

Airline Overview

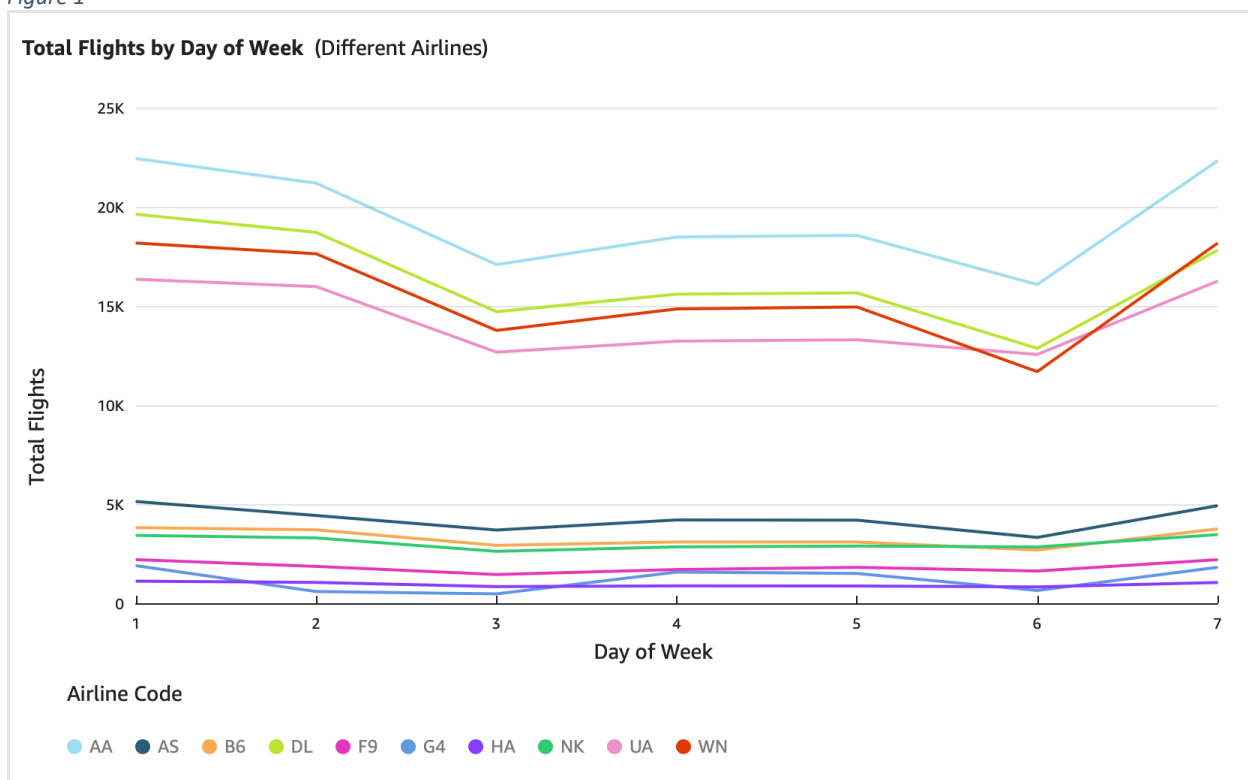
In the provided dataset on the on-time performance of flights, an in-depth analysis reveals that the most popular airlines are AA, DL, WN, and UA. Regarding flight delays, the data indicates that Fridays experience the lowest number of delayed flights. Furthermore, the hours between 11 p.m. and 4 a.m. see a significantly reduced number of delays. The airports with the highest percentage of weather-related delays are MQT, BJI, BRD, ABR, and VEL. However, the dataset does not include information on the age of planes. In hence, it is impossible to determine whether older planes suffer more to delays. Nevertheless, the data does reveal the most frequented routes, which include HNL-OGG, LAS-LAX, BOS-DCA, LAX-SFO, LGA-ORD, and JFK-LAX, each having over 1.5K flights.

