A1: Analysis of Carrier On-Time Performance by Awale Abdi

Table of Contents

Word Count excluding executive summary and references: 1,366

Executive Summary

This executive summary encapsulates this report's in-depth analysis of airline on-time performance. Utilizing a dataset from the US Department of Transportation, this report focus ranges from ranking airlines based on flight frequency to assessing optimal travel times and exploring the impact of weather on delays. Despite challenges in data's quality, key insights emerged, including industry-wide trends and unexpected findings. While certain limitations exist, this report's recommendations centre on refining data collection methods and fostering industry collaboration for a more robust understanding of aviation dynamics. Ultimately, this report aims to empower executive decision-making by providing actionable insights into enhancing overall airline operational efficiency despite the limited data it was forced to work with.

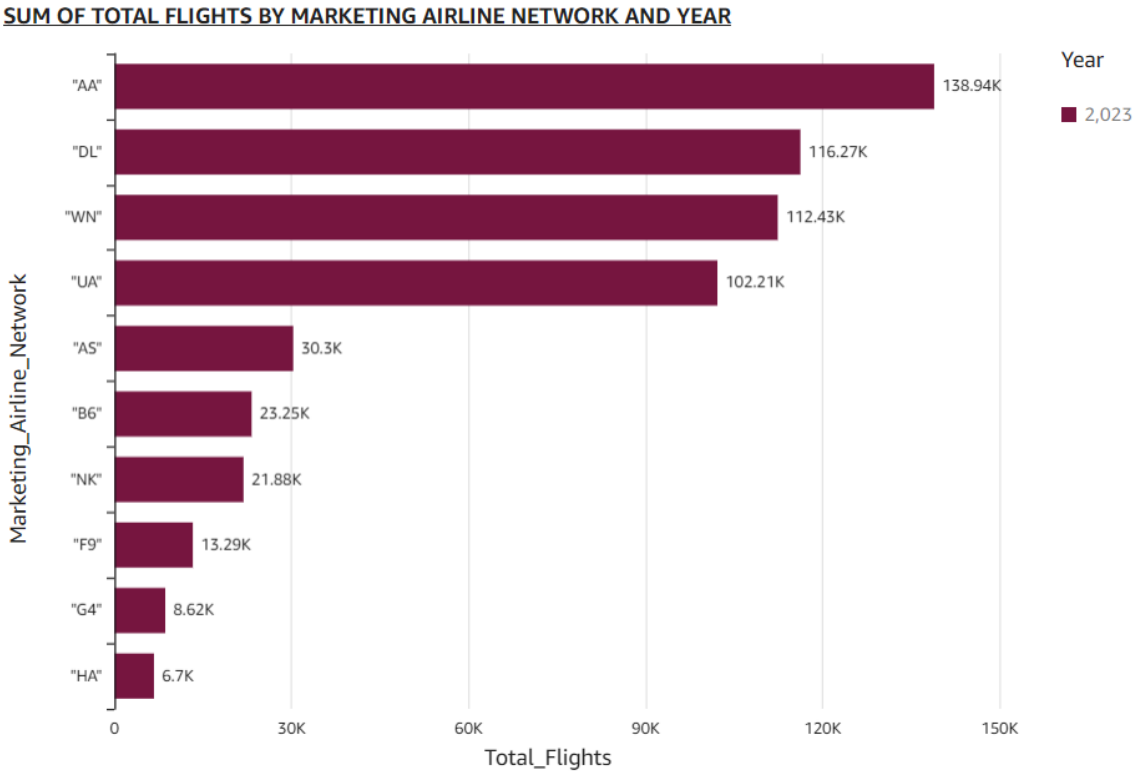
Introduction

Tasked by an executive team, this report aims to showcase a keen understanding of both business needs and the intricacies of data sources. Focused on evaluating airlines' on-time performance, the report’s objective is to dissect a dataset from the US Department of Transportation, procured via the Bureau of Transportation Statistics. The report has been directed to answer five critical questions concerning airline carrier performance, with each question guiding a specific analysis section. From ranking airlines based on flight frequency to discerning trends and identifying the optimal times to minimize delays, this report’s analysis aims to provide insights into pivotal aspects of the aviation industry. The deliverables will hopefully not only encompass a comprehensive understanding of airline operations but also offer actionable insights to inform strategic decisions. The end goal is to provide an executive team with a nuanced perspective on the dynamics influencing on-time performance in the airline industry.

