Big Data and Business Intelligence

Module Code: CIS4008-N-BF1-2023

Module Leader: Mansha Nawaz

Topic: Analysis of US Domestic Flight Delays and Cancellations in 2023.

Summary Executive BI Report for Stakeholders

Name: Konda Reddy Thokala

Student number: C2994587

Submission date: 10 Jan 2024

MSc Data Science
School of Computing, Engineering & Digital Technologies
Teesside University
Middlesbrough
Tees Valley, TS1 3BX
United Kingdo

Table of Contents

Summary Executive BI Report for Stakeholders

1. Introduction:

The presented dataset contains critical information about flight operations and can be used to create an insightful Executive Business Intelligence (BI) Report for aviation stakeholders. The dataset delves into key performance indicators such as departure and arrival times, delays, cancellations, and diversions, and provides a comprehensive overview of flights across different dates, airlines, and routes. Stakeholders can gain valuable insights into the airline industry's operational efficiency and challenges, which can aid in strategic decision-making.

Power BI emerges as an invaluable tool for harnessing the potential of this data and delivering a dynamic, interactive, and visually compelling BI Report. Power BI's robust data visualization capabilities enable the creation of intuitive charts, graphs, and dashboards, transforming raw data into actionable insights. Its seamless integration with diverse data sources and ability to handle large datasets enable stakeholders to investigate trends, identify patterns, and draw meaningful conclusions. By utilizing Power BI's functionalities, this Executive BI Report ensures that stakeholders can easily comprehend complex data, fostering a data-driven approach to improve operational efficiency and optimize decision-making within the aviation industry.

Audiences:

- Director of Civil Rights Advocacy for the Office of Aviation Consumer Protection
- Director of Consumer Advocacy for the Office of Aviation Consumer Protection
- Deputy Assistant General Counsel for the Office of Aviation Consumer Protection
- DOT US., Senior Advisors and other Officials of DOT US.

Key Highlights:

1. **Date and Time Analysis:** Because the dataset includes flights on various dates, stakeholders can identify patterns and trends in flight operations. This data is critical for improving scheduling and resource allocation.
2. **Airport Connectivity:** Information about the airline's origin and destination airports highlights its route network. Stakeholders can evaluate the airline's presence in various regions, which can aid in strategic decision-making for future expansions.
3. **Punctuality Metrics:** Accurate information on departure and arrival times, as well as delays, provides a comprehensive picture of the airline's punctuality. Based on these metrics, stakeholders can assess operational efficiency and customer satisfaction.
4. **Delay Attribution:** The dataset categorizes delays by carrier, weather, National Airspace System (NAS), security, and late aircraft. This breakdown allows stakeholders to identify areas for improvement and evaluate external factors influencing performance.
5. **Cancellations and Delays:** Understanding the frequency of cancellations and delays is critical for determining the airline's dependability. This data is critical for customer satisfaction and can help guide decision-making to improve operational resilience.

Purpose of the Report:

The primary goal of this Executive BI Report is to provide stakeholders with actionable insights into the airline's operational performance. Stakeholders can identify strengths, weaknesses, and areas for improvement by delving into the complexities of each flight's journey. This report is an excellent resource for strategic planning, resource optimization, and providing a positive customer experience.

2. Business Questions:

- How many miles, in total, were covered by the flights in the dataset?
- What is the cumulative airtime, in minutes, for all the flights in the dataset?
- Which carriers top the leaderboard based on the number of flights they operated?
- How many months are encompassed in the provided dataset?
- During the specified time period, what is the total count of carriers that operated flights?
- How many distinct destinations were operated during the specified time period?
- What is the overall count of flights operated during the specified time period?
- What are the top 5 airports based on the number of flights they handled?
- Which airlines rank in the top 5 based on the unique destinations they handled?
- What is the percentage of flights cancelled in that time period?
- What the overall count of the cancelled flights?
- What are the major reasons for the flight cancellations?
- What are the top 8 airlines which are having highest number of cancellations?
- Which month is the having highest and lowest number of cancellations?
- What is the average arrival delay?
- What is the average departure delay?
- What is the on time departure percentage?
- What are major factors which are involved in the flight delays?
- Which month is having highest and lowest average arrival delay and departure delay?
- What are the top 5 origin cities based on average departure delay?
- What is the percentage of diverted flights?
- What is the total number of diverted flights?
- What is the average departure delay for diverted flights in mins?
- Which month is having highest and lowest diverted flights?
- Which airlines are highest having diverted flights?
- Which Destination City is having the highest number of diverted flights?

