A1: Analysis of Carrier On-Time Performance

Yoshiyuki Watabe

Executive Summary

This report summarizes the results of an investigation and analysis of what factors affect flight delays using the data on airlines' on-time performance. The data for January 2023 revealed that particularly large delays occurred on days with a day-of-week of 3 and on January 11. In addition, larger delays were found to occur on aircraft with more than 400 flights. In addition, two airports ('DEN' and 'DFW') were found to experience larger delays with higher frequency.

Most Operated Airlines in January 2023

The number of flights operated in January 2023, counted by airline, is shown below. The airline with the largest number of flights was WN, with 112,430. This represents 19.59% of the total number of flights for all 21 airlines, or 1.50 times the number of flights for the second largest airline, DL. There is a large difference in the number of flights between the top 5 and other airlines, with the top 5 accounting for 64.41% of all flights, indicating that there is a bias in the number of airlines used. Since the dataset only includes data for January 2023, it is impossible to analyze trends, but by obtaining time-series data, trends can be analyzed (ref. SQL code 1 and Fig.1).
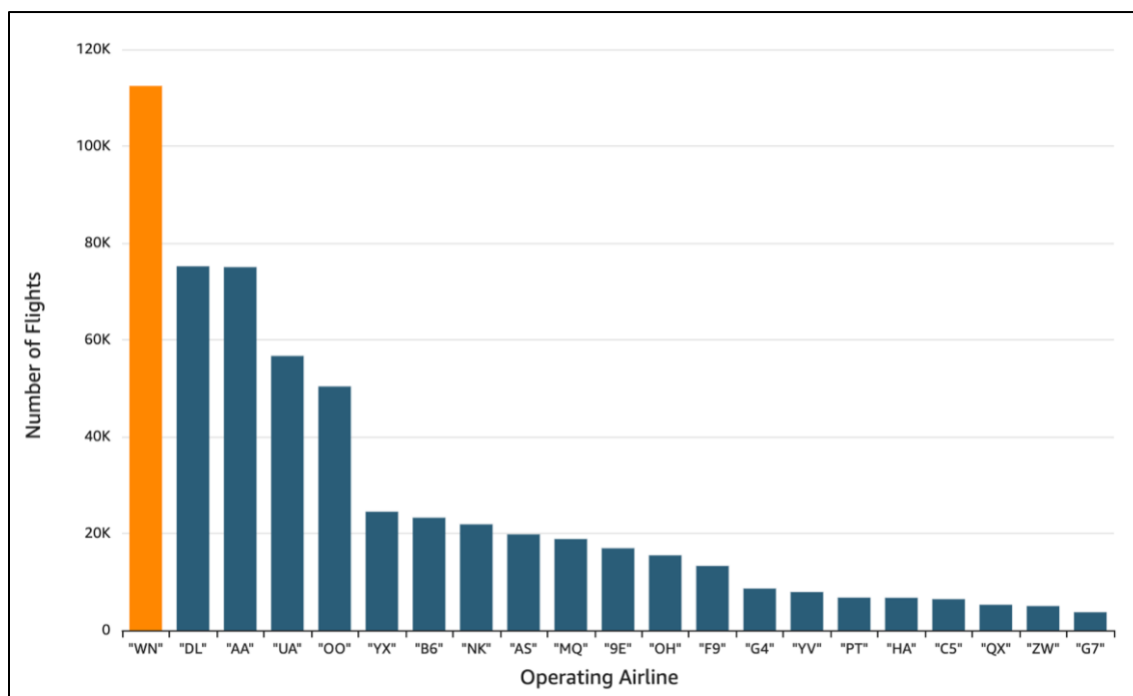