Section 1: Ranking Airlines by year, identifying the airlines with the most flights and outlining any trends worth noting**.**

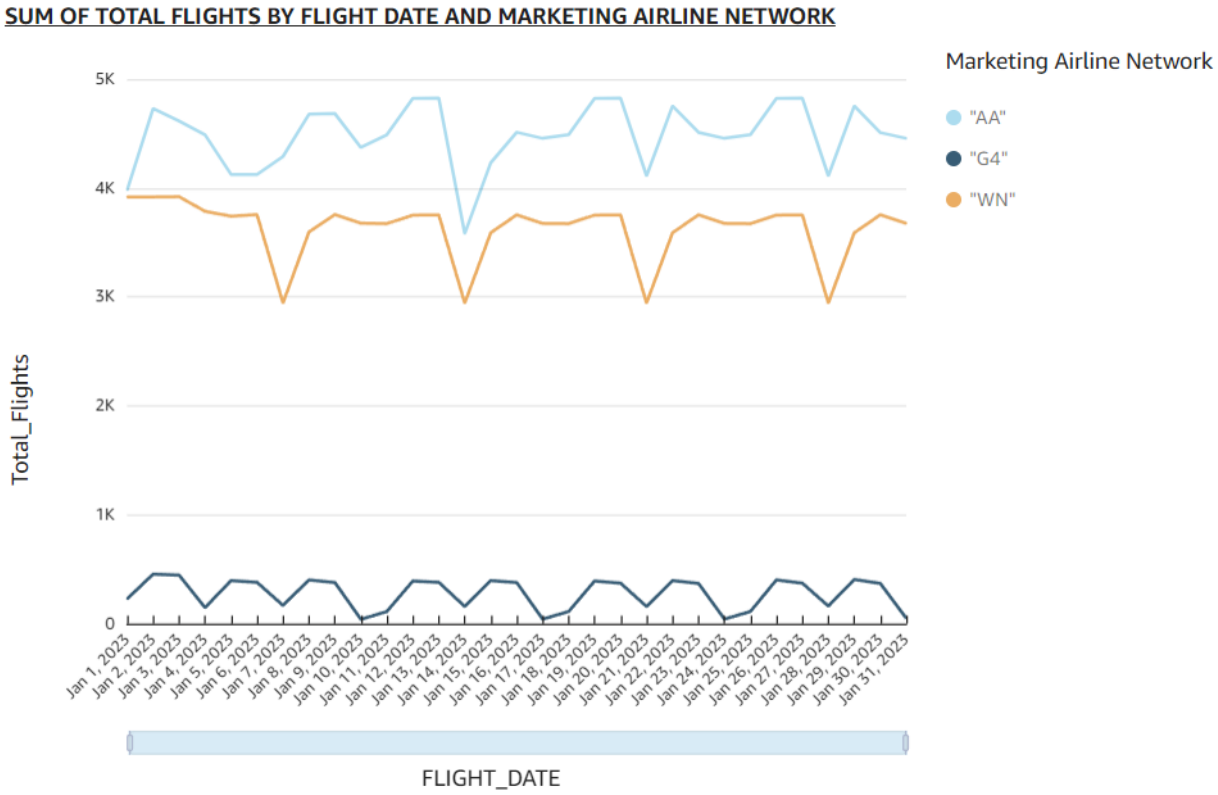
For this section of the analysis, it was found that only one year was made available in the dataset, meaning that any output would have to be limited entirely to that singular year which is admittedly less informative, but it is the best that can be done with the data available. The year in question that all the data throughout this report informs upon is **2023**.

Nevertheless, it was possible to identify the airlines with the most flights during 2023:



However, this is where another limitation of the dataset supplied was encountered. There is no field/column present where the names of Airlines are outright stated. The closest one could get is a field such as “Market Airline Network” where they are abbreviated as can be seen above. **Hence the most that can be gleaned is that “AA” is the airlines with the most flights at 138.9K flights.**

