TakeOff Timing: Dissecting Airline Delay Data

Rutwik Sanjay Guntoorkar

Rutgers Business School, Newark

Table of Contents

Table of Figures

Introduction

In an era where the pace of life accelerates by the day, the efficiency of air travel has become a linchpin for global connectivity and economic growth. "TakeOff Timing: Dissecting Airline Delay Data" is an ambitious project that dives deep into the intricacies of airline delays, aiming to unravel the complex tapestry of factors contributing to flight tardiness. This project stands at the intersection of data analytics and aviation management, where every second counts. Understanding the minutiae can lead to monumental improvements in travel planning and operational efficiency.

Imagine a dynamic dashboard that doesn't just display data but tells the story of millions of journeys - it's a narrative woven from the threads of time, weather, and the symphony of logistics that is air travel. At the heart of this endeavor lies a robust database that supports a rich analysis of over two decades of airline delay data sourced from the authoritative Bureau of Transportation Statistics of the U.S. Department of Transportation. The database will harness the power of SQL to distill vast amounts of data into actionable insights, allowing for a granular examination of delay causes, patterns, and trends.

This project's vision is to equip stakeholders, from passengers to policymakers, with the foresight to navigate the often unpredictable skies. By sifting through data from June 2003 to June 2023, the dashboard will reveal when and where delays occur and why. It will elucidate the intricate dance between scheduled timings and actual arrivals, dissecting the myriad reasons that turn timely takeoffs into delayed departures. Whether it's the logistical labyrinth within an airline's control, the wrath of nature unleashing extreme weather, or the cascading effects of the National Aviation System - every variable is scrutinized.

The project's ingenuity lies in its commitment to present data and contextualize it within the larger aviation narrative. It will explore the temporal ebbs and flows of flight punctuality, identifying top delay causes and unearthing seasonal trends that have, until now, remained hidden beneath raw figures and spreadsheets. With Tableau at the helm of visualization, the data will be transformed into a vivid tableau that tells a tale of time – time saved, lost, and managed, across the skies of the United States.

"TakeOff Timing" promises to be more than a tool; it's poised to be a compass for navigating the complexities of air travel. By interrogating variables like arrival delay, air carrier efficiency, and security-related interruptions, it will comprehensively understand the factors that lead to delays. Moreover, the project is not just backward-looking but forward-facing, offering insights that could potentially shape the future of airline delay mitigation strategies.

In a broader sense, this project is a testament to the transformative power of data analytics in contemporary decision-making processes. It encapsulates the journey from data to decisions, charting a course for more informed and efficient air travel. As we embark on this analytical expedition, "TakeOff Timing" will elevate our understanding of airline delays from a mere statistic to a strategic asset, enabling a future where time, in the realm of air travel, is not lost but optimized.

Objectives

The "TakeOff Timing: Dissecting Airline Delay Data" project aims to harness the power of data analytics to illuminate the intricacies of airline delays and flight arrival timings. This project endeavors to construct a comprehensive dashboard by leveraging a vast dataset spanning two decades (June 2003 to June 2023) from the Bureau of Transportation Statistics. This tool will visualize flight punctuality and latency and delve into the causative factors of such delays.

The Project will answer the following questions:

1. Flight arrival performance over the years?
2. What is the top cause for delays over the years?
3. Do the delays follow a seasonal trend?

The **variables of interest** are as follows: -
Arrival Delay: Difference in minutes between scheduled and actual arrival time. Early arrivals show negative numbers.
Air Carrier: The cause of the cancellation or delay was due to circumstances within the airline's control (e.g., maintenance or crew problems, aircraft cleaning, baggage loading, fueling, etc.).
Extreme Weather: Significant meteorological conditions (actual or forecasted) that, in the carrier's judgment, delay or prevent the operation of a flight, such as a tornado, blizzard, or hurricane.
National Aviation System (NAS): Delays and cancellations attributable to the national aviation system that refers to a broad set of conditions, such as non-extreme weather conditions, airport operations, heavy traffic volume, and air traffic control.
Late-arriving aircraft: A previous flight with the same aircraft arrived late, causing the present flight to depart late.
Security: Delays or cancellations caused by evacuation of a terminal or concourse, re-boarding of aircraft because of a security breach, inoperative screening equipment, and or long lines over 29 minutes at screening areas.

Methodology

The project "TakeOff Timing: Dissecting Airline Delay Data" methodology involves several strategic steps to analyze flight arrival and delay data over a 20-year period from June 2003 to June 2023. The project will use advanced data analytics techniques to develop a dynamic dashboard that offers comprehensive insights into flight timings, delay patterns, and their underlying causes.

1. The project will utilize Databricks to host and conduct SQL queries on a comprehensive dataset from the Bureau of Transportation Statistics. This setup will be the core system for data management, enabling intricate data filtering and compilation in real-time as users engage with the dashboard. The use of Databricks not only boosts the dashboard's responsiveness but also supports its integration with Tableau. This synergy ensures a live, interactive visual representation of data, enhancing user experience and facilitating shared access via online publishing.

2. The visualization component of the project will be executed using Tableau, a leading tool in the industry known for its powerful and user-friendly interface. Tableau will be employed to craft interactive dashboards and stories that succinctly convey the data analysis findings. Through a careful design process, these visualizations will be tailored to intuitively answer critical questions about flight arrival performance, the predominant causes of delays over the years, and any discernible seasonal trends.

3. The project's analytical approach will dissect key variables of interest, including arrival delays, categorized by minute differences from the scheduled time; causes of delays within the control of air carriers; impacts of extreme weather conditions; operational challenges within the National Aviation System (NAS); repercussions of late-arriving aircraft; and security-related delays. Each of these variables will be scrutinized to understand their contribution to overall flight delays.

4. To ensure a thorough analysis, the methodology also encompasses examining temporal trends to identify whether delays are influenced by seasonality, an aspect that can guide airlines and passengers in anticipation and management of potential disruptions. By addressing these focal points, "TakeOff Timing" aims to produce actionable insights that can inform decision-making processes for stakeholders in the aviation industry, offering a significant contribution to enhancing on-time performance and overall customer satisfaction.

Understanding the data

We have collected the data from the Bureau of Transportation Statistics (BTS) under the U.S. Department of Transportation, which provides extensive data on transportation and travel. The website hosts data on the causes of flight delays, as collected since June 2003. This data includes detailed statistics on flight delays, their causes, and other related information that would be valuable for analysis.

The data we have extracted is in raw format and can be defined as follows:

Data Field	Definition	Data Type	Possible Values / Range	Example / Notes
year	The calendar year when the flight operation took place.	Integer	2003–2023	2023
month	The calendar month when the flight operation took place.	Integer	1–12	6 (June)
carrier	A two-letter or three-digit code representing the airline carrier as per IATA and/or ICAO airline codes.	String	AA, DL, UA, etc.	9E (Endeavor Air Inc.)
carrier_name	The full name of the airline carrier.	String	Varies	Endeavor Air Inc.
airport	A three-letter code representing the airport of operation according to IATA airport codes.	String	ATL, LAX, ORD, etc.	ABE (Lehigh Valley International Airport)
airport_name	The full name of the airport, typically including the city or region served.	String	Varies	Allentown/Bethlehem/Easton, PA: Lehigh Valley International
arr_flights	The total number of flights that have arrived.	Float	>= 0	86
arr_del15	The number of flight operations delayed by more than 15 minutes.	Float	>= 0	15
carrier_ct	Delays attributed to the carrier, including maintenance, crew issues, etc.	Float	>= 0	6.03
weather_ct	Delays caused by significant meteorological conditions.	Float	>= 0	2.23
nas_ct	Delays and cancellations due to factors controlled by the National Aviation System.	Float	>= 0	May include non-extreme weather, airport operations, etc.
security_ct	Delays or cancellations caused by security issues such as terminal evacuations or breaches.	Float	>= 0	0
late_aircraft_ct	Delays due to a previous flight with the same aircraft arriving late.	Float	>= 0	2.63
arr_cancelled	The total number of cancelled flights.	Float	>= 0	0

arr_diverted	The total number of flights that were diverted from their scheduled destination.	Float	>= 0	0
arr_delay	The aggregate amount of time (in minutes) that arriving flights were delayed.	Float	>= 0	793
carrier_delay	The total amount of time (in minutes) that delays attributed to the carrier amounted to.	Float	>= 0	278
weather_delay	The total amount of time (in minutes) that delays caused by weather amounted to.	Float	>= 0	141
nas_delay	The total amount of time (in minutes) that delays associated with the National Aviation System amounted to.	Float	>= 0	151
security_delay	The total amount of time (in minutes) that delays caused by security issues amounted to.	Float	>= 0	0
late_aircraft_delay	The total amount of time (in minutes) that delays due to late-arriving aircraft amounted to.	Float	>= 0	223

For analytical purposes, we had to create multiple calculated fields in Tableau, such as the following, since the data collected was aggregated data for the specific month, year, airport, and carrier. We created aggregated percentage fields.

Calculated Field	Definition	Formula/Calculation	Purpose/Use
State	The U.S. state where the airport is located.	Extracted from airport_name or another relevant field.	Grouping data by state for regional analysis.
City	The city where the airport is located.	Extracted from airport_name or another relevant field.	Grouping data by city for localized analysis.
% Carrier Delay Count	Proportion of delays caused by the carrier to total delays.	(carrier_ct / arr_del15) * 100	To analyze the impact of carrier-related issues on overall delays.
% Delays caused by Carrier	Proportion of flights delayed due to the carrier.	(carrier_delay / arr_delay) * 100	To understand the extent to which carrier issues contribute to total delay time.
% Weather Delay Count	Proportion of weather-related delays to total delays.	(weather_ct / arr_del15) * 100	To quantify the impact of weather on flight punctuality.
% Delays caused by Weather	Proportion of flights delayed due to weather.	(weather_delay / arr_delay) * 100	To assess weather as a factor in flight delays.
% NAS Delay Count	Proportion of NAS-related delays to total delays.	(nas_ct / arr_del15) * 100	To measure the effect of NAS on delays.

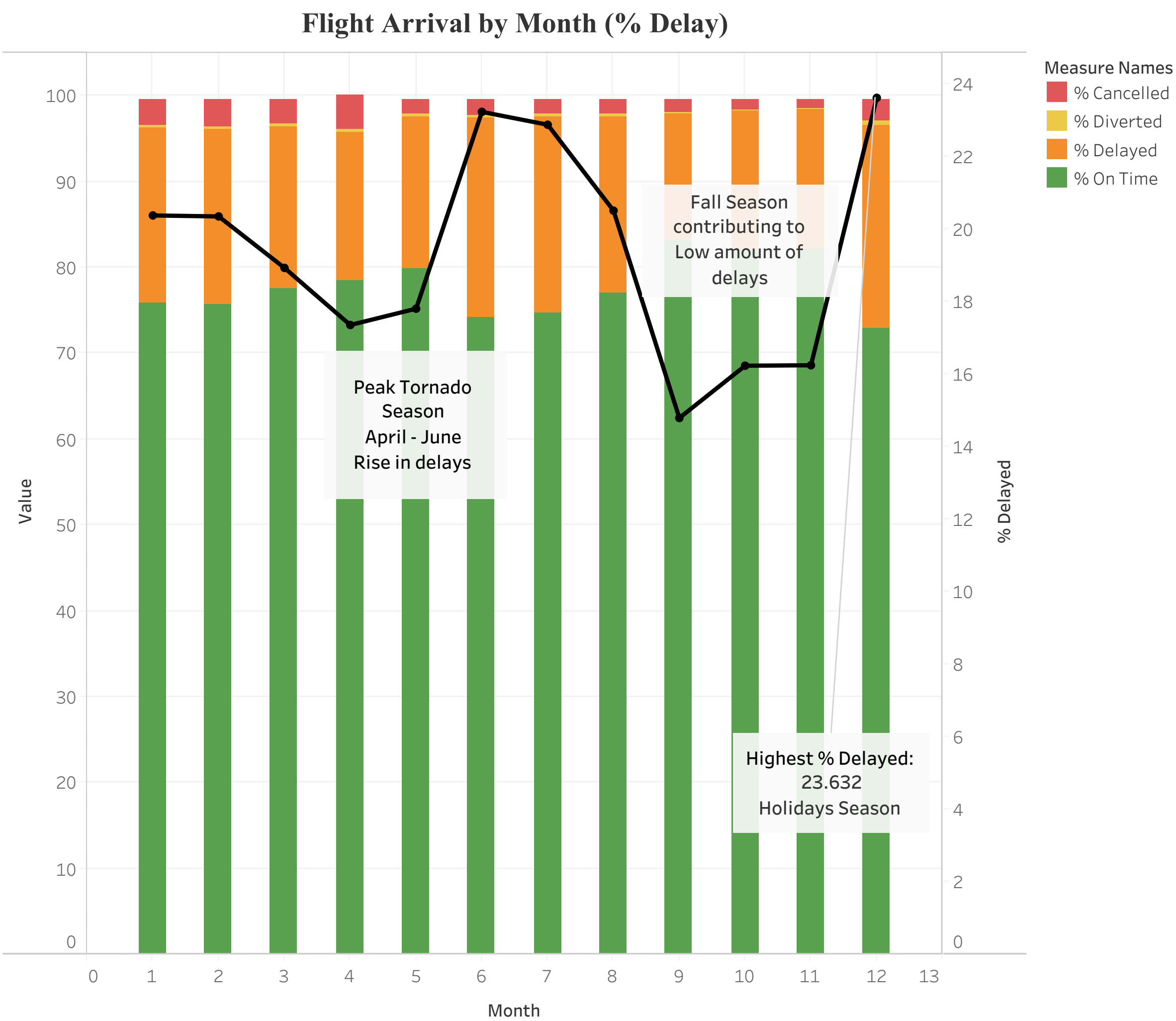
% Delays caused by NAS	Proportion of flights delayed due to NAS.	(nas_delay / arr_delay) * 100	To evaluate the significance of NAS operations on delay durations.
% Security Delay Count	Proportion of security-related delays to total delays.	(security_ct / arr_del15) * 100	To gauge the influence of security issues on flight delays.
% Delays caused by Security	Proportion of flights delayed due to security issues.	(security_delay / arr_delay) * 100	To examine the role of security in overall delay time.
% Late Aircraft Delay Count	Proportion of delays due to late aircraft to total delays.	(late_aircraft_ct / arr_del15) * 100	To analyze how previous flight delays contribute to current delays.
% Delays caused by Late Aircrafts	Proportion of flights delayed due to late aircraft.	(late_aircraft_delay / arr_delay) * 100	To understand the cumulative effect of late arrivals.
% Delayed	Proportion of delayed flights to total flights.	(arr_del15 / arr_flights) * 100	To identify the overall delay rate.
% On Time	Proportion of on-time flights to total flights.	(On Time Flights / arr_flights) * 100	To determine the punctuality rate.
On Time Flights	The number of flights that arrived on time.	arr_flights - arr_del15	To count flights that met scheduled arrival times.
% Diverted	Proportion of diverted flights to total flights.	(arr_diverted / arr_flights) * 100	To track the frequency of flight diversions.
% Cancelled	Proportion of cancelled flights to total flights.	(arr_cancelled / arr_flights) * 100	To monitor cancellation rates.
% Time Delay by Weather (In Minutes)	Proportion of total delay time caused by weather.	(weather_delay / arr_delay) * 100	To assess the impact of weather on delay durations.
% Time Delay by NAS (In Minutes)	Proportion of total delay time caused by NAS.	(nas_delay / arr_delay) * 100	To evaluate NAS's contribution to delay time.
% Time Delay by Security (In Minutes)	Proportion of total delay time caused by security.	(security_delay / arr_delay) * 100	To analyze the part of security delays in total delay time.
% Time Delay by Late Aircraft (In Minutes)	Proportion of total delay time caused by late aircraft.	(late_aircraft_delay / arr_delay) * 100	To determine the effect of late arrivals on overall delay time.
% Time Delay by Carrier (In Minutes)	Proportion of total delay time caused by the carrier.	`(carrier_delay / arr_delay)	