Section 1: Rank Airlines by Year

According to the data that I have been given, I cannot tell that which airline has the most flights by year since the data itself includes only a month within a year. However, if we investigate flights by the day of week as shown in graph (Fig. 1), we can see that the top four popular airlines are “AA” (136129 flights), “DL” (114945 flights), “WN” (109196 flights) and “UA” (100291 flights), and which these airlines’ flights has significantly higher than the rest of the airlines. This means that these airlines are dominating the industry of airline.

In Fig. 1, we can understand how many flights each airline has. The top four airlines show that the total flights decrease on Sunday and increase on Saturday.

Figure 1



Section 2: Overall Delayed Flights

The best day of week (Fig.2) to fly is Friday since it has the lowest flights delayed (20.24K). The best time of day (Fig.3) to fly is from 11 p.m. to 4 a.m. since these hours has significantly low delay flights. For the best time of year to fly, I do not have enough data to answer this question since the data includes only a month of the flights' data.

Figure 2

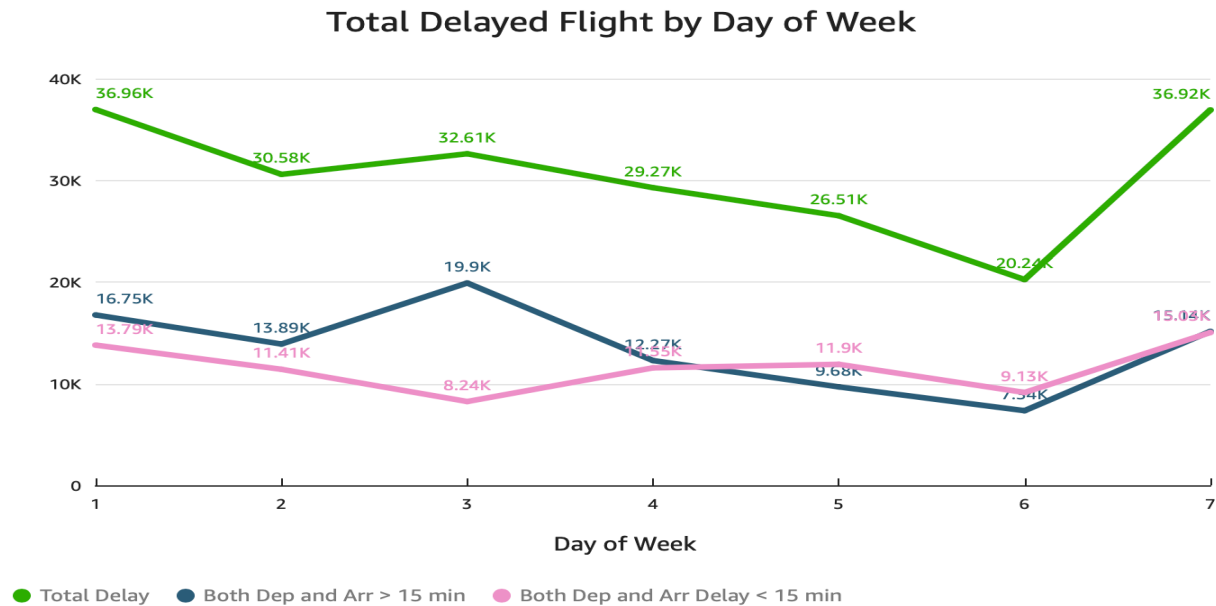
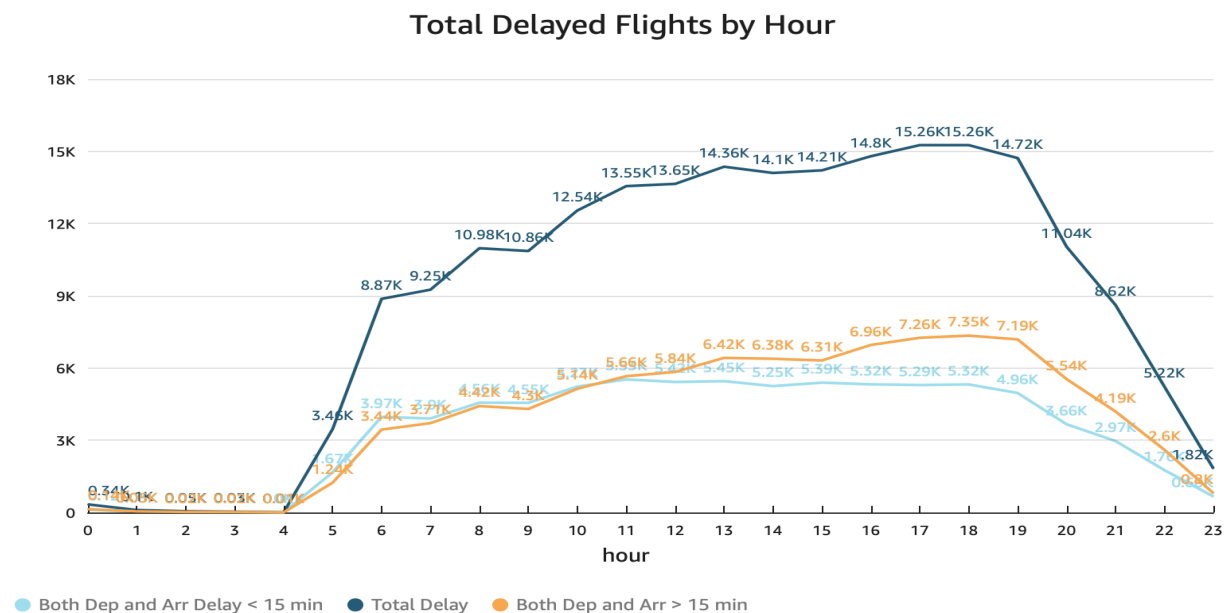


