Introduction

For this project, we wanted to analyze the effect of the covid-19 pandemic on the aviation industry. For this I used the OpenSky Network's public datasets (https://opensky-network.org/data/datasets) to plot data from 2019 to March 2022 and visually represent the changes in flight volume over that time period. I also looked at the changes in the price of jet fuel during this same interval, and the sizes of aircraft in use.

Data Wrangling

The first step in setting up this project was to connect to the OpenSky Network's live API, which provides live airspace information for research and non-commercial use. I called the json, cleaned up the dataframe, and then joined it to the metadata dataset on the icao24 number, so I would be able to reference the metadata for every flight in the air, via the unique icao24.

I then pulled every file provided in the OpenSky covid-19 dataset, which provided historical daily flight data in a series of one-month csv files, merged them into one large dataframe with all flight data from 2019 to March 2022, and cleaned up the column names.

I also created a dataframe that provided WTC (wake turbulence category, a way of categorizing aircraft size) by icao code. This data came from AvCode's Aircraft Type Codes list (https://www.avcodes.co.uk/acrtypes.asp).

All of these dataframes were then uploaded to Snowflake, so further analysis could be done using SQL in DBT. Within DBT, I created staging models of all my datasets, and then used those to create the models I would use for my data visuals. I created datasets for the outgoing flight volume of the top five airports in every inhabited continent, a list of the aircraft currently in the air (according to the latest pull from the api), a dataframe of historical flights over time that includes the WTC code (both worldwide and USA-specific), and the total count of arrivals and departures per airport per year.

All of these were then uploaded back to Snowflake as new tables, which were called from a jupyter notebook via snowflake connector.

Data Analytics

To start, I plotted the outgoing flight volume for the top five busiest North American Airports (figure 1), which do all happen to be airports in the US. As can be seen in the figure below, flight volumes dropped dramatically in March 2020 due to the covid-19 pandemic, and, as of late 2021, had nearly returned to pre-pandemic levels, although there is another, smaller dip starting in January of 2022, likely due to the emergence of the omicron variant of covid-19.