3. KPIs and Dashboards:

3.1 KPIs

Key Performance Indicators (KPIs) are measurable values that show how well an organization is meeting its key business objectives. KPIs are critical metrics in Power BI for measuring and monitoring performance. These quantifiable indicators, such as total distance travelled, airtime, and operational counts, guide the development of insightful Power BI dashboard questions. Stakeholders can focus on specific aspects of flight operations by aligning questions with KPIs, allowing for a more targeted analysis of data trends. This method ensures that Power BI queries directly address organizational goals, providing actionable insights and facilitating informed decision-making in the aviation domain.

To analysis the KPIs and get the insights for the business questions, the visualization is divided into 4 dashboards as mentioned below.

- Overview
- Delayed Flight Analysis
- Cancelled Flights Analysis
- Diverted Flight Analysis.

3.2 Dashboards:

Overview Dashboard:

The Overview dashboard provides a comprehensive source of high-level insights into the dataset by presenting key information as well as ranking metrics for airlines and airports based on specific parameters. This dashboard gives stakeholders a quick and comprehensive understanding of key dataset characteristics, enabling efficient decision-making and strategic analysis at a glance.

The business questions answered in this dashboard are as follows,

- **How many miles, in total, were covered by the flights in the dataset?**

The dashboard includes a Card Visual to show that the flights in the dataset collectively covered an impressive 3,789,417,823 miles between January and August 2023. This concise representation gives stakeholders a quick, at-a-glance overview of the vast distance travelled by flights during the specified time period.

- **What is the cumulative airtime, in minutes, for all the flights in the dataset?**

  The dashboard includes a Card Visual that shows that, from January to August 2023, the total airtime completed by all flights in the dataset is 512,574,528 minutes. This simple representation provides stakeholders with a quick, impactful overview of the total time spent in the air, providing insights into the scope and magnitude of flight operations during the specified period.

- **How many months are encompassed in the provided dataset?**
  The dashboard includes a Card Visual to effectively communicate that the dataset spans 8 months, from January to August 2023. This concise representation gives stakeholders a quick overview of the temporal scope, allowing them to understand the time span covered by the dataset.

- **During the specified time period, what is the total count of carriers that operated flights?**
  Within the time frame specified, the dashboard includes a Card Visual to highlight the presence of 15 domestic carriers that operate flights solely within the United States. This concise representation provides stakeholders with a quick overview of the number of carriers contributing to the dataset, emphasizing the emphasis on domestic air travel.

- **How many distinct destinations were operated during the specified time period?**
  A Card Visual on the dashboard effectively communicates that the 15 airlines operated 349 distinct destinations during the specified time period. This clear representation provides stakeholders with a quick overview of the extensive network coverage, highlighting the breadth of destinations served within the dataset. The Card Visual is a useful tool for summarizing and highlighting key metrics, allowing for a thorough understanding of the diversity and reach of airline operations during the specified timeframe.

- **What is the overall count of flights operated during the specified time period?**
  The dashboard uses a Card Visual to show that the 15 airline companies efficiently handled a total of 4.55 million flights over the specified time period. These flights connected to 349 distinct destinations, demonstrating the dataset's broad reach and operational scale. This concise representation gives stakeholders a quick, at-a-glance summary of overall flight activity, providing valuable insights into the scope and magnitude of air travel from January to August.

- **Which carriers top the leaderboard based on the number of flights they operated?**
  A Table Visual on the dashboard effectively displays the leaderboard of airlines, ranking them based on the number of flights handled during the specified time period. Southwest Airlines Co. (WN) takes the top spot, having flown a total of 941,733 flights. On the other end of the spectrum, Hawaiian Airlines Inc. (HA) ranks last with 54,011 flights. This tabular presentation, organized by rank, airline name, and unique code, gives stakeholders a comprehensive overview of how flight operations are distributed among the various carriers.