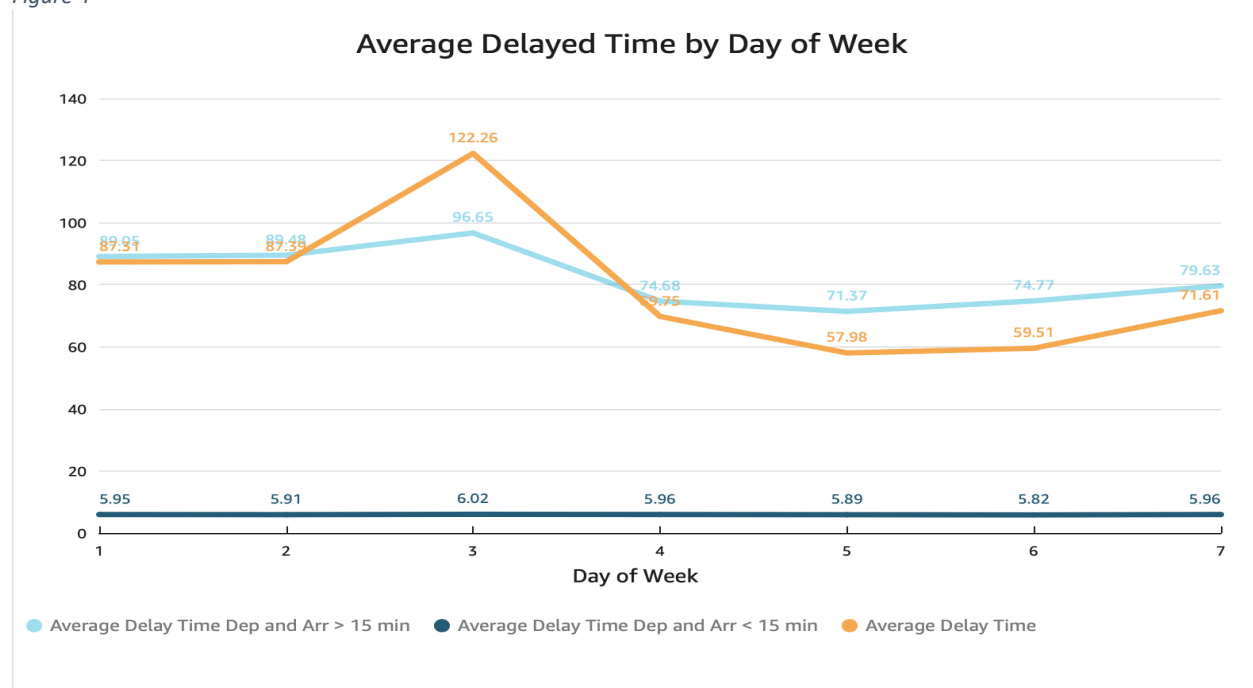
Figure 3



When we look at the average delayed time by day of week (Fig.4), when both departure and arrival delayed are less than 15 minutes, the average of the delayed time is around 6 minutes, while the average time of delayed when the departure and arrival delay time are both more than 15 minutes can be varied. The longest wait when both departure and arrival delay are more than 15 minute is 122.26 minutes which is more than two hours.