Fig.1 Airline Ranking of January 2023

Differences in Flight Delays by Date

<u><Day of Week></u>

The number of flights and average of arrival delay group for each day of the week are summarized below. The delay group is an indicator which value increases every 15 minutes and multiplying the value of delay group by 15 gives the approximate delay time. The average of arrival delay group is lowest on day of week 6, and the average of arrival delay group is highest on day of week 3, which is 2.87 times higher than on day of week 6. Additional analysis is required to determine why the average of arrival delay group is larger on day of week 3. Therefore, if delays are to be minimized, it is preferable to depart on day of week 6 (ref. SQL code 2_1 and Fig.2).
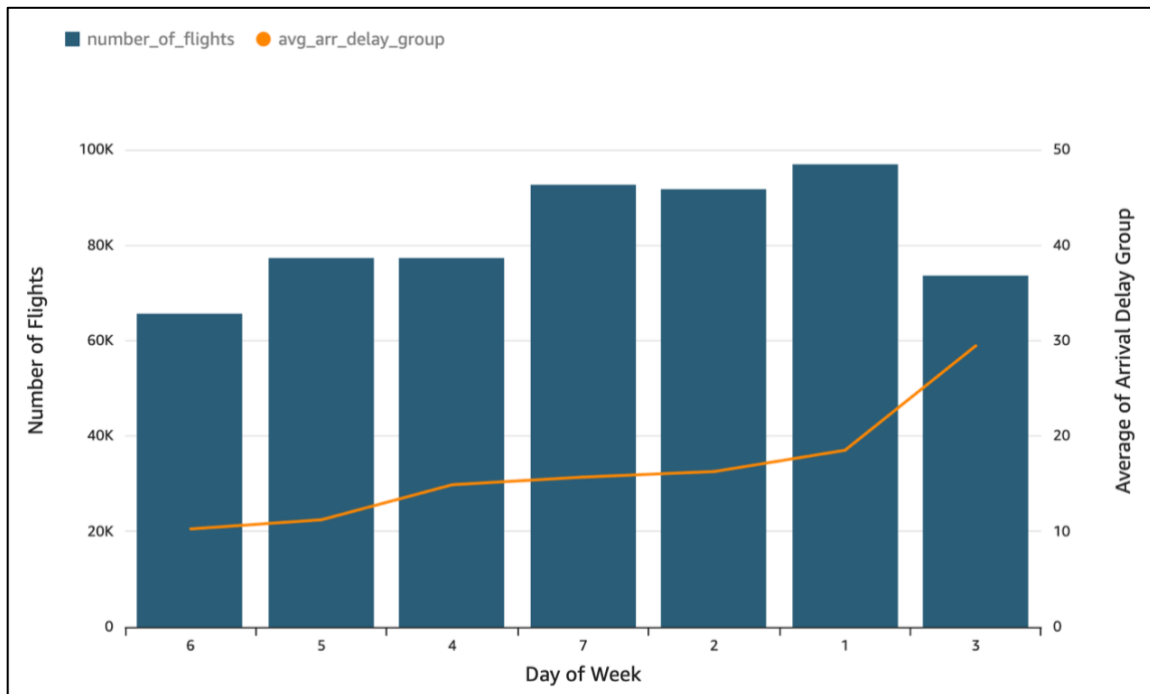

