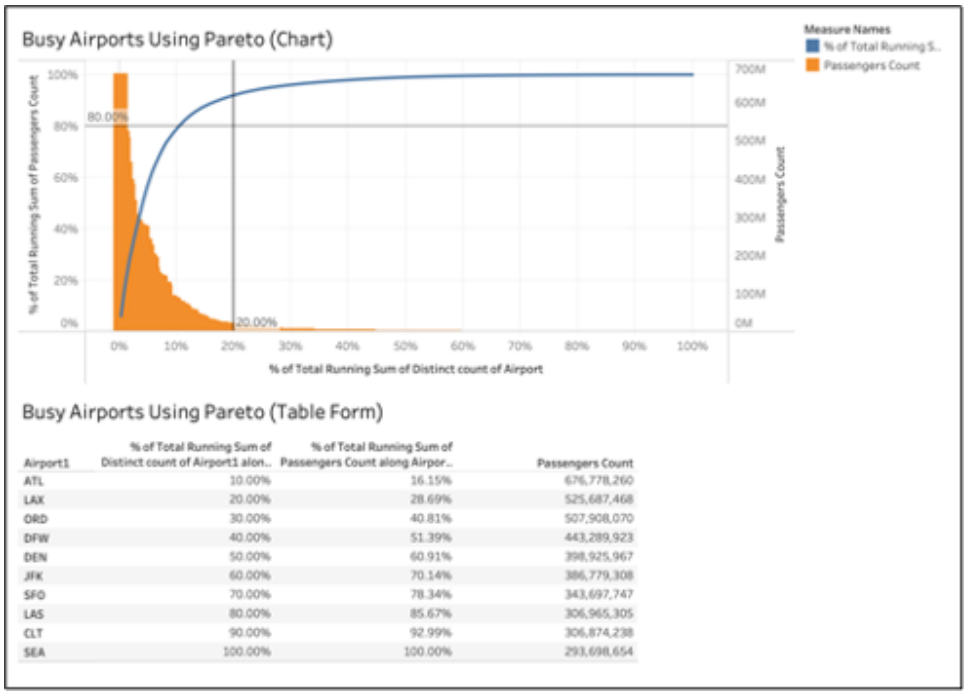
Weather Delay

In today's dynamic business environment, where uncertainties like weather-related delays can significantly impact operational efficiency, the importance of proactive decision-making cannot be overstated. Recognizing this, our company has developed a robust scorecard tailored for the visualization and meticulous monitoring of weather-related delays. This strategic tool allows for agile responses to emerging trends, empowering decision-makers with the insights needed to optimize operations and enhance overall performance.



Unveiling the Scorecard's Architecture:

1. Top Left (Busy Airports Using Pareto)

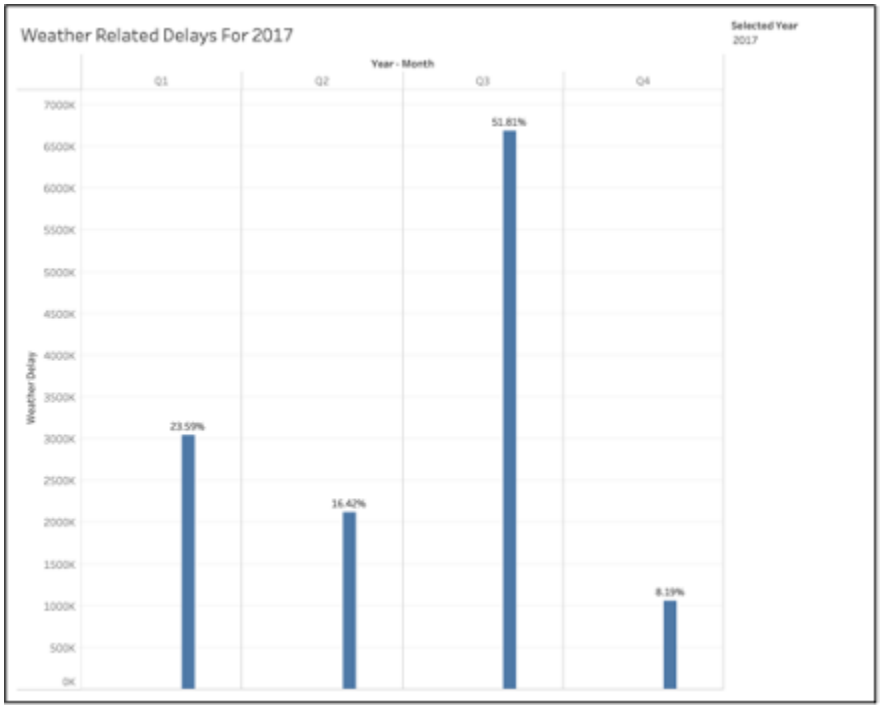


Busy Airports Using Pareto (Table Form)