Moreover, Fig.3 shows that Saturday (36.92K) and Sunday (36.96K) have the most delayed flights and Friday (20.24K) has the least delayed flights which means if companies are trying to minimize the delayed time, Friday will be the best day to fly.

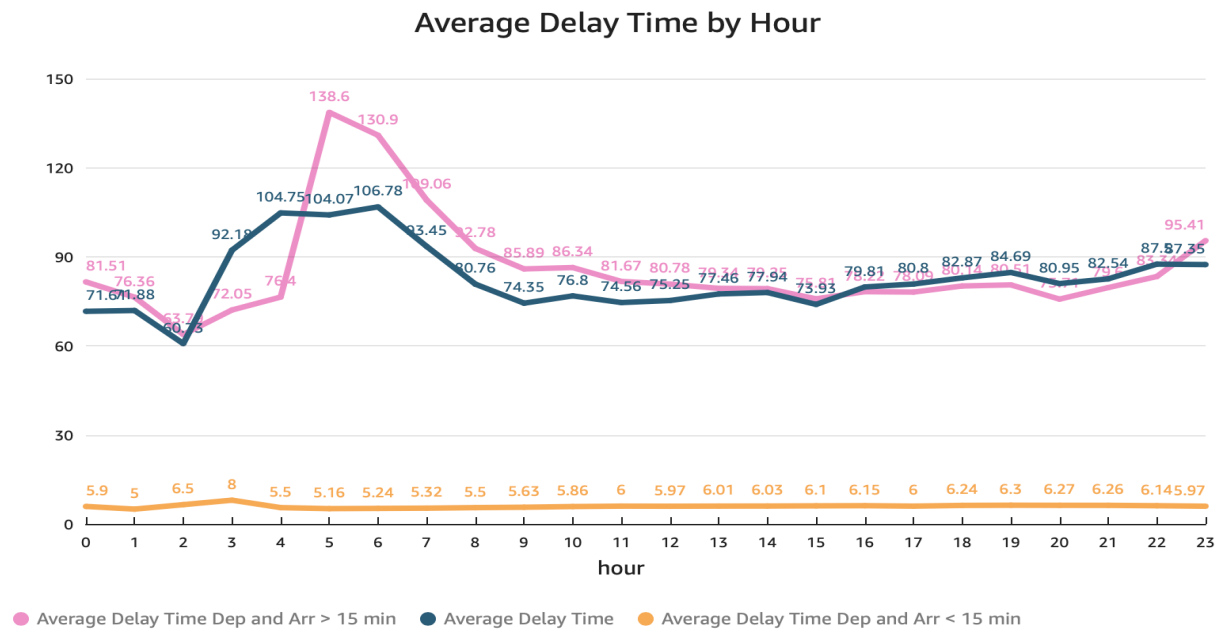
Figure 4



Furthermore, when we look at the delayed flights by time of day (Fig. 3), there is a significant increase after 4 a.m. up until 6 p.m. and starts to decrease until 11 p.m. Therefore, companies should have the flights departure after 11 p.m. until the next morning at 4 a.m. so that they can minimize the delayed flights.