Data Source Link: https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp

Analysis

We have created multiple sheets and charts to create dashboards, which will help us analyze the delay caused by the airline operations.

Flight Arrival by Month (% Delay):



The trends of % Cancelled, % Delayed, % Diverted, % On Time and % Delayed for Month. For pane Measure Values: Color shows details about % Cancelled, % Delayed, % Diverted and % On Time. The data is filtered on State, Carrier Name and Year. The State filter keeps 54 of 54 members. The Carrier Name filter keeps 33 of 33 members. The Year filter ranges from 2003 to 2023.

Figure 1: Flight Arrival by Month (% Delay)

The "Flight Arrival by Month (% Delay)" chart depicts the monthly breakdown of flight statuses over an extended period, likely from 2003 to 2023. Here's a detailed analysis based on the visible information:

- **Seasonal Variations in Delays:** The chart shows a clear pattern of seasonal variations, with higher delays in certain months, which can be attributed to weather-related impacts such as tornado season.
- **Holiday Season Impact:** The highest percentage of delays occurs during the holiday season, as annotated on the chart. This is likely due to increased travel demand and operational pressures.
- **Tornado Season Trends:** The peak tornado season, from April to June, is marked by a rise in delays, indicating the influence of severe weather conditions on flight punctuality.
- **Fall Season Observations:** The annotation indicates fewer delays during the fall season, possibly due to more stable weather conditions and less travel congestion than summer and holiday seasons.
- **Consistent On-Time Performance:** Despite fluctuations in delays, the on-time performance seems relatively consistent throughout the year, suggesting robust scheduling and operational management.
- **Cancellations and Diversions:** While not the primary focus of the chart, cancellations and diversions are present throughout the year with minimal variation, indicating that seasonal factors less influence these events.
- **Operational Implications:** Airlines and airports may need to allocate more resources during peak delay periods to mitigate the impact of seasonal weather and increased passenger volumes.
- **Strategic Planning for Peak Seasons:** The data could inform strategic planning, such as staffing levels, maintenance schedules, and customer service provisions, particularly during periods known for increased delays.

Flight Arrivals Performance - State-wise:

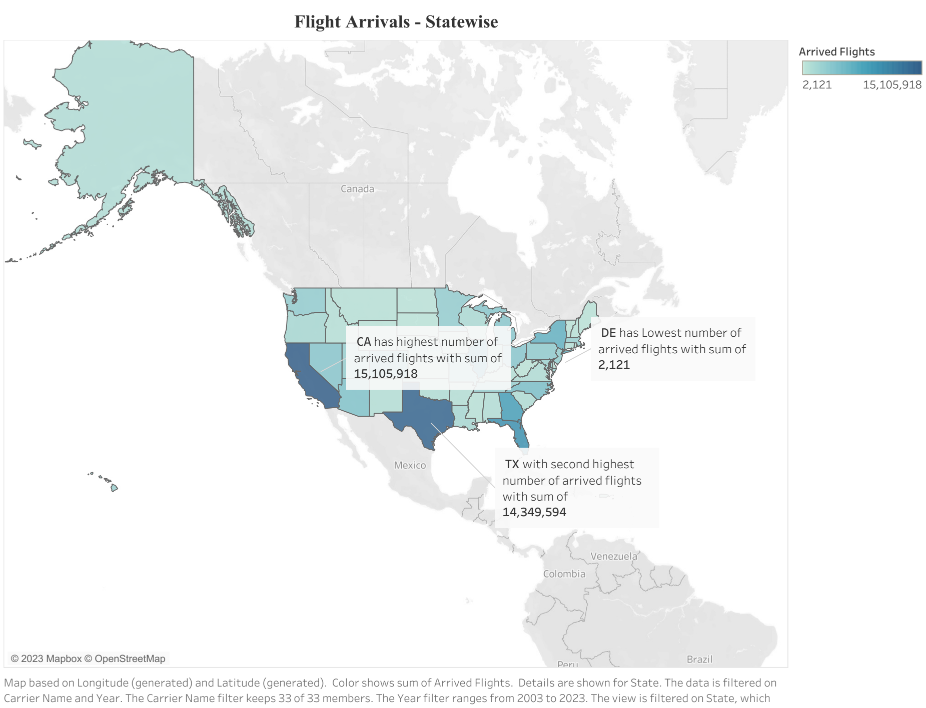


Figure 2: Flight Arrivals Performance - State-wise

The "Flight Arrivals Performance - Statewise" map provides a geographic visualization of flight arrival data across the United States. Here is an analysis and key observations based on the chart:

1. **Geographic Distribution of Flight Arrivals:** California (CA) has the highest number of flight arrivals, with Texas (TX) following closely behind, indicating these states are major transportation hubs.
2. **Lowest Arrival Numbers:** Delaware (DE) has the lowest number of flight arrivals, which may reflect its smaller size, lower population, or fewer airport facilities than other states.
3. **Regional Flight Traffic Disparities:** There's a clear disparity in flight traffic among the states, with coastal and larger states having higher numbers of arrivals.
4. **Impact on State Economies:** States with higher flight arrivals might experience more economic benefits from the airline industry due to tourism and business travel.
5. **Infrastructure and Capacity:** The high numbers in CA and TX suggest they have substantial airport infrastructure and capacity to handle large air traffic volumes.
6. **Potential for Market Expansion:** States with low arrival numbers might represent markets with potential for expansion or increased airline service offerings.
7. **Strategic Planning for Airlines:** Airlines can use this data for strategic planning, route development, and resource allocation to optimize flight schedules and meet demand.
8. **Consideration for Environmental Impact:** States with high air traffic might need to consider the environmental impact of flights and explore sustainability measures.