Airport1	% of Total Running Sum of Distinct count of Airport1 alon..	% of Total Running Sum of Passengers Count along Airpor..	Passengers Count
ATL	10.00%	16.15%	676,778,260
LAX	20.00%	28.69%	525,687,468
ORD	30.00%	40.81%	507,908,070
DFW	40.00%	51.39%	443,289,923
DEN	50.00%	60.91%	398,925,967
JFK	60.00%	70.14%	386,779,308
SFO	70.00%	78.34%	343,697,747
LAS	80.00%	85.67%	306,965,305
CLT	90.00%	92.99%	306,874,238
SEA	100.00%	100.00%	293,698,654

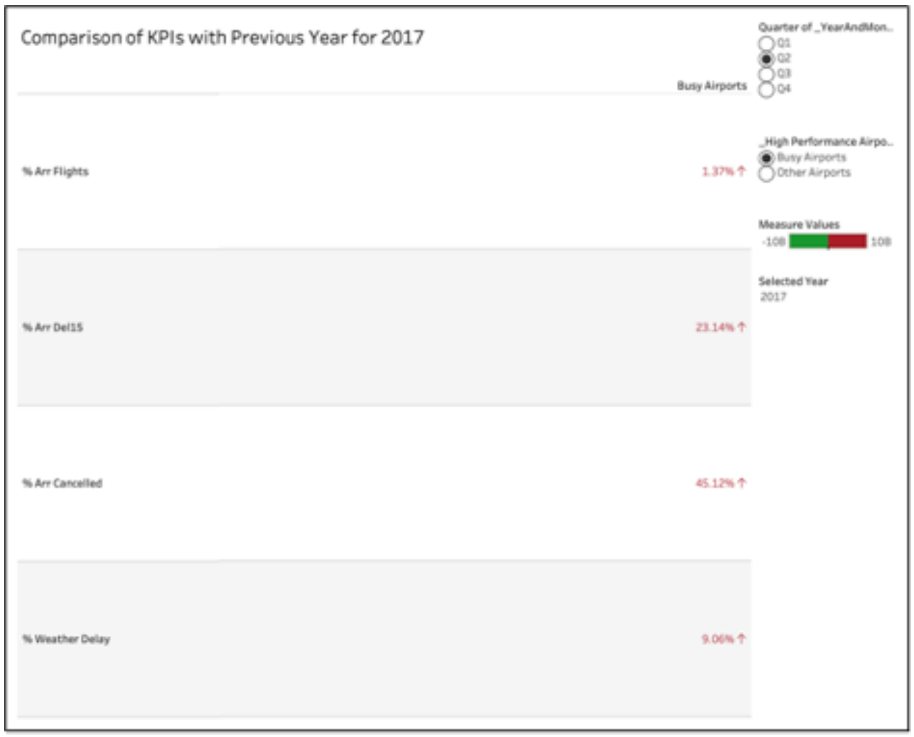
Rule-of-thumb figures, including overall on-time performance, total arrival flights, delayed flights, and cancellations, are conveniently situated at the top right corner. This central location facilitates quick and comprehensive insights into the overall operational situation, streamlining decision-making for executives.