When we look at the average delayed time by time of day (Fig. 5), we can see that if both departure and arrival delay are less than 15 minutes, the delayed time are mostly around 6 minutes. On the other hand, if both departure and arrival delay time is more than 15 minutes, we can see that there is a huge increase from 4 a.m. to 5 a.m. which also related to the previous graph meaning that this hour is the turning point of the day from off-peak hours to peak hours. In hence, companies should find out the reasons why there are an increasing delay during the peak hours and solve the problems to reduce the delay flights.

Figure 5



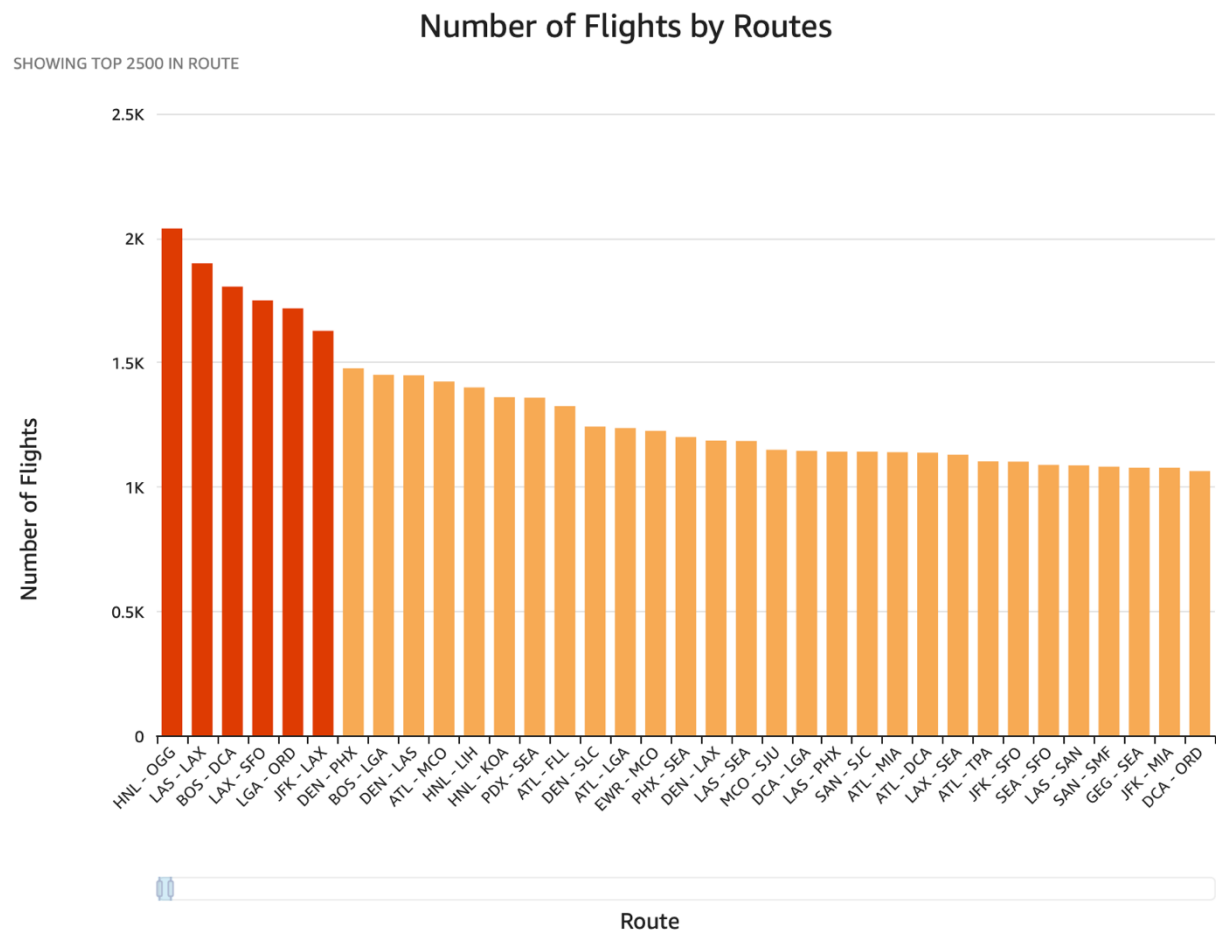
Section 3: Plane Age vs. Flight Delays

I am unable to determine whether older planes experience more delays, as the dataset provided does not include information about the age of the planes. If the data were to include the ages of planes, it would enable a comprehensive analysis to ascertain whether there is a correlation between a plane's age and the frequency of its delays. Additionally, understanding this relationship could be very important for airlines in planning maintenance schedules, optimizing fleet operations, and improving overall on-time performance. Without this information, however, any conclusions about the impact of aircraft age on delay frequency remain speculative. It would also be beneficial to consider other factors like aircraft model, maintenance history, and usage patterns, which together with age, could provide a more clearly understanding of flight delays.

Section 4: Changes in Passenger Numbers on Flights Between Various Locations Over Time

Since there are no data to clarify how many passengers there are in each flight, I am not sure about the changes in passenger numbers of flights between various locations over time. Assuming that each of the flights carries the same number of passengers, we can see that (Fig. 6) the top popular routes are “HNL-OGG” (2039 flights), “LAS-LAX” (1899 flights), “BOS-DCA” (1805 flights), “LAX-SFO” (1750 flights), “LGA-ORD” (1718 flights) and “JFK-LAX” (1628 flights), which means these are the most popular places and have more than 1.5K flights within a month where people travel in January in 2023.

Figure 6



Section 5: Flights Delayed by Weather per Airport

In the provided dataset, there is no indication of whether weather delays occur at the origin or destination airport. Therefore, the accuracy of the analysis might be uncertain. However, I have assumed that each weather delay affects both the departure and arrival airports if there is a delay. Under this hypothesis, the top five airports with the highest percentage of delays caused by weather are as follows:

Airport MQT: 16 flights weather delayed with 38 total delayed (42.11%)

Airport BJI: 13 flights weather delayed with 37 total delayed (35.14%)