- **What are the top 5 airports based on the number of flights they handled?**
  A table on the dashboard effectively depicts the top 5 airports based on the number of flights they handle. Atlanta, GA (ATL) takes the lead with an impressive 223,368 flights in 2023, according to the dataset. Denver, CO (DEN), Dallas/Fort Worth, TX (DFW), Chicago, IL (ORD), and Los Angeles, CA (LAX) are close behind with 189,716, 187,141, 172,653, and 129,688 flights, respectively. This tabular representation provides stakeholders with a detailed breakdown of airports and their associated flight volumes, allowing for more in-depth insights into air traffic distribution.

- **Which airlines rank in the top 5 based on the unique destinations they handled?**

  A table on the dashboard represents the top five airlines in terms of the number of unique destinations they serve. SkyWest Airlines Inc. emerges as the leading carrier, having managed an impressive 264 unique destinations across the United States, according to the dataset. The table gives stakeholders a detailed breakdown of airlines and their corresponding counts of unique destinations, allowing them to gain a thorough understanding of each carrier's network coverage.

Delayed Flight Analysis:

The Delayed Flight Analysis dashboard is useful for investigating delayed flights during the specified time period, providing information on both arrival and departure delays. This analysis explores the major causes of delays, providing stakeholders with valuable information to improve operational efficiency and address specific areas contributing to flight delays. It is calculated as the difference in minutes between scheduled and actual arrival/departure times. The dashboard is a powerful tool for understanding and mitigating the impact of delays.

Dashboard:

The business questions answered in this dashboard are as follows,

- **What is the average arrival delay?**
  A card visual on the dashboard effectively communicates that, according to the dataset, the average arrival delay is 9.28 minutes within the specified time period. This computation is derived by averaging the arrival delay across all flights, providing stakeholders with a clear and concise view of flight performance in terms of arrival punctuality.

- **What is the average departure delay?**
  A card visual in the dashboard communicates succinctly that, based on the dataset for the specified time frame, the average departure delay is 14.41 minutes. This calculation is based on averaging the departure delay across all flights, and it provides stakeholders with a concise yet informative metric on overall departure punctuality.

- **What is the on-time departure percentage?**
  A card visual on the dashboard effectively conveys that approximately 60% of flights are left on time, indicating instances where the actual departure time aligned precisely with the scheduled time. This brief representation provides stakeholders with a quick and impactful overview of the dataset's punctuality performance.

- **What are major factors which are involved in the flight delays?**
  To effectively communicate the factors contributing to flight delays within the dataset, the dashboard employs a Card Visual. Notably, delays caused by late flights are the most significant factor, accounting for nearly 70% of all delays. Delay due to security, on the other hand, has the least influence, accounting for only 0.35%. This clear representation provides stakeholders with a quick overview of the relative impact of different delay factors, guiding strategic insights into operational improvements.

- **Which month is having highest and lowest average arrival delay and departure delay?**
  A Multi-Line Chart is used in the dashboard to effectively illustrate the temporal variation in average arrival and departure delays within the dataset. Notably, the month of July had the highest average arrival delay of 16.64 minutes and the longest average departure delay of 21.25 minutes. In comparison, May month had the shortest average arrival delay of 3.81 minutes and the longest departure delay of 10.03 minutes. This visualization method enables stakeholders to identify trends, patterns, and seasonal variations in average delay times, allowing for a deeper comprehension of the dataset's temporal dynamics.

- **What are the top 5 origin cities based on average departure delay?**
  A Bar Chart is effectively utilized in the dashboard to display the average departure delays for various airport cities based on the dataset. Notably, Santa Maria, CA has the longest average departure delay (56.53 minutes), followed by Pago Pago, TT (44.25 minutes), Ashland, WV (34.97 minutes), Sault Ste. Marie, MI (31.45 minutes), and Aguadilla, PR (30.50 minutes). This visual representation gives stakeholders a clear and comparative overview of departure delays across these airport cities, assisting them in understanding performance variations.