Fig.2 Number of Flights and Average of Arrival Delay Group by Day of Week

<u><Day of Month></u>

Similarly, the number of flights and average of arrival delay group are summarized by date of the month. The average of arrival delay group is lowest on days with a day of month of 17, while the average of arrival delay group is highest on days with a day of month of 11, which is nine times higher than on days with a day of week of 17. Therefore, if delays are to be minimized, it is preferable to depart on a day of month 17. Additional analysis is required to determine why the average of arrival delay groups is greater on days with a day of month of 11. (ref. SQL code 2_2 and Fig.3).
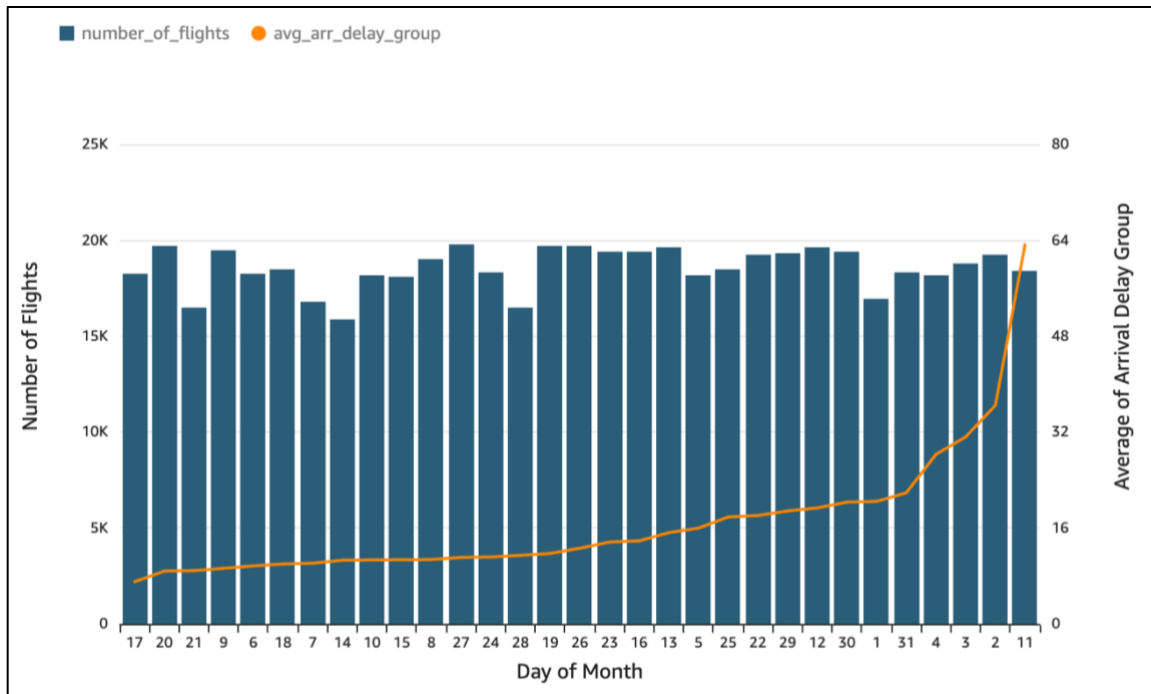
Fig.3 Number of Flights and Average of Arrival Delay Group by Day of Month

<u><Time of Day></u>

Since there was no information on specific departure times from the data set, it was not possible to analyze the time at which delays could be minimized. Collecting data on departure times would allow for analysis.

Relationship between Number of Flights and Flight Delays

The relationship between the number of flights and average of arrival delay group was then examined. Since information on the year of manufacture of the aircraft was not included in the data set, the number of flights is used to investigate the effect of long-term use on delay. First, bins were created for the number of times the flight had been in operation since its inception, in increments of 100. The corresponding number of flights and average of arrival delay group were then calculated for each bin. From the 0-99 bin to the 300-399 bin, the average of arrival delay group does not change significantly, and these bins contain 99.56% of the total number of flights. On the other hand, from bins 400-499, the average of arrival delay group begins to increase, and in bins 700-799, the value is 80.2, 6.35 times higher than in bins 0-99.