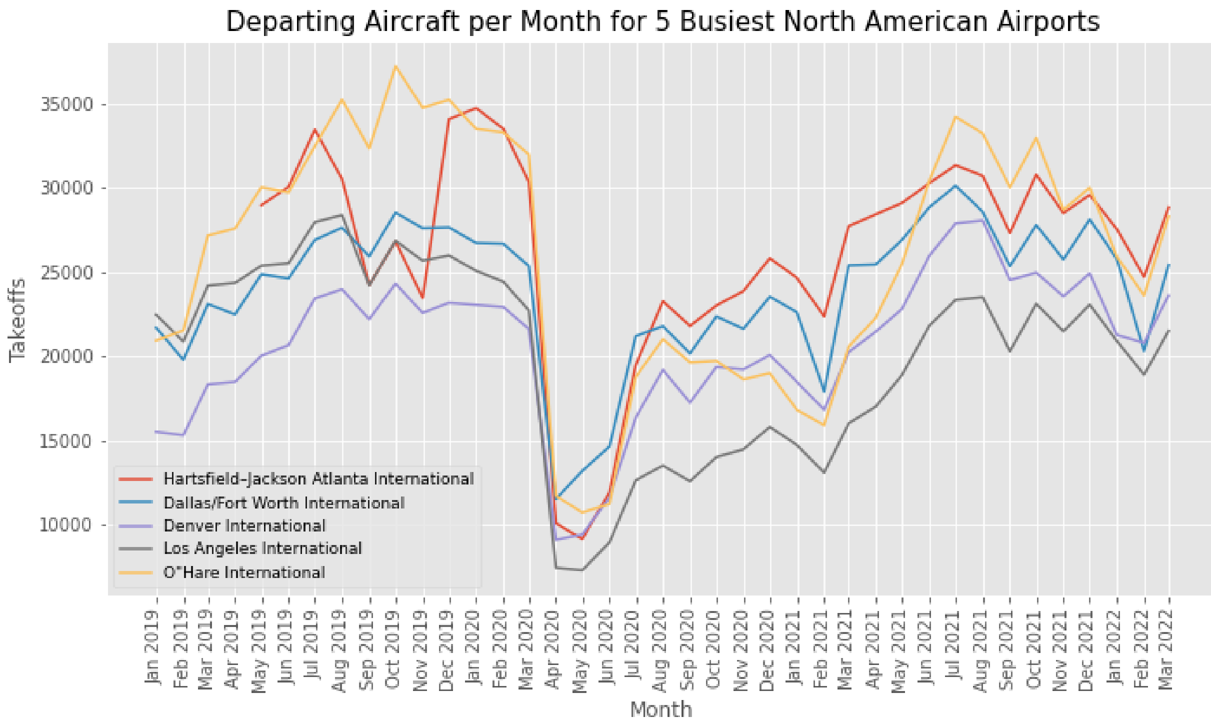

Figure 1

Data from the top five European airports (figure 2) shows a similar trend, with flight volume cratering in March of 2020 and reaching its low point in April 2020. Interestingly, the European airports recovered at a similar rate to the US airports, but then had another, wider dip from August 2020 to May 2021. There are a few reasons this might have happened, but the most likely seem to be that many countries implemented international travel restrictions in response to the covid-10 pandemic, and while the US has a large proportion of domestic flights, just by virtue of being a large country with little high-speed rail, european flights tend to be international, so the travel restrictions would have a greater impact on flight volume of european airports.

Another interesting data trend here is the sudden decline of air traffic at Sheremetyevo International Airport starting in February of 2022. This, of course, corresponds to the beginning of the Russian invasion of Ukraine.

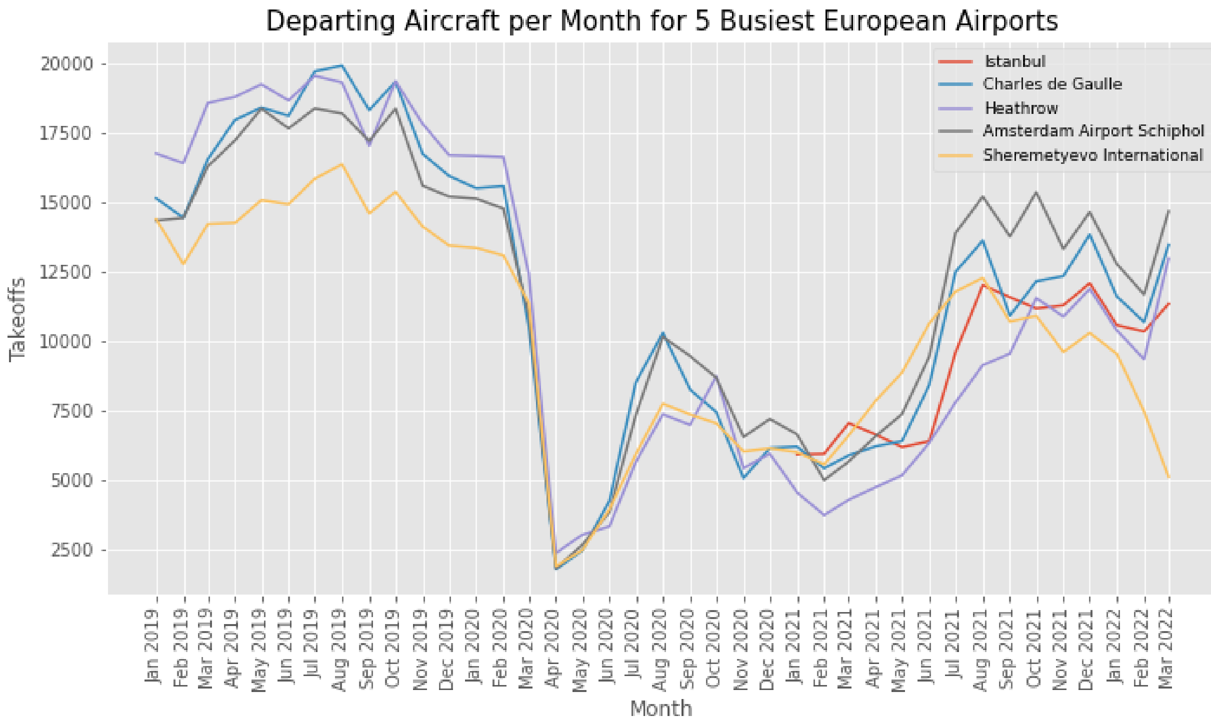

Figure 2

The next three datasets I will cover simultaneously, since the data is much less clean but still shows some basic trends. The flight volume plot for South America (figure 3) clearly shows the same dip in March-April of 2020 that we saw in the North American and European plots, followed by a modest recovery until the data gets extremely messy in August 2021. The flight volume for Africa (figure 4) shows a dip in response to the covid pandemic, although their minimum occurs in June instead of April. The flight volume plot for Asian Airports (figure 5) shows that its decline in response to covid-19 begins in February of 2020 rather than March (though the low point is still in April).