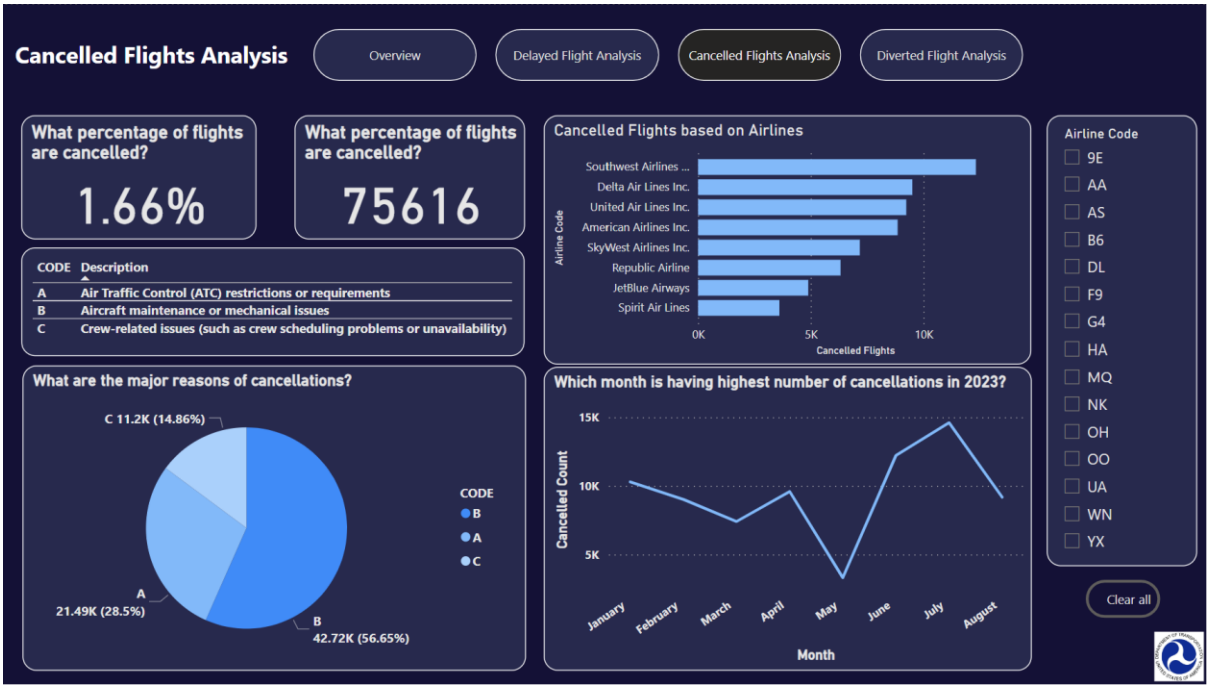
- **Slicer**

  Using a slicer in the dashboard for carrier selection allows for in-depth analysis of patterns and trends specific to individual carrier companies. Stakeholders gain a comprehensive understanding of performance metrics by studying insights based on each carrier, allowing for a complex assessment of factors influencing operational efficiency, delays, and overall business performance within the aviation dataset. This focused approach enables focused exploration and meaningful comparisons across different carriers, increasing the dashboard's effectiveness in providing actionable insights for strategic decision-making.

Cancelled Flights Analysis:

The Cancellation Flight Analysis dashboard is extremely useful for investigating cancelled flights during the specified time period, providing insights into both the frequency and causes of cancellations. This investigation looks into the primary causes of flight cancellations, providing stakeholders with critical information to improve operational efficiency and address specific factors that contribute to cancelled flights. The dashboard, which is based on recorded cancellation instances, is a powerful tool for comprehending and mitigating the impact of cancelled flights.

Dashboard:



The business questions answered in this dashboard are as follows,

- **What is the percentage of flights cancelled in that time period?**

  A Card Visual in the dashboard effectively communicates that the overall cancellation rate for flights, according to the dataset, is 1.66%. This calculation is based on the "cancelled" column, where a value of 1 indicates that the flight was cancelled. The Card Visual gives stakeholders a quick and impactful snapshot of the percentage of cancelled flights, allowing for a concise understanding of the dataset's cancellation dynamics.