Delay Cause & Time Overview:

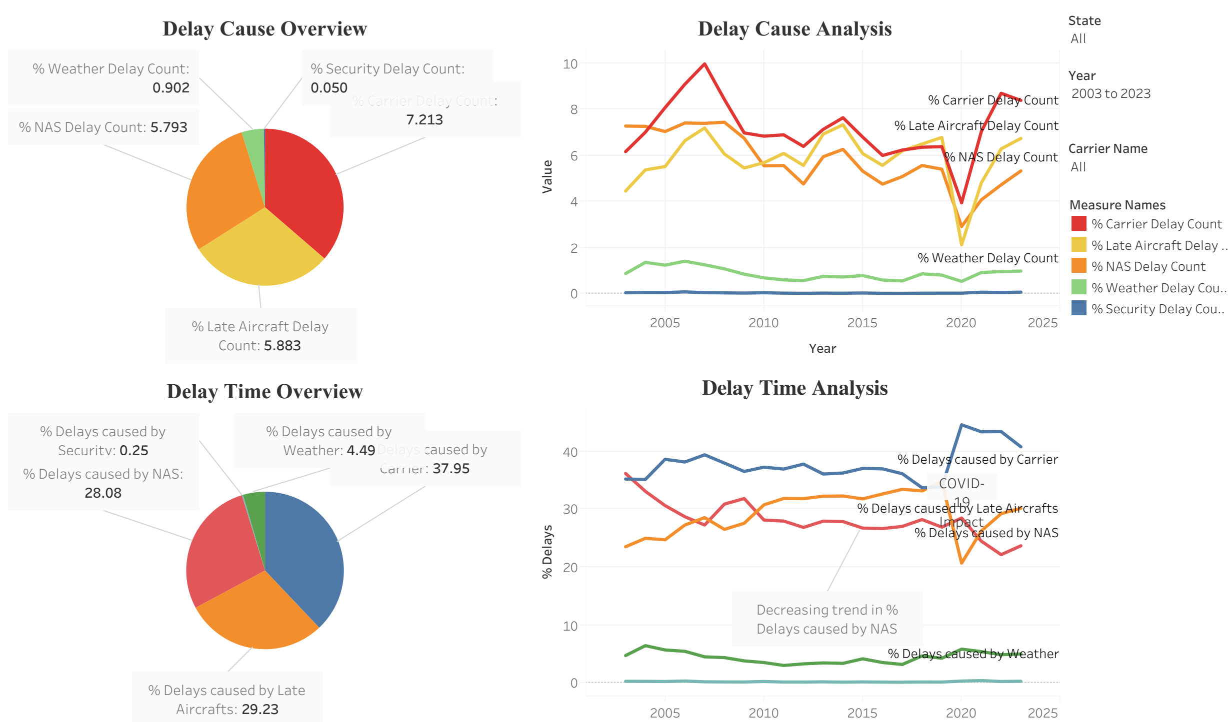


Figure 3: Delay Cause & Time Overview

The dashboard presents a multi-faceted view of airline delay causes and delay time distribution across a period from 2003 to 2023. Here's a detailed analysis of the given dashboard components:

Delay Cause Overview (Pie Chart - Upper Left)

The "Delay Cause Overview" pie chart illustrates the distribution of delay causes, represented as a percentage of total delays. Here are the key points derived from the analysis:

- **Carrier Delay (Red Segment):** The most significant portion of the pie chart, at 7.213%, indicates that carrier-related issues are the predominant cause of delays. This could include maintenance, crew scheduling, or mechanical problems that are within the airline's control to manage.
- **Late Aircraft Delay (Orange Segment):** Representing 5.883%, this is the second largest cause of delays. This usually occurs when a previous flight with the same aircraft is delayed, causing a domino effect on subsequent flights.
- **NAS Delay (Yellow Segment):** Making up 5.793%, NAS delays are significant. This category includes a variety of issues, such as air traffic control, airport operations, and other systemic delays that are not attributed to weather, security, or the airlines' control.
- **Weather Delay (Green Segment):** Accounting for 0.902%, weather delays are relatively minor compared to other factors. These delays occur due to adverse weather conditions like storms, fog, or extreme weather events that can affect flight schedules.
- **Security Delay (Purple Tiny Segment):** At 0.050%, the smallest segment represents delays caused by security issues. This might include delays due to TSA operations, security breaches, or other incidents requiring additional security measures.
- **Analysis of Proportions:** The chart shows that most delays are caused by factors that could be managed or mitigated by the airlines (carrier and late aircraft delays) or the aviation system (NAS delays), with weather and security playing smaller roles.
- **Focus Areas for Improvement:** The data suggests that airlines could potentially reduce overall delay figures by focusing on operational efficiency, particularly with respect to factors under their direct control.
- **Implications for Strategic Planning:** Understanding the proportion and causes of delays can help airlines and airports in strategic planning and resource allocation to target the most impactful areas for delay reduction.

Delay Time Overview (Pie Chart - Lower Left)

The "Delay Time Overview" pie chart shows the percentage of flight delays attributed to various causes. Here's an analysis of key points:

- **Carrier-Related Delays (Blue Segment):** The largest segment at 37.95% indicates that most delays are due to factors within the airline's control, such as maintenance or crew issues.
- **Late Aircraft Delays (Orange Segment):** Constituting 29.23%, this is the second-largest cause of delays, suggesting that many delays are due to the domino effect of delayed earlier flights.
- **NAS Delays (Red Segment):** Making up 28.08% of delays, this shows the impact of the National Aviation System on delay times, which could include air traffic control issues, airport operations, or volume of air traffic.
- **Weather Delays (Green Segment):** At 4.49%, weather is a less frequent cause of delays, which may be due to airlines' effective weather prediction and management.
- **Security Delays (Purple Tiny Slice):** The smallest slice at 0.25% indicates that security rarely causes significant flight delays.

- **Proportional Impact:** The chart clearly shows that carrier, late aircraft, and NAS are the most significant contributors to delays, suggesting these are the areas where improvements could have the most substantial impact on reducing overall delay times.
- **Minor Causes:** Both weather and security, while important for overall airline operations, contribute relatively little to the overall delay percentages, indicating that current systems to manage these factors are relatively effective.
- **Data Representation:** The pie chart effectively represents this type of categorical data, giving immediate insight into the relative importance of each cause of delay.

Delay Cause Analysis (Line Chart - Upper Right)

The "Delay Cause Analysis" chart provides a visual representation of the various causes of airline delays over a period of two decades, from 2003 to 2023. Here's a detailed analysis of several key points:

- **Carrier-Related Delays (Red Line):** The most significant peaks in carrier-related delays occurred around 2007, indicating a period of high operational challenges for airlines. Following this peak, there was a general declining trend until about 2012, after which the trend stabilized with slight fluctuations.
- **Late Aircraft Delays (Orange Line):** Late aircraft delays closely follow the pattern of carrier delays, which suggests a relationship between airline operational issues and subsequent flight delays. The peaks and valleys in this line indicate that improvements or deteriorations in carrier performance directly affect the punctuality of subsequent flights.
- **NAS Delays (Yellow Line):** There was a pronounced peak in NAS-related delays around 2007, which significantly declined afterward. This trend indicates that measures taken to improve the National Aviation System efficiency have positively impacted, although there is a minor resurgence around 2015.
- **Weather Delays (Green Line):** Weather delays show less variability than other factors and maintain a low level throughout the years. This steadiness implies that weather, while uncontrollable, has a consistently low impact on overall delay percentages.
- **Security Delays (Light Green Line):** Security delays remain the least common cause and show very low variability over the years, implying that security measures, while critical, have not been a significant source of delay in airline operations.
- **Interconnectedness of Delay Factors:** The chart shows that carriers, late aircraft, and NAS delays often move in tandem, which could indicate that they are interconnected. Improvements or deteriorations in one area could have cascading effects on the others.
- **Impact of External Events:** The sharp decline in all types of delays around the 2020 mark correlates with the onset of the COVID-19 pandemic, highlighting the significant impact of external events on airline operations. This period is marked by reduced flight volumes, which likely contributed to the decrease in delays.
- **Overall Trend:** While there have been improvements in delay management, particularly with NAS delays, the data suggests that airline and late aircraft delays remain challenging for the industry. Ongoing efforts to improve turnaround times and airline operations efficiency are essential to reduce these types of delays further.