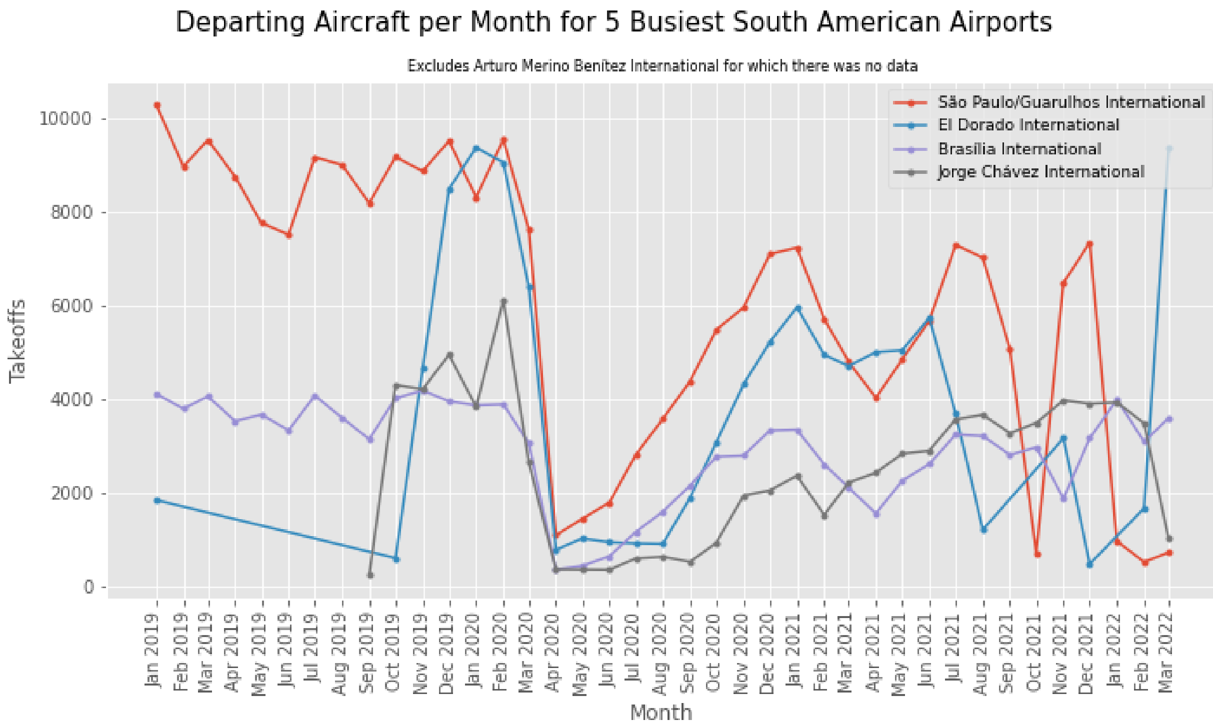

Figure 3

Figure 4

Figure 5

Figures 6 & 7 show the overall flight volume for the busiest five US airports (figure 6) and all world airports (figure 7). These clearly show again that the aviation industry has recovered much faster in the US than in the world more generally, with the US almost returning to pre-pandemic levels of air traffic, while the world more generally is at about 2/3 of previous levels. These graphs also separate out the flights by aircraft size (based on WTC classification) to show the proportion of aircraft that are affected most by the changes in flight volume. Initially, it was expected that with the rising price of jet fuel (figure 8), aviation companies might be switching from heavy to medium aircraft, however, the plots do not bear this out, likely because since an airplane carries a large upfront cost, there is a lag between changes in market forces and the response to that change. Interesting further analysis could be done by tracking the proportion of medium-aircraft over time as US jet fuel prices continue to rise, and noting how long it takes to respond, if it does at all.