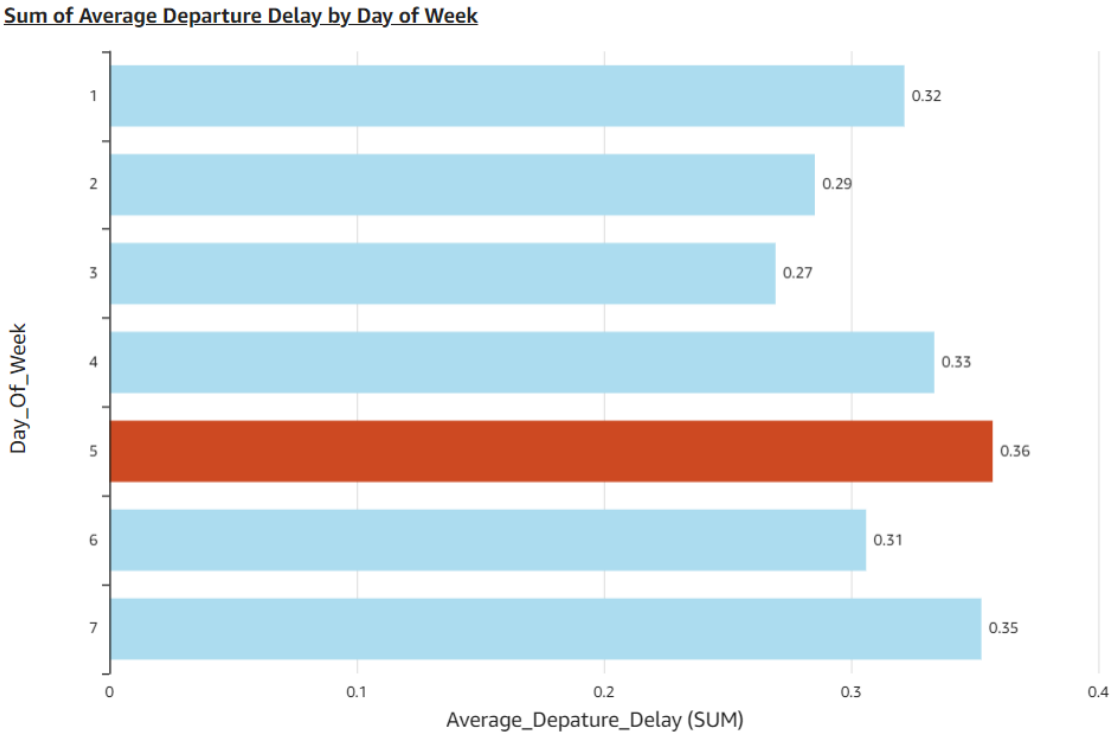
There was also something worth noting regarding trends and that is that virtually all the airlines seem to follow the same trend in terms of number of flights throughout the month except for one titled “HA” with the lowest number of trends. This can be seen below in that an airline with a low number of flights, an airline with a middling number of flights and airline with a high number of flights generally show the same trend throughout the month:



There seems to be an overall trend in terms of total flights that holds true regardless of airlines throughout the month.

Section 2: The best time of day, day of week and time of year to fly to minimize delays