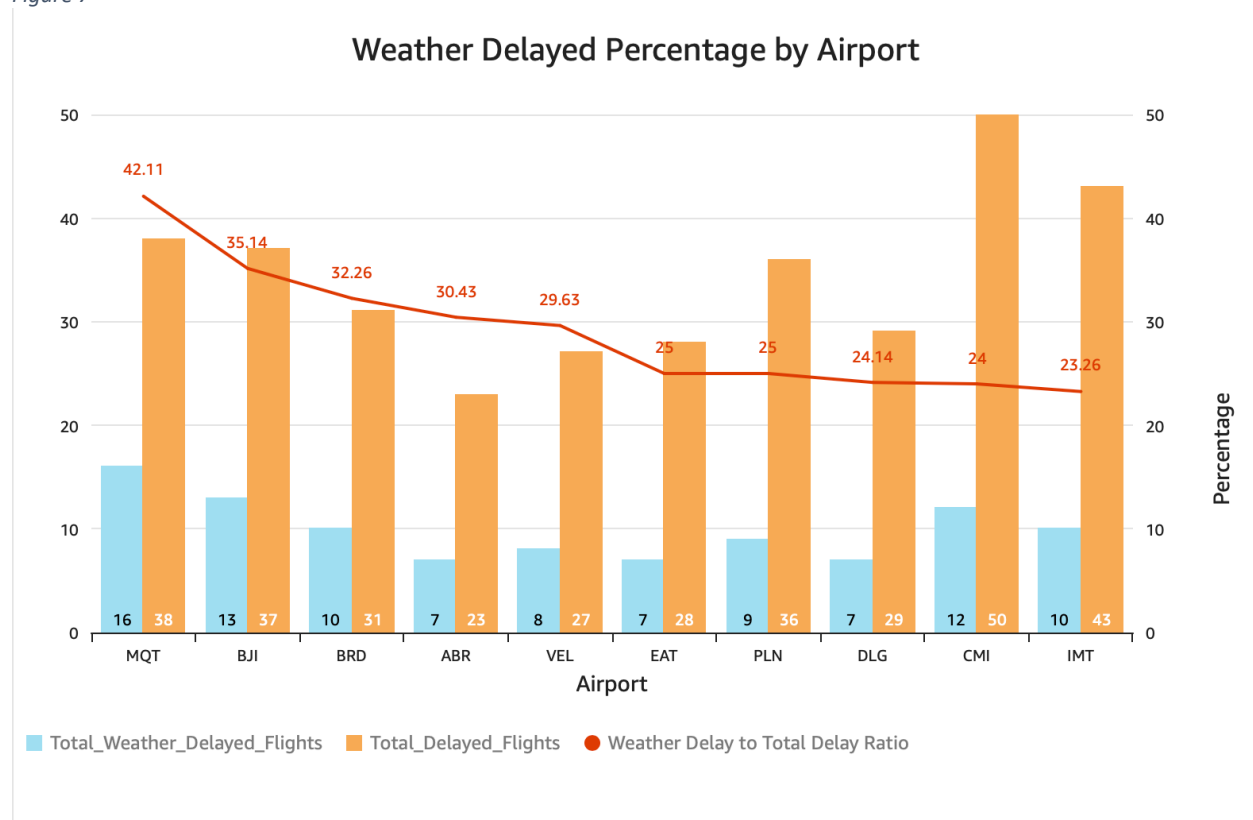
Airport BRD: 10 flights weather delayed with 31 total delayed (32.26%)

Airport ABR: 7 flights weather delayed with 23 total delayed (30.43%)

Airport VEL: 8 flights weather delayed with 27 total delayed (29.63%)