Therefore, to minimize delays, it should not be used for aircraft with more than 400 flights in service. If data on the year of manufacture of the aircraft is available, a more detailed analysis of the impact of time changes on delays can be performed (ref. SQL code 3 and Fig.4).
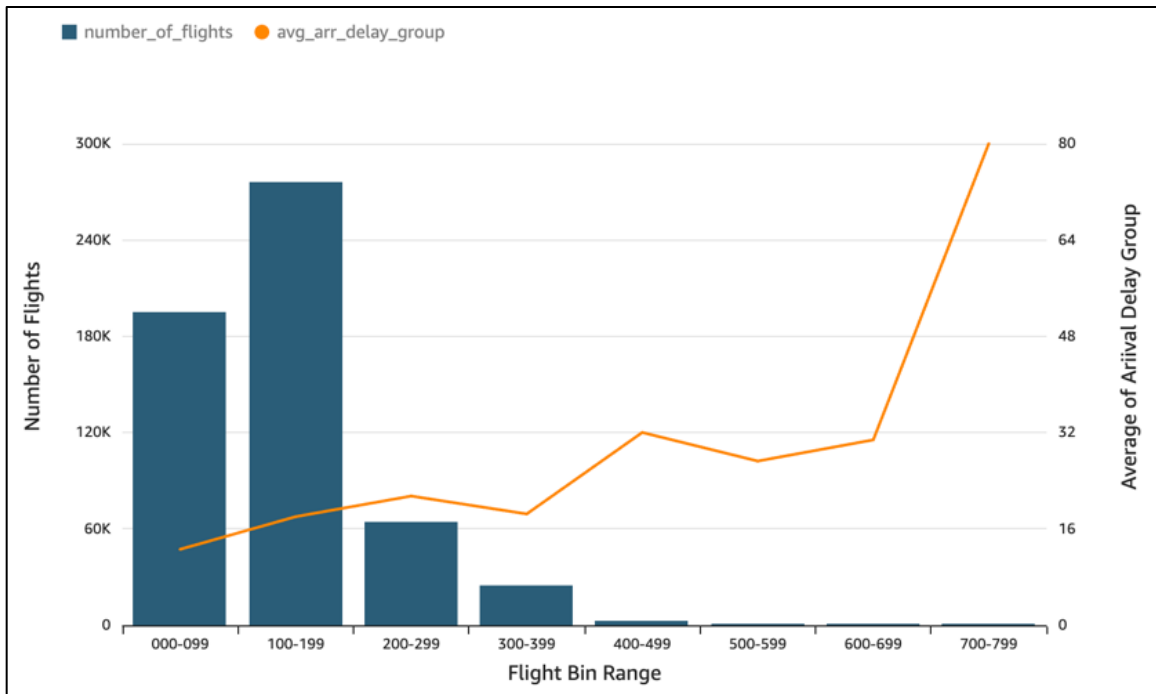
3

Fig.4 Number of Flights and Average of Arrival Delay Group by Flight Bin Range

Difference in Flight Distance for each Day of Week

The following is a summary of the changes in flight distance depending on the day of the week. Bins were created for every 100 miles, the number of flights falling into each bin was calculated, and the results were summarized for each day of the week.

The results show that for all days of the week, flights of 99 miles or less account for about 50% of the total, and if flights of 100-199 miles are included, they account for about 85% of the total. The most obvious difference is in the 600–699-mile bin, where day of week bins 1, 2, and 3 account for about 3-5% of the total number of flights, while day of week bins 4, 5, 6, and 7 decrease this value to about 1%.

The results show that no significant differences in flight distance were observed between days of the week, but that long-distance flights of 600-699 miles were more frequent when the day of week was 1, 2, or 3 (ref. SQL code 4 and Fig.5).

By obtaining time-series data, a similar analysis can be performed at different scales, such as monthly, quarterly, or yearly, which allows for more detailed analysis of flight distance trends.
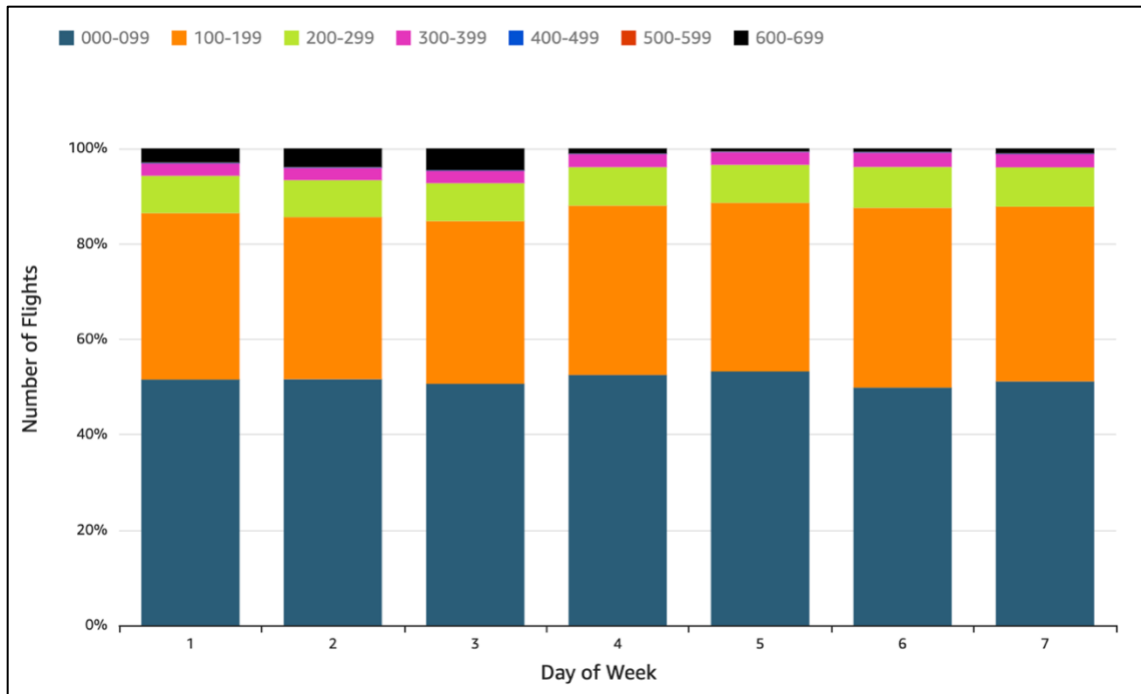
Fig.5 Number of by Day of Week and Distance Bin Range