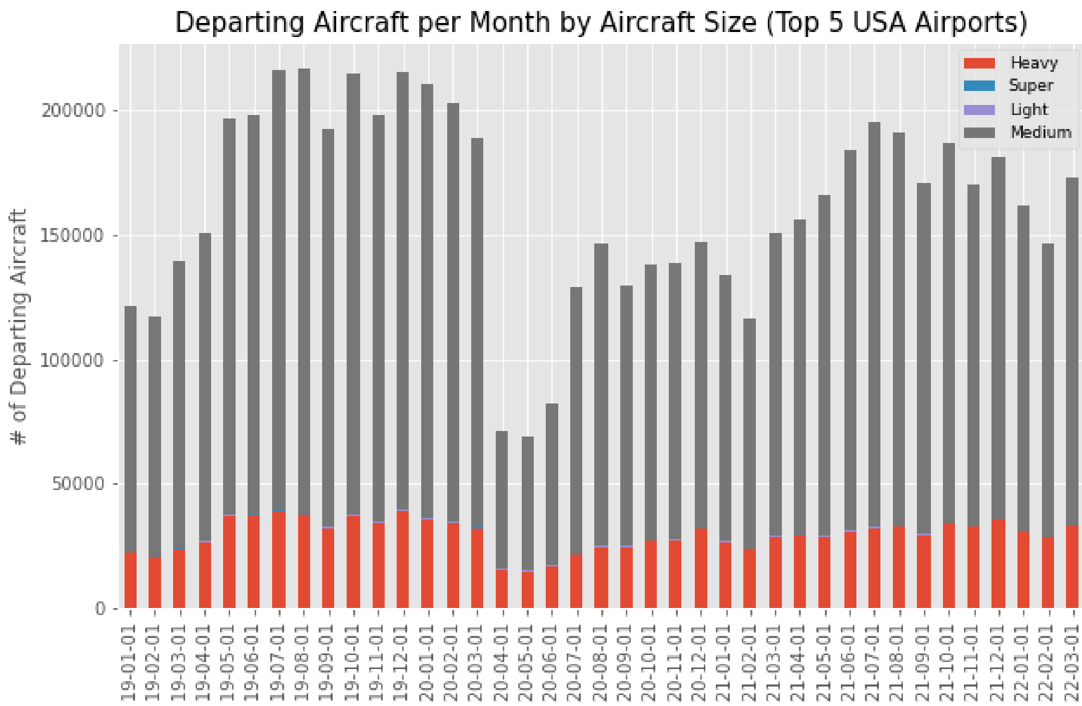

Figure 6

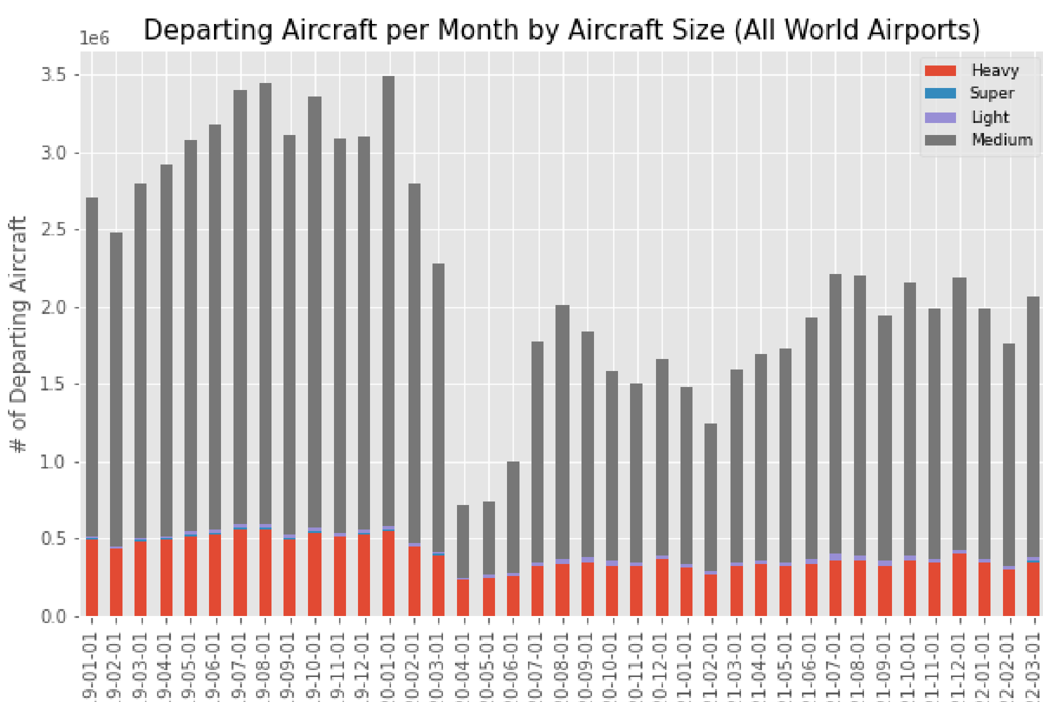

Figure 7

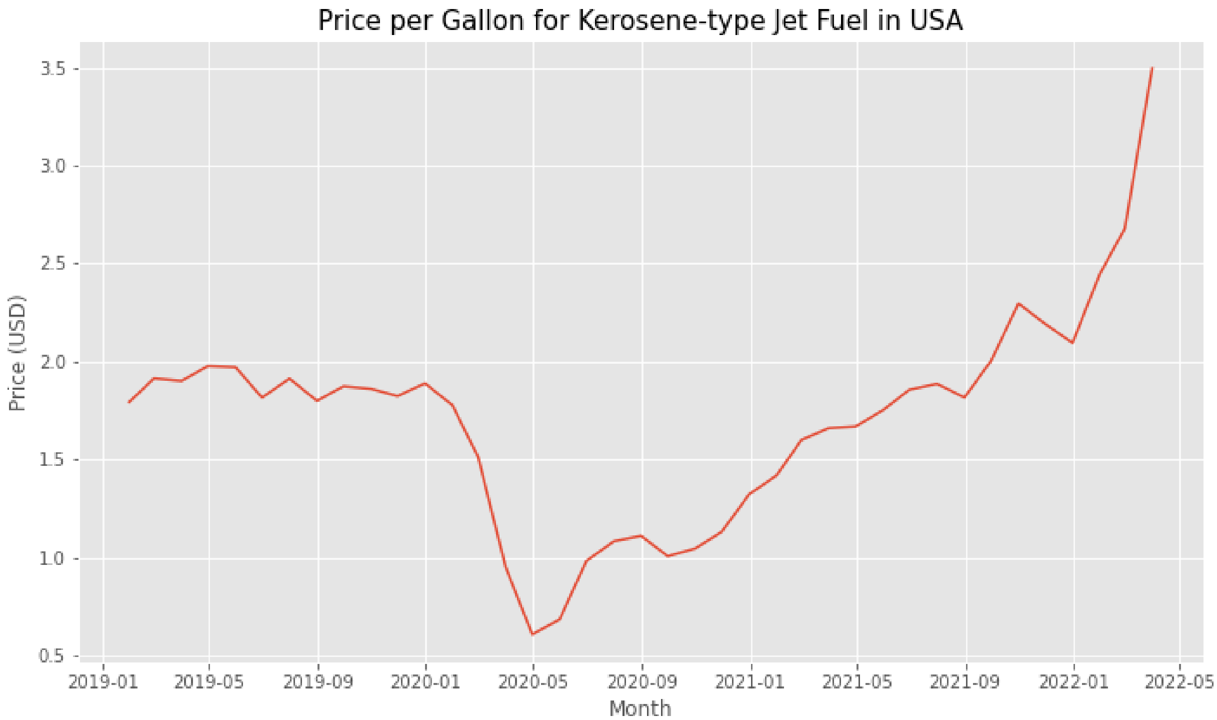

Figure 8

Next Steps

As time allows, I'd like to confirm the "US flight volume rebound is driven by domestic flights" theory by analyzing the proportion of US flights that are domestic vs international, and then comparing volume of US domestic flights to non-US flights of similar length, to see how each was impacted by the covid-19 pandemic. I could also limit the comparison group to only include european countries with high-speed rail systems to see if that is a cause of the reduced rebound in europe.