Delay Time Analysis (Line Chart - Lower Right)

The "Delay Time Analysis" chart visualizes the percentage of flight delays by cause from 2002 to around 2023. Here are the key points derived from the analysis:

- **Carrier Delays (Blue Line):** This is the highest line on the chart, suggesting that carrier-related issues are consistently the primary cause of delays. The line trends downward after 2010, indicating some improvement in managing carrier-caused delays.
- **Late Aircraft Delays (Red Line):** The second leading cause of delays, these have decreased since around 2010, with some fluctuations. This line closely follows the trend of carrier delays, which implies a connection between airline operations and the subsequent impact on the aircraft's ability to maintain schedules.
- **NAS Delays (Orange Line):** NAS-related delays show a significant decreasing trend from around 2007, suggesting improvements in the system or operational changes that have positively affected delay rates.
- **Weather Delays (Green Line):** Weather-related delays remain relatively low and stable throughout the years, with slight peaks that likely correlate with seasonal weather patterns or extreme weather events.
- **Security Delays (Purple Line):** Security delays are the lowest and exhibit minimal variability over the period, indicating that security checks and procedures have a negligible impact on overall delay percentages.
- **COVID-19 Impact (Noted on the Chart):** There's a marked decrease in all types of delays around 2020, coinciding with the COVID-19 pandemic. This suggests the pandemic's significant impact on flight operations is likely due to reduced flight volumes and changes in airline operations.
- **Decreasing Trend in NAS Delays:** A notable observation is the continuous decline in NAS delays, hinting at effective management and operational strategies to mitigate these delays over the years.
- **Comparison of Trends:** Comparing the trends, carrier and late aircraft delays show a parallel pattern, while NAS, weather, and security delays seem to be less influenced by the factors affecting the former two.

Key Observations:

- Carrier and Late Aircraft as Major Delay Causes: Both the "Delay Cause Overview" and "Delay Time Analysis" charts show that a significant proportion of delays are attributed to carriers and late-arriving aircraft. This indicates operational inefficiencies within the airlines' control.
- NAS Delays Show Improvement: The "Delay Cause Analysis" and "Delay Time Analysis" charts reveal a decrease in NAS-related delays over time, suggesting improvements in air traffic management and airport operations.
- Weather and Security Delays are Minimal: All charts illustrate that weather and security are less frequent causes of delays, implying that current management strategies for these factors are relatively effective.
- Impact of COVID-19 Pandemic: A noticeable decline in delays across all categories is observed in the "Delay Time Analysis" chart around 2020, highlighting the impact of the pandemic, most likely due to reduced flight operations.
- Variability in Delay Causes Over Time: The "Delay Cause Analysis" chart shows that the relative impact of different delays causes changes over time, suggesting that airlines and airports must be adaptable in their management strategies.

Implications for Airlines and Airports:

- Operational Efficiency: There's a clear need for airlines to improve operational aspects like maintenance, crew management, and aircraft turnaround times to reduce carrier and late aircraft delays.

- Strategic Focus on NAS Improvements: While there have been improvements, focusing on NAS-related issues can further reduce delays, improving overall efficiency and passenger satisfaction.
- Contingency Planning for Weather: Despite weather delays being a smaller portion of the total, having robust contingency plans for adverse weather can prevent these from becoming a larger issue.
- Adaptability to External Factors: The COVID-19 pandemic showed the importance of being able to adapt operations in response to external impacts quickly. Airlines and airports should develop flexible operational plans to manage sudden changes in flight volumes.
- Data-Driven Decision Making: The insights from these visualizations can drive targeted strategies to mitigate delays, such as investing in technology and infrastructure that supports more efficient airline and airport operations.
- Customer Communication and Service: With a better understanding of delay causes, airlines and airports can improve communication with passengers regarding expected delays and provide better customer service during such events.
- Collaboration Between Stakeholders: The interconnected nature of delay causes suggests a need for collaborative efforts between airlines, airports, air traffic control, and other stakeholders to address the multifaceted challenges leading to delays.

The dashboard provides a clear representation of the data and allows for an at-a-glance understanding of trends and patterns in airline delays, which stakeholders can use to target improvements in operations and passenger experience.

Arrival & Delay Overview Dashboard:

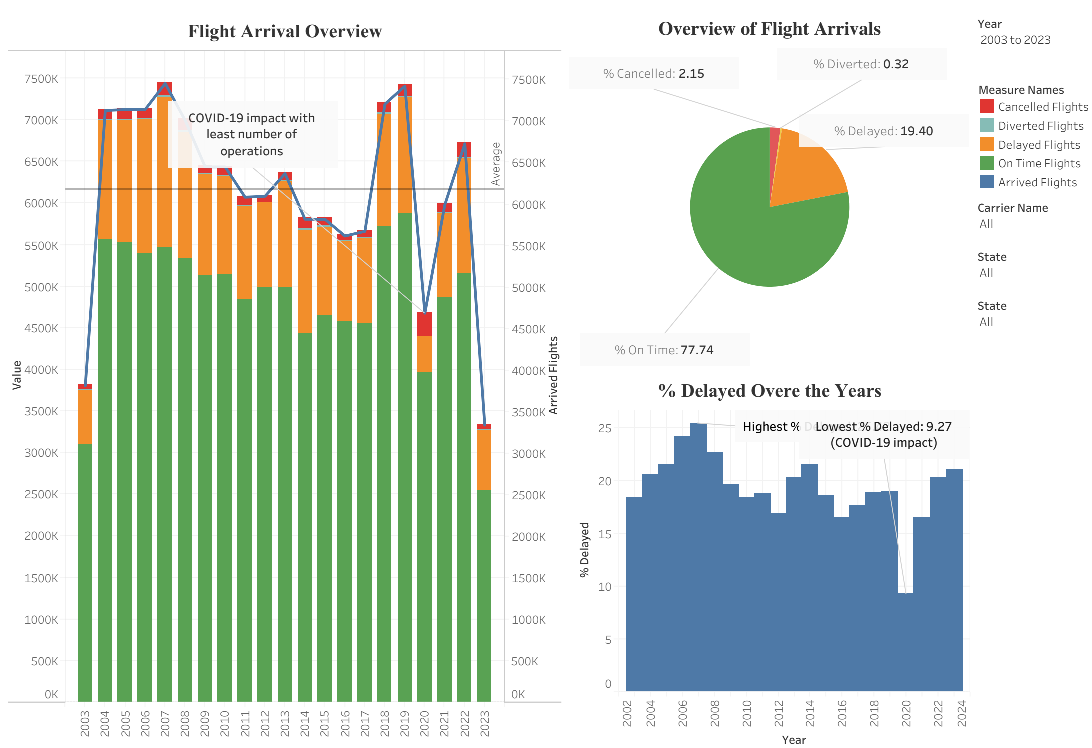


Figure 4: Arrival & Delay Overview Dashboard

The dashboard titled " Arrival & Delay Overview Dashboard" is segmented into three main parts: a bar and line chart titled "Flight Arrival Overview," a pie chart titled "Overview of Flight Arrivals," and a bar chart titled "% Delayed Over the Years." Here's a combined analysis of the presented data:

Flight Arrival Overview:

The "Flight Arrival Overview" chart presents the trends of flight statuses over the years, including arrived, delayed, canceled, diverted, and on-time flights. Here's the analysis in key points:

- **General Trend of Arrivals:** The total number of flights (represented by green bars) generally shows an increasing trend from 2003 to 2019, indicating growth in the airline industry.
- **COVID-19 Impact:** There's a sharp decline in the number of flights in 2020, marked on the chart, which is due to the COVID-19 pandemic. This significantly impacted, resulting in the lowest operational volume within the observed period.
- **Delayed Flights (Orange Bars):** The number of delayed flights seems to fluctuate yearly, with no clear upward or downward trend before the pandemic.
- **Cancelled Flights (Red Bars):** Cancelled flights show spikes in some years, but these do not appear to follow a consistent pattern over time.
- **Diverted Flights (Blue Bars):** Diverted flights are relatively low compared to other flight statuses and show minimal change over the years, suggesting that diversions are less common.
- **Resilience and Recovery:** After the initial drop in 2020, there is a visible recovery in the number of flights in 2021, though not to pre-pandemic levels. This suggests resilience in the industry and a degree of recovery.
- **On-Time Flights (Green Bars):** The chart might imply that on-time flights have maintained a relatively steady proportion of the total, though actual values are not discernible from this representation.
- **Implications for Planning:** The data emphasizes the need for robust contingency planning to manage large-scale disruptions, such as those caused by the COVID-19 pandemic.