- **What the overall count of the cancelled flights?**
  A Card Visual in the dashboard communicates simply that the total number of cancelled flights within the specified time period, according to the dataset, is 75,616. This representation provides stakeholders with a quick, at-a-glance overview of the magnitude of flight cancellations, as well as a key metric for understanding the impact of cancellations within the dataset.

- **What are the major reasons for the flight cancellations?**
  A Pie Chart in the dashboard effectively shows the distribution of reasons for flight cancellations within the dataset. Aircraft maintenance or mechanical issues emerge as the leading cause, accounting for 42% of total cancellations, followed by ATC restrictions or requirements at 28% and crew-related issues at 15%. This visual representation gives stakeholders a clear and proportionate understanding of the primary factors influencing flight cancellations, assisting in strategic decision-making to address specific challenges in the aviation dataset.

- **What are the top 8 airlines which are having highest number of cancellations?**
  Based on the dataset, the dashboard includes a Horizontal Bar Chart to effectively display the number of flight cancellations among airlines. Southwest Airlines Co. leads the pack with the most flight cancellations, totalling 12,328. This visualisation optimises the representation of categorical values, allowing stakeholders to quickly identify and compare the cancellation counts of different airlines.

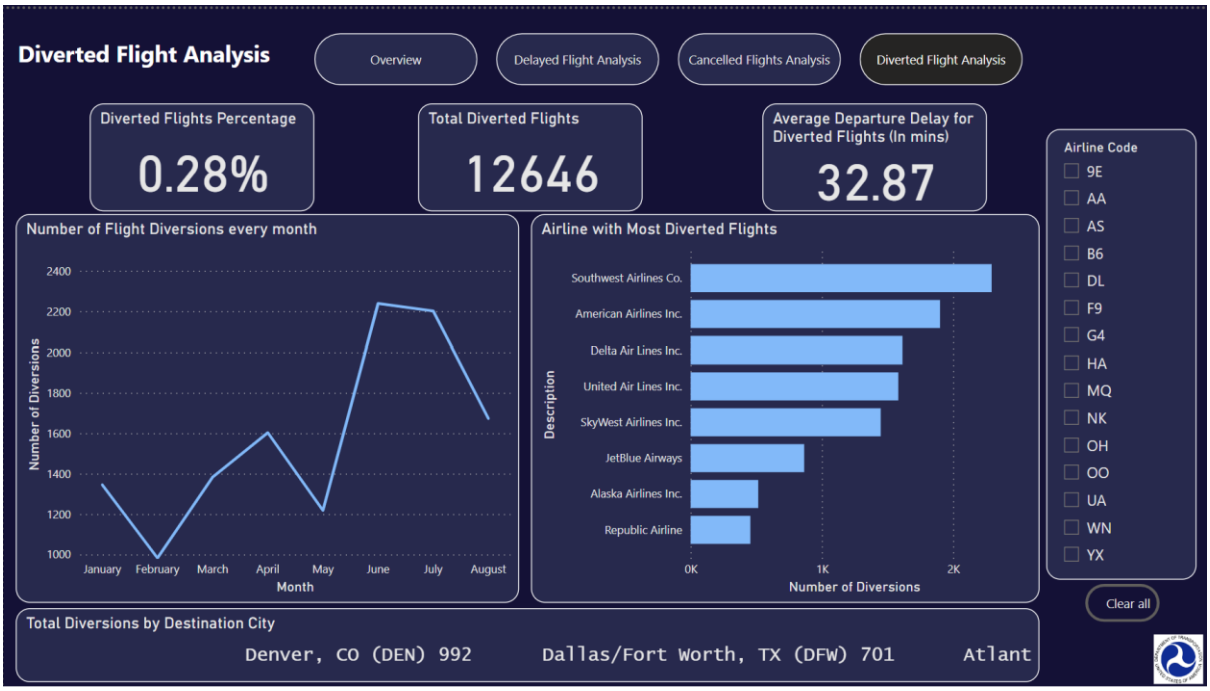
- **Which month is the having highest and lowest number of cancellations?**
  Based on the dataset, the dashboard employs a Line Graph to effectively visualise the fluctuation in the number of flight cancellations across different months. Particularly, July has the most cancellations with 14,606 flights, while May has the fewest with 3,310 flights. Using a line graph to illustrate this time series data allows stakeholders to identify trends and variations in cancellation counts over the specified months.

