Dataset Analysis – Flights Timing

By Dotan Entin

Introduction and Assignment Goals

This assignment's goal is to analyze the flights timing of a number of American airlines, originated or destined to New York on March 2013.

The data was manipulated in Python 3 through Jupyter-Notebook and visualized in Python & PowerBI.

Data source and description

- flights_data.csv file – raw flights data including full flight data of more than 48,000 flights on March, 2013
- airports-lookup.csv file – airports and airports codes
- flights_db.csv file – summary file of the data from which we can collect the flights' company names and some basic statistics.

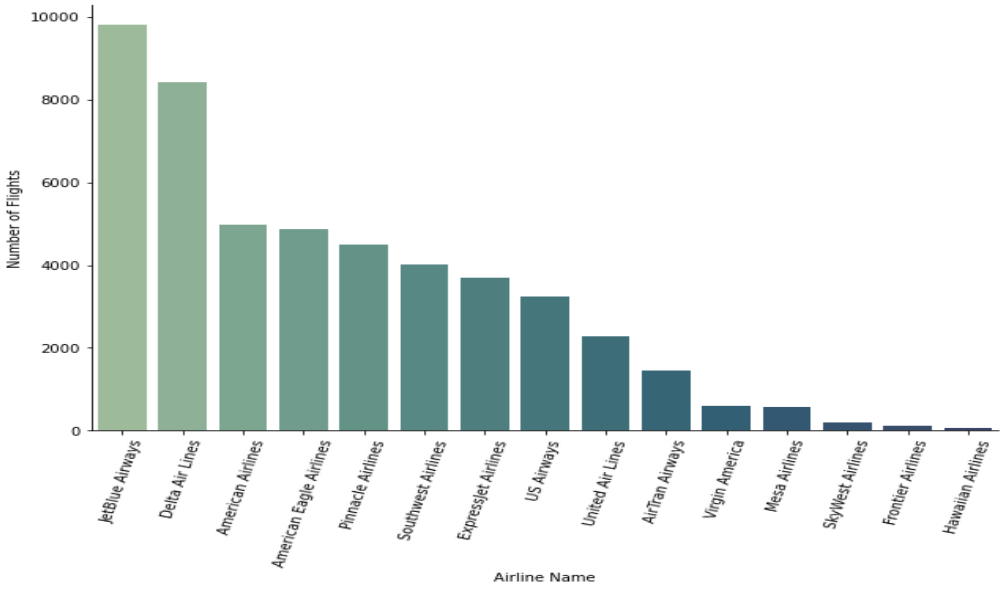
Table of contents

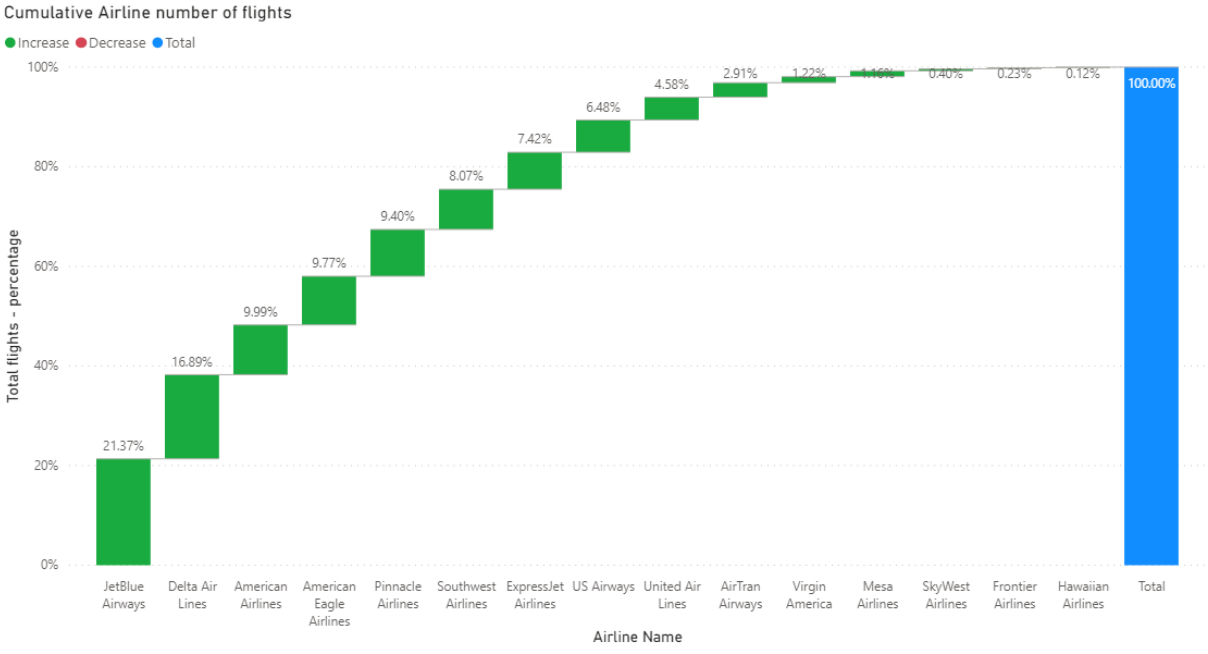
Which airlines are the leading flight operators?