Overview of Flight Arrivals:

The "Overview of Flight Arrivals" pie chart shows flight statuses over a specified period, likely from 2003 to 2023. Below is an analysis with key observations:

- **High On-Time Arrival Rate:** The largest portion of the chart is green, representing on-time flights, which make up 77.74% of the total. This indicates a strong performance in maintaining scheduled arrival times.
- **Delayed Flights Proportion:** A significant portion, 19.40%, is orange, indicating delayed flights. While not the majority, this is a considerable fraction that could impact passenger satisfaction and airline operations.
- **Low Cancellation and Diversion Rates:** The small slices for cancellations (2.15%) and diversions (0.32%) suggest that these events are relatively rare compared to delays.
- **Operational Efficiency:** The significant on-time arrival rate showcases operational efficiency but also suggests that there is room for improvement, especially in reducing delays.
- **Customer Service Focus:** Given that nearly one-fifth of flights are delayed, there is a potential need to enhance customer service and communication strategies to address passenger inconvenience.
- **Strategic Planning:** The data provides a basis for strategic planning and resource allocation, particularly in addressing the causes of flight delays.

- **Benchmarking for Industry Standards:** This chart could benchmark performance against industry standards or competitors, helping identify areas for competitive advantage or improvement.
- **Impact of External Factors:** While not explicitly shown on this chart, the knowledge of external factors such as weather or air traffic control impacts could be correlated with this data to plan for mitigation strategies.

<u>% Delayed Over the Years:</u>

The bar chart "% Delayed Over the Years" provides a historical perspective on the percentage of flights delayed from 2002 through the projected year of 2023. Here's the analysis of key points:

- **Trend of Delays:** There's a notable fluctuation in the percentage of delays over the years, with no consistent upward or downward trend until 2020.
- **Peak Delay Period:** The highest percentage of delays was recorded in 2007 at 25.42%, indicating a challenging year for the airline industry.
- **COVID-19 Impact:** A significant drop in delays occurred in 2020, marked as the "Lowest % Delayed: 9.27" due to the COVID-19 pandemic's impact, likely resulting from reduced flight volumes and less congested airspaces.
- **Post-Peak Variability:** After the peak in 2007, there's a general downward trend with some variability, suggesting that the industry may have taken steps to address the factors contributing to delays.
- **Recovery Post-Pandemic:** The years 2021 and 2022 show an increase in delays compared to 2020, possibly indicating a recovery in flight volumes as the pandemic's impact lessens.
- **Data for Strategic Decisions:** The trend data can inform airlines and airports about historical performance, enabling them to make strategic decisions to manage and reduce flight delays.
- **Benchmark for Improvement:** The lowest point in 2020 sets a new benchmark for the industry, showcasing the potential for operational efficiency with reduced delays.
- **Long-term Analysis for Policy Making:** Such historical data is crucial for policymakers and industry regulators to evaluate the effectiveness of past and current policies and to formulate future strategies for delay mitigation.

Key Observations:

- The integrated dashboard provides a detailed view of flight arrivals, delays, cancellations, and diversions over a span of two decades. Here are seven key observations and the inferences that can be drawn from this dashboard:
- Overall Trend of Flight Operations:
  - The bar graph indicates an overall increasing trend in the number of flights up until 2019, suggesting growth in the airline industry.
  - Inference: Air travel demand has grown consistently, necessitating scalable operational solutions.
- Impact of COVID-19:
  - There is a pronounced dip in 2020 across all metrics, representing the substantial impact of the COVID-19 pandemic on air travel.
  - Inference: Airlines and airports should prepare for swift responses to future global crises to maintain operational stability.
- High On-Time Performance:

- The pie chart shows that most flights (77.74%) are on time, indicating a strong overall performance in flight punctuality.
- Inference: Despite challenges, airlines generally manage to maintain a high on-time arrival rate.

- Consistent Delay Rates:
  - The "% Delayed Over the Years" bar graph showcases a relatively consistent rate of delays over the years, with a slight downward trend.
  - Inference: Although the rate of delays is consistent, there is room for improvement in delay management.
- Lowest Delay Rates During the Pandemic:
  - 2020 shows the lowest percentage of delays, which correlates with fewer flights due to pandemic restrictions.
  - Inference: With fewer flights, airlines could manage schedules more effectively, suggesting that operational congestion contributes significantly to delays.
- Recovery post-pandemic:
  - The slight uptick in the number of flights in 2021 suggests the beginning of recovery from the pandemic's effects.
  - Inference: The industry is resilient but may need to adapt to new travel patterns and passenger behaviors post-pandemic.
- Canceled and Diverted Flights:
  - The percentages of canceled (2.15%) and diverted (0.32%) flights are relatively low.
  - Inference: Airlines effectively minimize cancellations and diversions, which are costly and disruptive.

Conclusion

After analyzing the presented graphs, here are detailed conclusions, inferences, and suggestions for business decisions:

- **Carrier and Late Aircraft as Primary Delay Causes:**
  - Carriers and late-arriving aircraft are significant contributors to delays.
  - Airlines should focus on improving turnaround times and maintenance efficiency.
- **Seasonal Patterns Affecting Delays:**
  - Delays peak during holiday seasons and during severe weather conditions like tornado seasons.
  - Airlines could implement seasonal staffing and adjust scheduling to mitigate these effects.
- **Geographic Disparities in Flight Volumes:**
  - Flight arrivals are concentrated in states like California and Texas, with hubs like Atlanta seeing the highest traffic.
  - Smaller airports or states with fewer arrivals could be targeted for growth and expansion strategies.
- **Impact of the COVID-19 Pandemic:**
  - The pandemic drastically reduced operations, providing an opportunity to reset and optimize scheduling and operational processes.
  - Airlines could reassess their route profitability and efficiency in the post-pandemic recovery phase.
- **Consistent On-Time Performance Despite Delays:**
  - A high percentage of flights consistently arrive on time, indicating effective overall operations.
  - Continued investments in operations research and optimization can further improve on-time performance.
- **Low Cancellation and Diversion Rates:**
  - Cancellations and diversions remain low, suggesting effective crisis management.
  - Continued training and investment in predictive analytics can help maintain these low rates.
- **Potential in Low Traffic Markets:**
  - States with lower flight arrivals may represent untapped markets or opportunities for new routes.
  - Airlines should analyze market demand and potential in these regions for possible expansion.
- **Environmental Considerations:**
  - High-traffic states and airports should consider the environmental impact of operations.
  - Investment in more fuel-efficient aircraft and sustainable practices could be beneficial.
- **Infrastructure Needs:**
  - High-traffic airports may require infrastructure upgrades to handle future growth.
  - Engaging with government and private investors for airport development projects could be strategic.
- **Customer Service and Communication:**
  - Enhancing customer service and communication is vital, given the impact of delays, especially during peak seasons.
  - Implementing comprehensive passenger communication strategies can improve customer satisfaction.
- In conclusion, while the airline industry is adept at managing its operations, there is always room for improvement, especially in areas affected by external factors like weather and seasonal demand. Airlines can further optimize their operations and expand their market presence by leveraging data analytics, investing in infrastructure, and focusing on customer experience.