Airports with the most delays

< Number of Routes by Route Delay Bin Range >

First, there is some weather delay on all flights included in the data set, and the values are listed as very small compared to other delay factors such as carrier delay. Therefore, without limiting the cause of delays to weather, we investigated which airports are more likely to experience delays. Index numbers were assigned to each route, and each average of arrival delay group was calculated and divided into 6 bins.

The results showed that 715 routes, or 12.36% of the total, had an average of arrival delay group greater than 50. Therefore, additional research is required on these two airports to determine why they are causing so many delays (ref. SQL code 5_0, SQL code 5_1 and Fig.6).

< The Number of Times each Airport was included as Origin or Destination>

Next, for the 715 routes with an average of arrival delay group of 50 or more, the number of times each airport was included as origin or destination was counted.

The results show that two airports, 'DEN' and 'DFW', were included in over 40 flights in one month. These two airports alone account for 6.15% of the 243 airports included in the average of arrival delay group 50+ flights (ref. SQL code 5_2 and Fig.7).
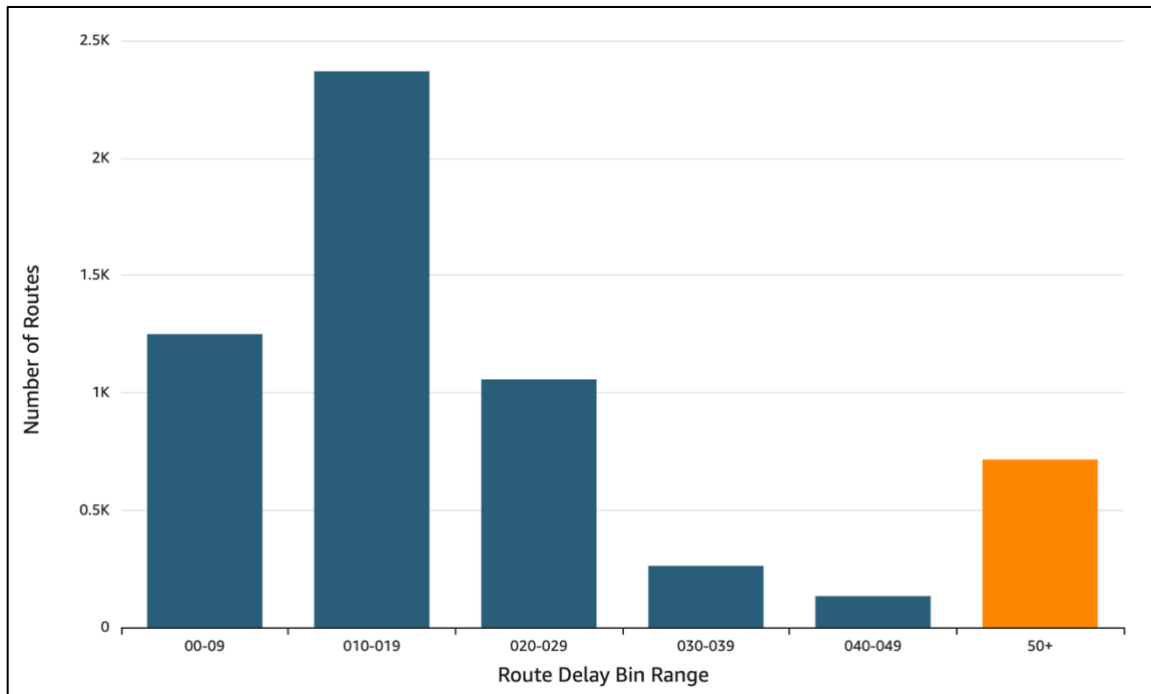
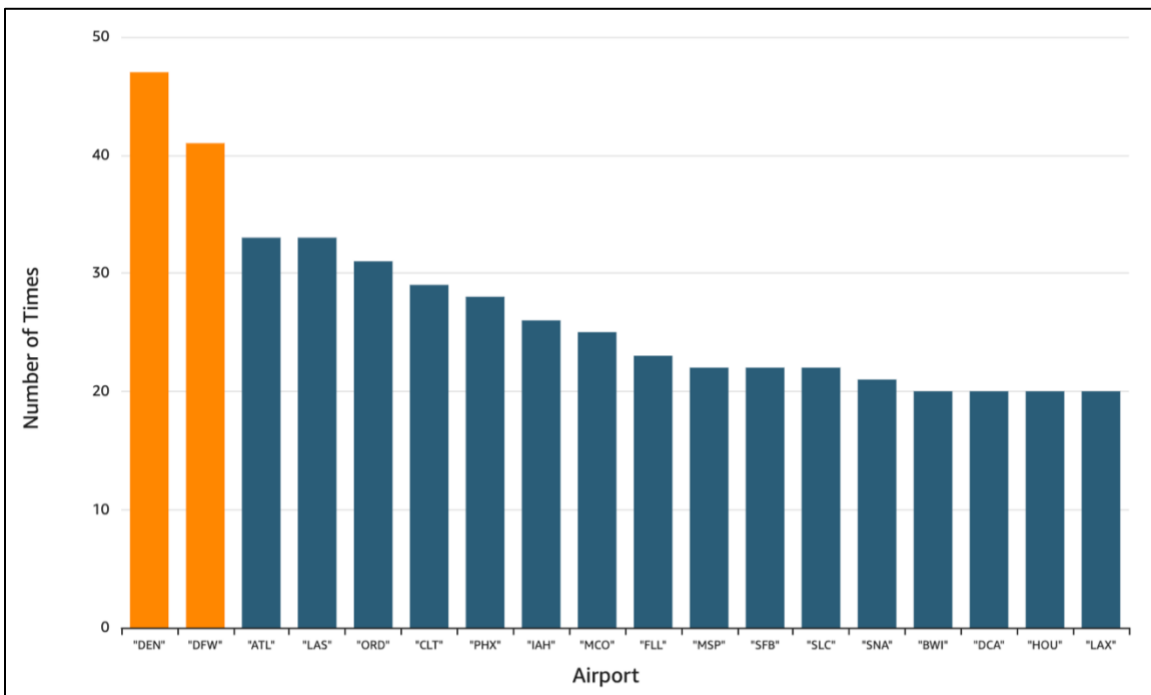Fig.6 Number of Routes by Route Delay Bin Range

Fig.7 Number of Times each Airport was used as Origin or Destination (>= 20)

Actionable Insights and Recommendations for Better On-Time Performance

1. Deeper Analysis about the Cause of Delay by Date:
   Additional research and analysis are needed to figure out why there were significant delays on day of week 3 and January 11, 2023.

2. Select the Aircraft that should be used:
   Avoid using aircraft with more than 400 flights as much as possible, as they tend to cause more delays.

3. Deeper Analysis about the Cause of Delay at Two Specific Airports:
   The airports of 'DEN' and 'DFW', where long delays are particularly frequent, should be additionally investigated and analyzed for their causes.