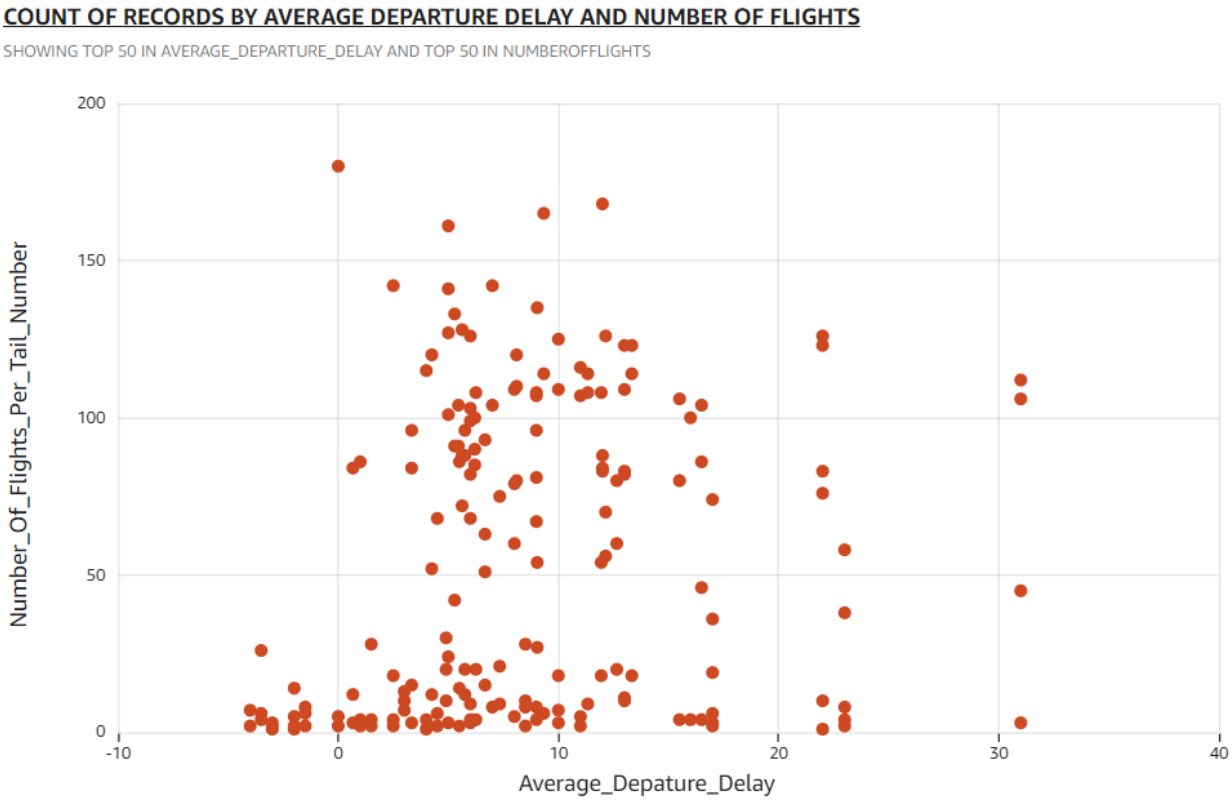
In addressing this section of the analysis problems with the dataset lacking data were once again encountered. There was no way to identify the best time of day as there was no such data visible once the data was being run through Athena Query Editor nor could such a field be found when the dataset's "ReadMe" was consulted. The time of year also proved redundant as all the data in the dataset is limited to one month of the year 2023, the first month (January). Therefore, the output had to be limited to the day of the week for that limited period:



The best day of the week to travel to minimize delays, during 2023's first month, by a small margin, is day 5. This was identified by noticing how many flights were delayed by 15 or more minutes per day of the week and generating an average. (consult question 2 in .txt file)

Section 3: Addressing whether older planes suffer more delays

This section's analysis, like those previous to it, encountered problems due to data not being present in the dataset. We do not have data showing directly how old each plane is at all through any of the fields therefore some amount of creativity had to be employed and that occurred through counting the number of times a distinct tail number (used to identify a plane) occurred as a measure of how old, at least within this limited period of one month within 2023, a plane was then noticing the average delay in minutes:

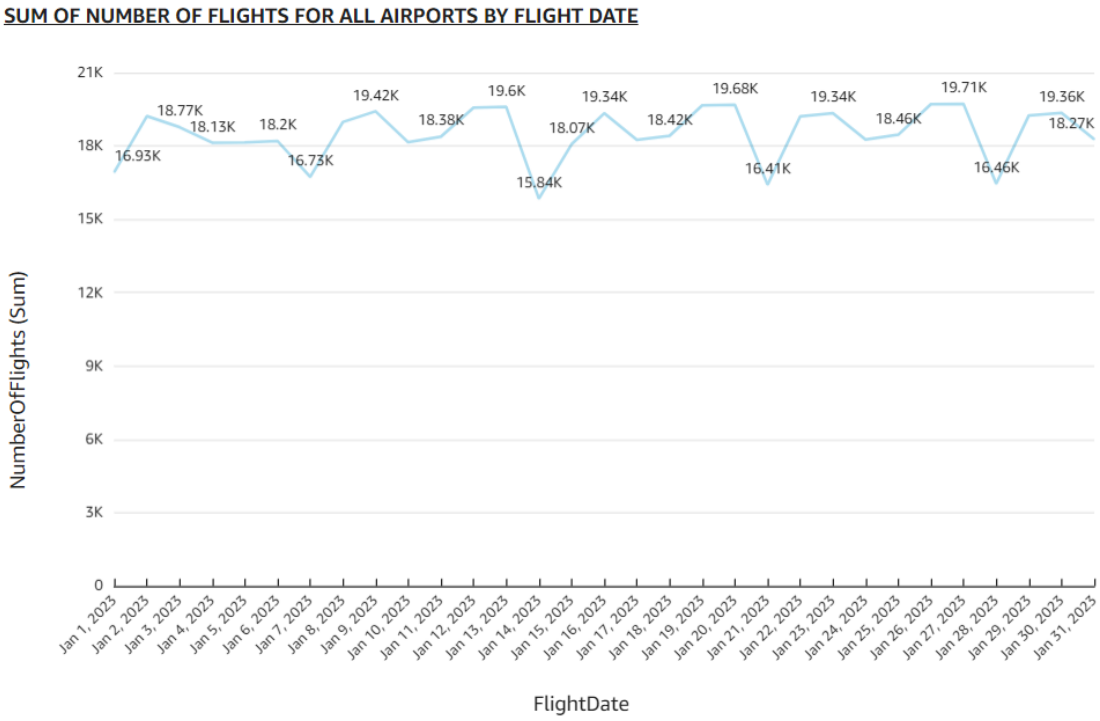


The data provided does not show that older planes, at least within the limited scope of this dataset, necessarily suffer more delays. In fact, the plane with the single highest number of flights as per tail_number had an average departure delay of "0". (consult question in .txt file)

Section 4: Identifying how the number of people flying between different locations changes over-time

Yet another problem was encountered in this case in that not a single field/column can be utilized to identify the number of people flying. There is no passenger field or anything remotely relating to that. Even combing through the ReadMe bore no results. Such a field simply does not exist and the author could not devise a clever way to parse out such information.

As a result, the output had to be notably reframed to **be about the number of flights between origin and destination locations across the one month period our dataset represents**. This data is however difficult and unsightly to visualize as there are far too many destination sets yet it can be viewed as a table utilizing SQL (question 4 of .txt file). Thus, it appeared prudent to narrow the results down to all flights and how their numbers change over time when creating a visualization:



The period represented by “over time” being the one month of 2023 we have to contend with. It however must be noted that the goal of this section of the analysis could not be reached. There is no way of truly knowing the number of people flying between different locations and how that changes over-time. This is merely an alternative visualization. **The true answer to this section of the analysis is that with the current dataset we can only know how many flights occur between locations over-time, not how many people are flying between different locations.**

Section 5: Identifying the percentage of flights delayed by weather per airport

The banner above illustrates the results perfectly. When a query was run utilizing the dataset (question 5 in .txt file) it was found that the percentage of flights delayed by weather per airport during the period of 2023 in question (1 month) 100% of flights were delayed by weather conditions. This seems surprising and almost unlikely, yet it is what the data shows when queried. It would be ridiculous to simply share a bar-chart of one hundred percents lined up hence the straightforward answer in large text above.

Conclusions & Actionable Insights

In summation, the analysis of the 2023 dataset aided in elucidating key trends despite the data's notable limitations. "AA" emerged as the leading airline with 138.9K flights, and a consistent industry-wide monthly trend was observed. Day 5 was identified as the number 1 day to minimize delays during January 2023 though it must be noted.

While challenges in data's accessibility hindered insights into the best time of day and seasonal variations, the examination of older planes revealed no conclusive evidence of increased delays. Improved data collection methods are recommended to capture critical details such as passenger numbers and airline names.

And finally, despite difficulties in quantifying passenger numbers, the analysis offered insights into the frequency of flights between various locations. Notably, a surprising 100% of flights were delayed by weather conditions, prompting a deeper investigation into specific weather events impacting airports.

Actionable Insights:

- **Data Quality Improvement:**
  Advocate for enhanced data collection methods to include crucial details like airline names and passenger numbers, facilitating more comprehensive analyses.
- **Deeper Delay Factor Exploration:**
  Further research into delay factors beyond aircraft age is recommended, considering maintenance practices, operational policies, and other external variables.
- **Better overall data is required:**
  Expand the dataset and analysis to cover multiple months and years and contain far fewer null values if possible for a more nuanced understanding of the data and travel patterns. It was quite difficult to work with what was provided.
- **Weather Impact Assessment:**
  Investigate specific weather events influencing the 100% delay rate observed, enabling better contingency planning for adverse weather conditions.

References

Reporting Carrier On-Time Performance (1987-present). (2023). Bureau of Transportation Statistics. https://www.transtats.bts.gov/Fields.asp?gnoyr_VQ=FGJ