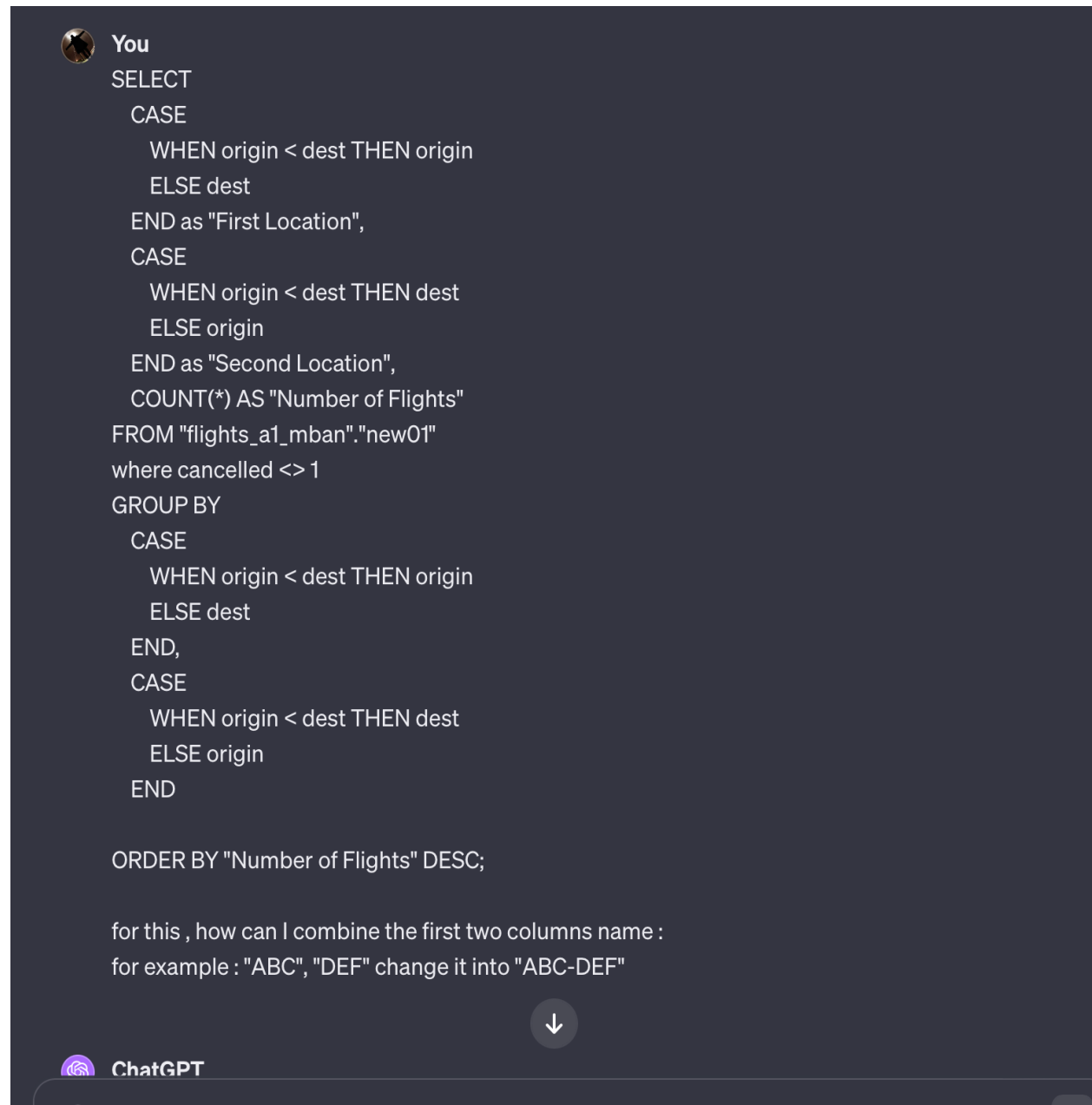
According to this graph, a higher percentage of weather delays at an airport does not necessarily correlate with a greater number of flights. This indicates that airports with more flights do not inherently experience a higher percentage of weather-related delays.

Figure 7



References

OpenAI. (2023). Personal communication [with ChatGPT-4]





ChatGPT

To combine the values of the "First Location" and "Second Location" into a single column with the format "ABC-DEF", you can concatenate these values within your `SELECT` statement. Here's how you can modify your SQL query to achieve this:

sql

Copy code

```
SELECT
  CASE
    WHEN origin < dest THEN origin || '-' || dest
    ELSE dest || '-' || origin
  END as "Route",
  COUNT(*) AS "Number of Flights"
FROM "flights_a1_mban"."new01"
WHERE cancelled <> 1
GROUP BY
  CASE
    WHEN origin < dest THEN origin || '-' || dest
    ELSE dest || '-' || origin
  END
ORDER BY "Number of Flights" DESC;
```



You

if it is a weather delay how will I know if its a departure delay or arrival delay



ChatGPT

In the context of analyzing flight data, distinguishing whether a weather delay is a departure or arrival delay typically depends on the specific columns available in your dataset. Most standard flight datasets, like the one you're working with, provide separate columns for departure and arrival delays. However, the indication of whether a delay is due to weather is usually only given for the departure.

Here's how you can interpret it:

1. **Departure Delays:** If your dataset has a column specifically for weather delays (like `WeatherDelay`), it typically refers to delays at the departure airport. This is because weather conditions at the origin airport directly impact the ability of the aircraft to depart.
2. **Arrival Delays:** Arrival delays can be caused by a variety of factors, including weather, but they are often not explicitly broken down by cause in standard datasets. If the dataset includes an arrival delay column (like `ArrDelay`), it might not specify if the delay was due to weather.

In your case, if you have a `WeatherDelay` column, it most likely refers to weather-related departure delays. To analyze weather-related arrival delays, you would typically need to infer this from overall arrival delay data and possibly cross-reference with weather data for the destination airports at the time of arrival, which can be more complex and may require external data sources.