References

1. **Bureau of Transportation Statistics - U.S. Department of Transportation. (2023). *Airline On-Time Performance Data*. Retrieved from https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp**

Appendix

1. Tableau File
   - To access the Tableau file, please click on the link and open the file.
   - Then connect the data to Databricks for the live connection using the following personal access token: **dapi529bfad95b7763f3ab8738bb8e167e14**
   - Once the connection is established, you can click on the extract data button for a faster and smoother experience.
2. Presentation File
3. SQL Queries:

   We use SQL queries to filter and aggregate the data for the analysis. I have added some reference queries as example.

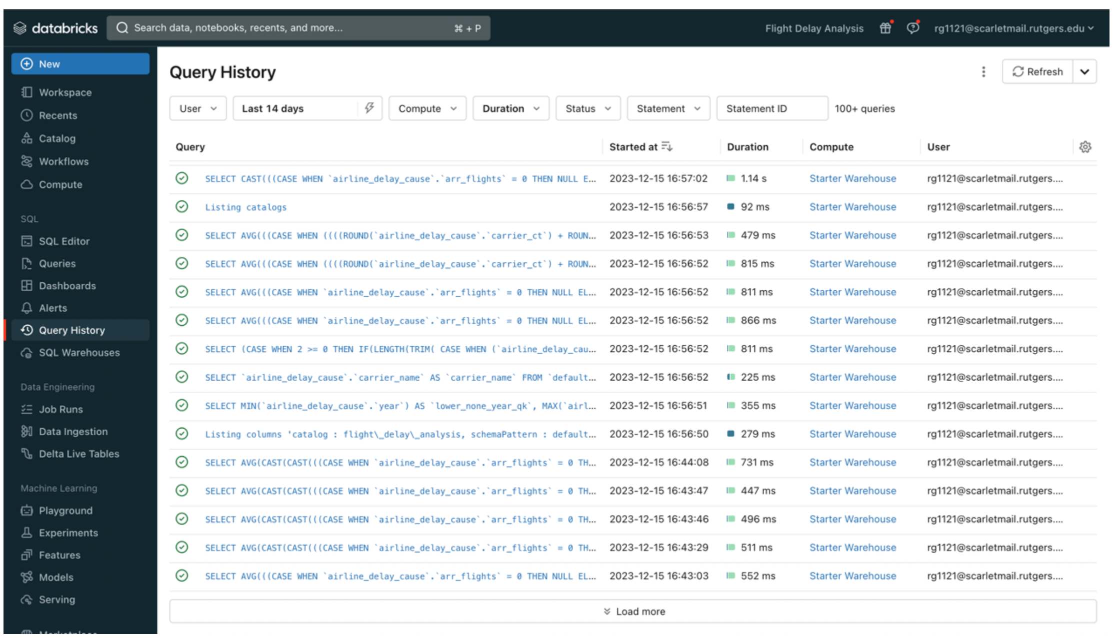


Figure 5: SQL queries on Databricks e.g. 1

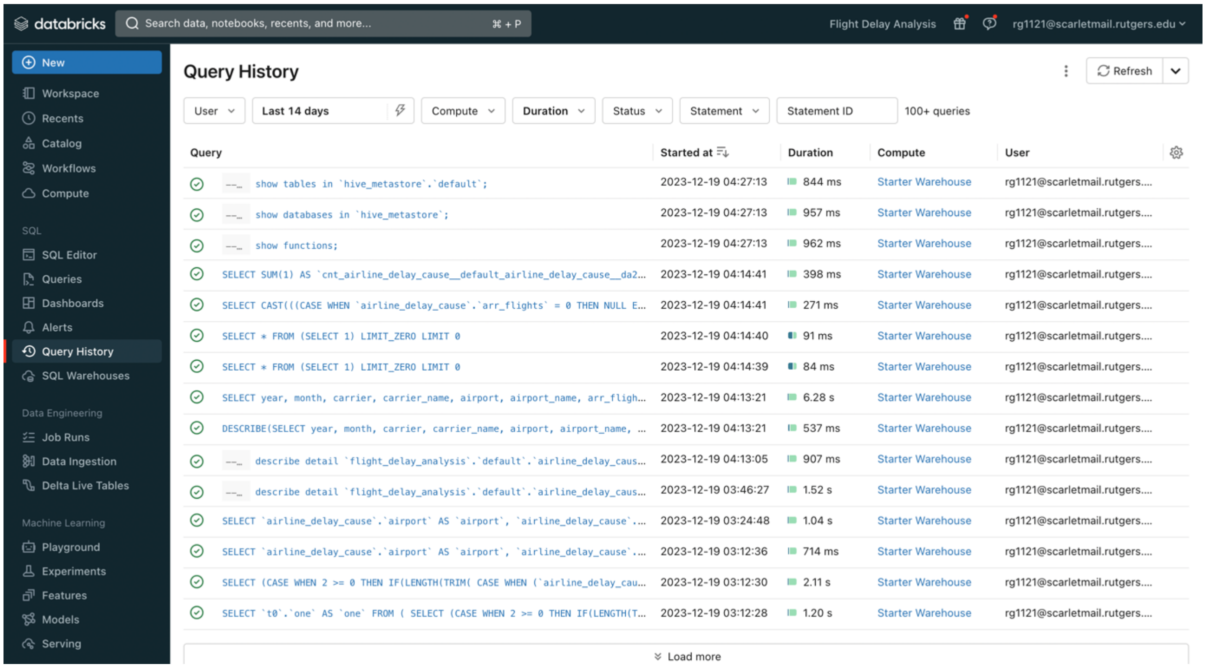


Figure 6: SQL queries on Databricks e.g. 2

```
SELECT
 (
  CASE
   WHEN 2 >= 0 THEN IF(
    LENGTH(
     TRIM(
      CASE
       WHEN (`airline_delay_cause`.`airport_name` IS NULL) THEN CAST(NULL AS STRING)
       WHEN NOT (1 IS NULL) THEN COALESCE(
        (
         CASE
          WHEN 1 > 0 THEN SPLIT(`airline_delay_cause`.`airport_name`, ':') [1-1]
          ELSE SPLIT(
           REVERSE(`airline_delay_cause`.`airport_name`),
           REVERSE(':')
          ) [ABS(1)-1]
         END
        ),
        ''
       )
       ELSE NULL
      END
     )
    ) < 2 + 1,
    TRIM(
     CASE
      WHEN (`airline_delay_cause`.`airport_name` IS NULL) THEN CAST(NULL AS STRING)
      WHEN NOT (1 IS NULL) THEN COALESCE(
       (
        CASE
         WHEN 1 > 0 THEN SPLIT(`airline_delay_cause`.`airport_name`, ':') [1-1]
         ELSE SPLIT(
          REVERSE(`airline_delay_cause`.`airport_name`),
          REVERSE(':')
         ) [ABS(1)-1]
        END
       ),
       ''
      )
      ELSE NULL
     END
    ),
    SUBSTRING(
     TRIM(
      CASE
       WHEN (`airline_delay_cause`.`airport_name` IS NULL) THEN CAST(NULL AS STRING)
       WHEN NOT (1 IS NULL) THEN COALESCE(
        (
         CASE
          WHEN 1 > 0 THEN SPLIT(`airline_delay_cause`.`airport_name`, ':') [1-1]
          ELSE SPLIT(
           REVERSE(`airline_delay_cause`.`airport_name`),
           REVERSE(':')
          ) [ABS(1)-1]
```

```
              END
            ),
            ''
          )
          ELSE NULL
        END
      ),
      LENGTH(
        TRIM(
          CASE
            WHEN (`airline_delay_cause`.`airport_name` IS NULL) THEN CAST(NULL AS STRING)
            WHEN NOT (1 IS NULL) THEN COALESCE(
              (
                CASE
                  WHEN 1 > 0 THEN SPLIT(`airline_delay_cause`.`airport_name`, ':') [1-1]
                  ELSE SPLIT(
                    REVERSE(`airline_delay_cause`.`airport_name`),
                    REVERSE(':')
                  ) [ABS(1)-1]
                END
              ),
              ''
            )
            ELSE NULL
          END
        )
      ) - CAST(2 AS INT) + 1
    )
  )
  ELSE NULL
 END
) AS `calculation_1206261045223809041`,
COUNT(DISTINCT `airline_delay_cause`.`airport`) AS `ctd_airport_ok`,
SUM(`airline_delay_cause`.`arr_diverted`) AS `sum_arr_diverted_ok`
FROM
 `default`.`airline_delay_cause` `airline_delay_cause`
 JOIN (
  SELECT
   (
    CASE
     WHEN 2 >= 0 THEN IF(
      LENGTH(
        TRIM(
          CASE
            WHEN (`airline_delay_cause`.`airport_name` IS NULL) THEN CAST(NULL AS STRING)
            WHEN NOT (1 IS NULL) THEN COALESCE(
              (
                CASE
                  WHEN 1 > 0 THEN SPLIT(`airline_delay_cause`.`airport_name`, ':') [1-1]
                  ELSE SPLIT(
                    REVERSE(`airline_delay_cause`.`airport_name`),
                    REVERSE(':')
                  ) [ABS(1)-1]
                END
              ),
              ''
            )
            ELSE NULL
          END
        )
      ) < 2 + 1,
      TRIM(
        CASE
          WHEN (`airline_delay_cause`.`airport_name` IS NULL) THEN CAST(NULL AS STRING)