- **Slicer**
  Using a slicer in the dashboard to select a carrier allows for a more in-depth examination of patterns and trends specific to individual carrier companies within the cancellation flight analysis. By studying insights associated with each carrier, stakeholders gain a thorough understanding of performance metrics, allowing for a complex assessment of factors influencing cancellation rates and overall operational efficiency. This targeted approach improves the dashboard's effectiveness in providing focused insights and meaningful comparisons across various carriers, enabling informed decision-making in dealing with cancellation-related issues.

Diverted Flight Analysis:

The Diverted Flight Analysis dashboard is extremely useful for examining diverted flights over a specified time period, providing insights into both the frequency and causes of diversions. This analysis explores into the primary causes of flight diversions, providing stakeholders with critical information to improve operational efficiency and address specific factors that contribute to flight diversions. The dashboard, by utilising recorded diversion instances, serves as a powerful tool for comprehending and mitigating the impact of diversions.

Dashboard:



The business questions answered in this dashboard are as follows,

- **What is the percentage of diverted flights?**
  A Card Visual on the dashboard shows that the percentage of diverted flights in the overall dataset is 0.28%. This representation provides stakeholders with a quick and impactful snapshot of the proportion of flights that experienced diversions, serving as a critical metric for understanding the impact of diversions within the dataset.

- **What is the total number of diverted flights?**
  A Card Visual in the dashboard communicates simply that the total number of flights diverted within the specified time period, according to the dataset, is 12,646. This representation provides stakeholders with a quick, at-a-glance overview of the extent of flight diversions, as well as a critical metric for understanding the impact of diversions within the dataset.

- **What is the average departure delay for diverted flights in mins?**
  The dashboard makes effective use of a Card Visual to convey that, according to the dataset, the departure delay for diverted flights is approximately 33 minutes. This clear representation provides stakeholders with a quick and impactful insight into the average departure delay associated with diverted flights, providing a critical metric for understanding the temporal dynamics of diversions within the dataset.

- **Which month is having highest and lowest diverted flights?**
  Based on the dataset, the dashboard includes a Line Chart to effectively visualise the variation in the number of diverted flights across different months. Particularly, June has the highest number of diverted flights with 2,239, closely followed by July with 2,202. A line chart is ideal for displaying this time series data, allowing stakeholders to

identify trends and variations in the number of diverted flights over the specified months.

- **Which airlines are highest having diverted flights?**
  Based on the dataset, the dashboard effectively uses a Horizontal Bar Graph to display the number of diverted flights among different carriers. Southwest Airlines Co. leads the pack with 2,293 diverted flights, followed by American Airlines Inc. with 1,900 diverted flights. This visualisation optimises the representation of categorical values, allowing stakeholders to quickly identify and compare the diversion counts of different carriers.

- **Which Destination City is having the highest number of diverted flights?**

  The dashboard effectively uses a scroller visual with the format set to Airport City (Airport Code) to highlight those flights destined for Denver, CO (DEN) have the most diversions in the given time period, with Dallas/Fort Worth, TX (DFW) following closely behind with 701 diversions. This interactive visual element enables stakeholders to explore and focus on specific airport destinations within the dataset in real time.

- **Slicer**
  Using a slicer in the dashboard for carrier selection allows for a more in-depth examination of patterns and trends specific to individual carrier companies within the Diverted Flight Analysis. Stakeholders gain a comprehensive understanding of performance metrics by delving into insights associated with each carrier, allowing for a complex assessment of factors influencing diversion rates and overall operational efficiency. This targeted approach improves the dashboard's effectiveness in providing focused insights and meaningful comparisons across various carriers, allowing for informed decision-making to address flight diversion issues.

4. Summary:

From January to August 2023, the aviation dataset analysis provides a comprehensive overview of flight operations, including key metrics and insights. Flights travelled a total of 3,789,417,823 miles and spent 512,574,528 minutes in the air. The 15 airlines flew to 349 different destinations, demonstrating a diverse network coverage.