2. Bottom Left (Weather Related Delays for Year X)



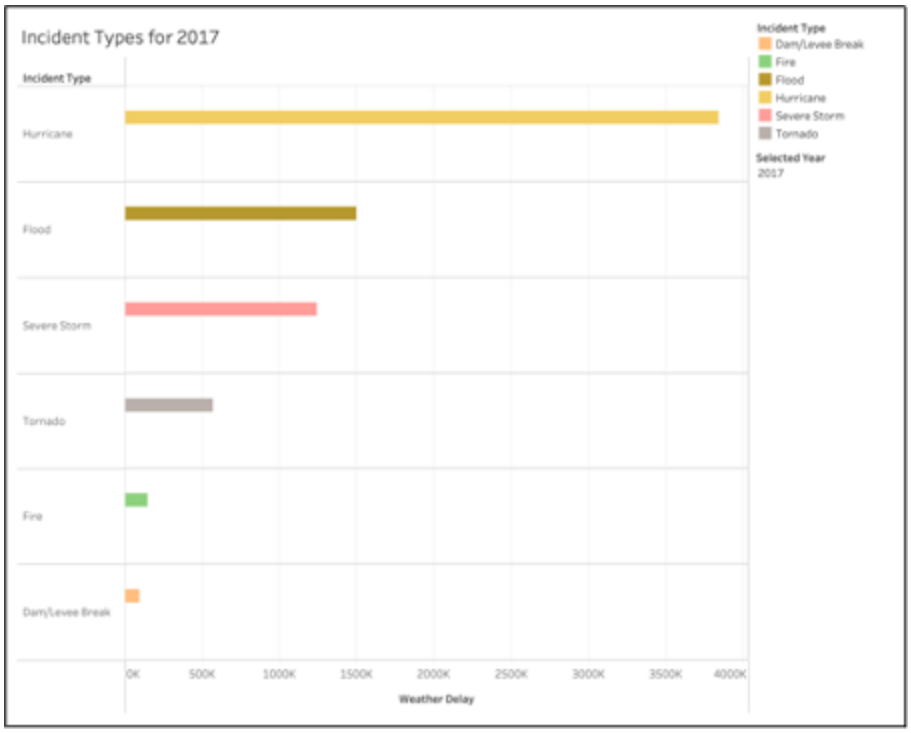
At the bottom left, a bar chart visualizes the percentage of weather delays by quarters of the year. Given the influence of seasonal factors on weather-related disruptions, this chart serves as a crucial tool for monitoring and identifying abnormal variations in this key performance indicator. It provides decision-makers with a timely overview of potential challenges arising from changing weather patterns.

3. Top Right (Comparison of KPIs with Previous Year for Year X)



The top right text table displays the year-by-year change in key KPIs, offering a historical perspective. This text table allows readers to comprehend both the delays distribution and the overall trend, aiding in strategic planning and forecasting. Understanding the annual changes provides valuable context for evaluating the effectiveness of on-time performance.

4. Bottom Right (Top Incident Types for Year X)



Delving deeper, the bottom right chart investigates the major causes of weather-related delays. This in-depth analysis provides readers with valuable insights, enabling them to pinpoint root causes. Armed with this knowledge, efforts can be strategically directed towards mitigating specific issues, contributing to more effective delay management.

4. Recommendations

Operate in Smaller Airport nearby Big Airports

Through Pareto analysis of the busiest airports and a correlation study between passenger numbers and total delays, we've identified *a direct proportionality between total delays and passenger volume*. This correlation is further underscored by our in-depth examination of security delays, revealing an exponential positive relationship with the number of passengers and total security delay.

Notably, delays are predominantly concentrated at specific airports rather than being regionally widespread. In response to these findings, our strategic approach involves avoiding major airports with the highest passenger traffic. Instead, *we choose to operate in smaller airports situated in close proximity to these major hubs.* This approach is designed to alleviate delays associated with larger airports while maintaining an optimized coverage of passengers.

From a financial perspective, our strategy is crafted to minimize the costs incurred from handling flight delays and, simultaneously, maximize revenue generated from potential customers.

Focus resource on Late Aircraft Delays

Following an analysis of delay types and an in-depth examination of Late Aircraft Delays, it's evident that Late Aircraft Delay significantly contributes to Total Delays, exhibiting an increasing trend (excluding 2020, impacted by COVID, and 2023, which is incomplete). Further exploration reveals a seasonal pattern, with improved performance in December & January and June & July, but heightened delays in March, April, and October.

To enhance our airline's on-time performance and surpass competitors, we will strategically allocate resources to minimize Late Aircraft Delays during the challenging months of March, April, and October. While acknowledging the myriad causes contributing to these delays, we will specifically address factors such as fleet availability and infrastructure deficiencies in our investigation of the processes leading to Late Aircraft Delays.

5. Conclusion

Following extensive analysis of flight data spanning from 2013 to 2023, insights into the intricate relationship among the number of passengers, airports, regions, and total flight delays have been unearthed. Additionally, a detailed examination has brought to light the significant contributions of various types of delays to the overall total.

Consequently, we have formulated two fundamental strategies for our company's operations: firstly, **to concentrate on smaller airports in close proximity to major hubs**, and secondly, **to channel resources towards addressing and minimizing Late Aircraft Delays**. By adhering to these strategic pillars, we aim to propel our airline towards robust profitability and enhance our competitive stance against other major players in the market.

---

i