          WHEN NOT (1 IS NULL) THEN COALESCE(
            (
              CASE
                WHEN 1 > 0 THEN SPLIT(`airline_delay_cause`.`airport_name`, ':') [1-1]
                ELSE SPLIT(
                  REVERSE(`airline_delay_cause`.`airport_name`),
                  REVERSE(':')
                ) [ABS(1)-1]
              END
            ),
            ''
          )
          ELSE NULL
        END
      ),
      SUBSTRING(
        TRIM(
          CASE
            WHEN (`airline_delay_cause`.`airport_name` IS NULL) THEN CAST(NULL AS STRING)
            WHEN NOT (1 IS NULL) THEN COALESCE(
              (
                CASE
                  WHEN 1 > 0 THEN SPLIT(`airline_delay_cause`.`airport_name`, ':') [1-1]
                  ELSE SPLIT(
                    REVERSE(`airline_delay_cause`.`airport_name`),
                    REVERSE(':')
                  ) [ABS(1)-1]
                END
              ),
              ''
            )
            ELSE NULL
          END
        ),
        LENGTH(
          TRIM(
            CASE
              WHEN (`airline_delay_cause`.`airport_name` IS NULL) THEN CAST(NULL AS STRING)
```

```
            WHEN NOT (1 IS NULL) THEN COALESCE(
              (
                CASE
                  WHEN 1 > 0 THEN SPLIT(`airline_delay_cause`.`airport_name`, ':') [1-1]
                  ELSE SPLIT(
                    REVERSE(`airline_delay_cause`.`airport_name`),
                    REVERSE(':')
                  ) [ABS(1)-1]
                END
              ),
              ''
            )
            ELSE NULL
          END
        )
      ) - CAST(2 AS INT) + 1
    )
  )
  ELSE NULL
  END
 ) AS `calculation_1206261045223809041`,
 COUNT(DISTINCT `airline_delay_cause`.`airport`) AS `x__alias__0`
FROM
 `default`.`airline_delay_cause` `airline_delay_cause`
GROUP BY
 1
ORDER BY
 `x__alias__0` DESC,
 `calculation_1206261045223809041` ASC
LIMIT
 5
) `t0` ON (
 (
  CASE
   WHEN 2 >= 0 THEN IF(
    LENGTH(
     TRIM(
      CASE
       WHEN (`airline_delay_cause`.`airport_name` IS NULL) THEN CAST(NULL AS STRING)
       WHEN NOT (1 IS NULL) THEN COALESCE(
        (
         CASE
          WHEN 1 > 0 THEN SPLIT(`airline_delay_cause`.`airport_name`, ':') [1-1]
          ELSE SPLIT(
           REVERSE(`airline_delay_cause`.`airport_name`),
           REVERSE(':')
          ) [ABS(1)-1]
         END
        ),
        ''
       )
       ELSE NULL
      END
     )
    ) < 2 + 1,
    TRIM(
     CASE
      WHEN (`airline_delay_cause`.`airport_name` IS NULL) THEN CAST(NULL AS STRING)
      WHEN NOT (1 IS NULL) THEN COALESCE(
       (
        CASE
         WHEN 1 > 0 THEN SPLIT(`airline_delay_cause`.`airport_name`, ':') [1-1]
         ELSE SPLIT(
          REVERSE(`airline_delay_cause`.`airport_name`),
          REVERSE(':')
         ) [ABS(1)-1]
        END
       ),
       ''
      )
      ELSE NULL
     END
    ),
    SUBSTRING(
     TRIM(
      CASE
       WHEN (`airline_delay_cause`.`airport_name` IS NULL) THEN CAST(NULL AS STRING)
       WHEN NOT (1 IS NULL) THEN COALESCE(
        (
         CASE
          WHEN 1 > 0 THEN SPLIT(`airline_delay_cause`.`airport_name`, ':') [1-1]
          ELSE SPLIT(
           REVERSE(`airline_delay_cause`.`airport_name`),
           REVERSE(':')
          ) [ABS(1)-1]
         END
        ),
        ''
       )
       ELSE NULL
      END
     ),
     LENGTH(
      TRIM(
       CASE
        WHEN (`airline_delay_cause`.`airport_name` IS NULL) THEN CAST(NULL AS STRING)
        WHEN NOT (1 IS NULL) THEN COALESCE(
         (
          CASE
           WHEN 1 > 0 THEN SPLIT(`airline_delay_cause`.`airport_name`, ':') [1-1]
           ELSE SPLIT(
            REVERSE(`airline_delay_cause`.`airport_name`),
```

```
            REVERSE(':')
           ) [ABS(1)-1]
          END
         ),
         ''
        )
        ELSE NULL
       END
      )
     ) - CAST(2 AS INT) + 1
    )
   )
   ELSE NULL
  END
 ) = `t0`.`calculation_1206261045223809041`
)
WHERE
 (
  (`airline_delay_cause`.`year` >= 2003)
  AND (`airline_delay_cause`.`year` <= 2023)
 )
GROUP BY
 1
```