Southwest Airlines Co. led the pack in terms of flight operations, with 941,733 flights, while Hawaiian Airlines Inc. had 54,011 flights, indicating a diverse range of operations among carriers. Atlanta, GA (ATL), Denver, CO (DEN), Dallas/Fort Worth, TX (DFW), Chicago, IL (ORD), and Los Angeles, CA (LAX) were the top five airports in terms of flight volume. SkyWest Airlines Inc. distinguished itself by managing 264 distinct destinations, emphasising its extensive network.

According to punctuality metrics, flights had an average arrival delay of 9.28 minutes and a departure delay of 14.41 minutes. Particularly, 60% of all flights took off on time. Late flights were identified as the primary cause of delays, accounting for nearly 70% of all delays.

An analysis of flight cancellations revealed an overall rate of 1.66%, with 75,616 flights cancelled. The leading cause (42%), followed by ATC restrictions or requirements (28%), and crew-related issues (15%), was aircraft maintenance or mechanical issues. Southwest Airlines

Company had the most cancellations (12,328), and July had the most cancellations (14,606 flights).

12,646 flights were diverted, accounting for 0.28% of all flights. The average departure delay for diverted flights was 33 minutes. Southwest Airlines Co. once again led in diversions with 2,293 flights, and June saw the most diversions (2,239).

5. Report Outcomes:

The report on the aviation dataset analysis provides a comprehensive understanding of various key performance indicators, challenges, and operational dynamics in the aviation industry from January to August 2023. The following are the report's main findings:

Operational Overview: The report provides a detailed operational overview, which includes total distance travelled, airtime, and flight destinations. It provides information about the extensive network coverage and the scale of flight operations.

Carrier Performance: Stakeholders gain insight into individual carriers' performance, with a leaderboard displaying the number of flights operated. Southwest Airlines Co. emerges as a leader, and carrier diversity is evident.

Airport Analysis: Based on the number of flights handled, the report identifies the top airports, emphasising major hubs such as Atlanta, Denver, and Dallas/Fort Worth. This data is critical for comprehending air traffic distribution.

Metrics for Punctuality: Metrics for punctuality, such as average arrival and departure delays, provide a clear picture of flight performance. According to the report, flights experienced an average arrival delay of 9.28 minutes and a departure delay of 14.41 minutes.

Cancellation Insights: The report delves into flight cancellations, highlighting the overall cancellation rate, the reasons for cancellations, and the airlines that have been most affected. Southwest Airlines Co. has the most cancellations, and aircraft maintenance issues are a major factor.

Analysis of Flight Diversions: The report investigates flight diversions, displaying the percentage of diverted flights, the reasons for diversions, and the carriers and airports most impacted. Southwest Airlines Co. is once again the leader in diversions, with June recording the highest number.

Visualisation Techniques: To present data in a clear and understandable manner, the report makes effective use of various visualisation techniques such as card visuals, tables, charts, and slicers. This improves stakeholders' ability to derive actionable insights.

Actionable Insights: The outcomes provide stakeholders with actionable insights to help them make informed decisions. Whether dealing with delays, cancellations, or diversions, the report provides valuable information to decision-makers for improving operational efficiency and addressing specific challenges within the aviation dataset.

Overall, the report's findings provide stakeholders with a thorough understanding of the aviation dataset, allowing them to strategize and implement improvements to various aspects of flight operations.

6. Recommendations:

Based on the findings, the report will make specific recommendations to improve overall operational efficiency, reduce delays, and improve customer satisfaction. These suggestions

will be consistent with the overarching goal of stimulating robust and dependable airline service.

This Executive Business Intelligence Report serves as a foundation for informed decision-making, allowing stakeholders to steer the airline towards greater success and resilience in the volatile aviation industry.

7. Reference:

- US Department of Transportation: https://www.transtats.bts.gov/DL_SelectFields.aspx?gnoyr_VQ=FGJ&QO_fu146_anzr=b0-gvzr
- Dataset source: https://www.kaggle.com/datasets/patrickzel/flight-delay-and-cancellation-dataset-2019-2023
- Snowflake Schema Reference: https://phoenixnap.com/kb/star-vs-snowflake-schema