The first question I was asking myself was a basic one: which airlines are operating the most flights in our database?

We have a total of 15 airlines recorded in our database for March 2013.

As we can see in the graph below, "JetBlue Airways" is leading in the amount of flights, with nearly 10,000 flights in March 2013, which is about 20% of the total flights operated this month. Next, we will find "Delta Airlines" with a bit over 8,400 flights which is around 17% of the total flights, and the rest are further behind, with "American Airlines" in the third place with nearly 5,000 flights and a bit over 10% out of the total flights in the database. Those 3 airlines operate 47% out of all the flights.

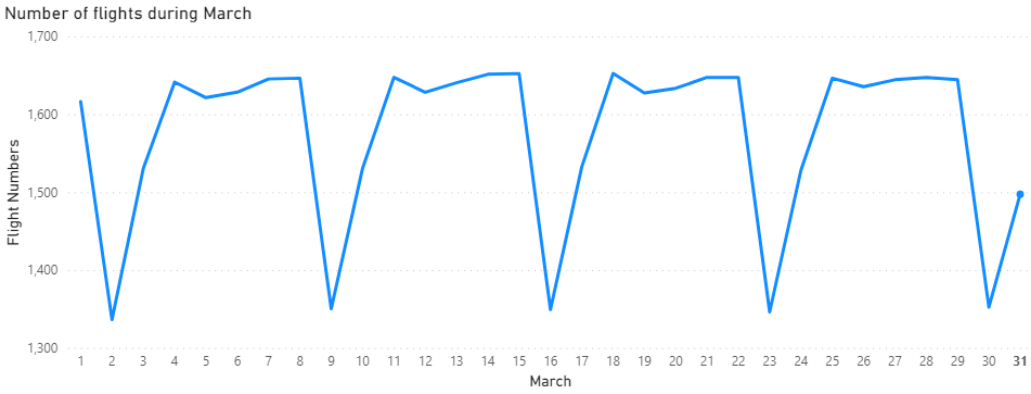




What can we learn about flight timing?

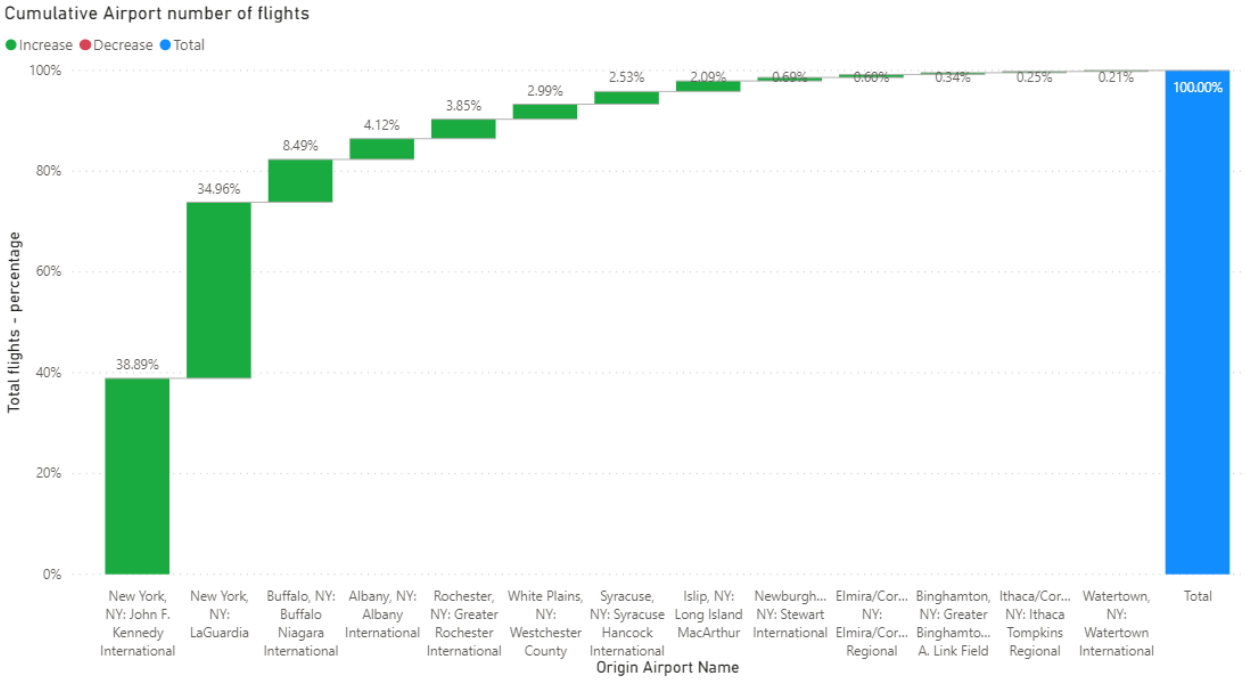
In our database, we have flights that operated on March 2013.

An interesting point is that if we look at the whole month, day by day and with the number of flights per day, we can see that the number is quite stable (between 1600 to 1650 flights a day) other than on Saturdays, where the numbers of flights descend by about 20% and raise back right afterwards:

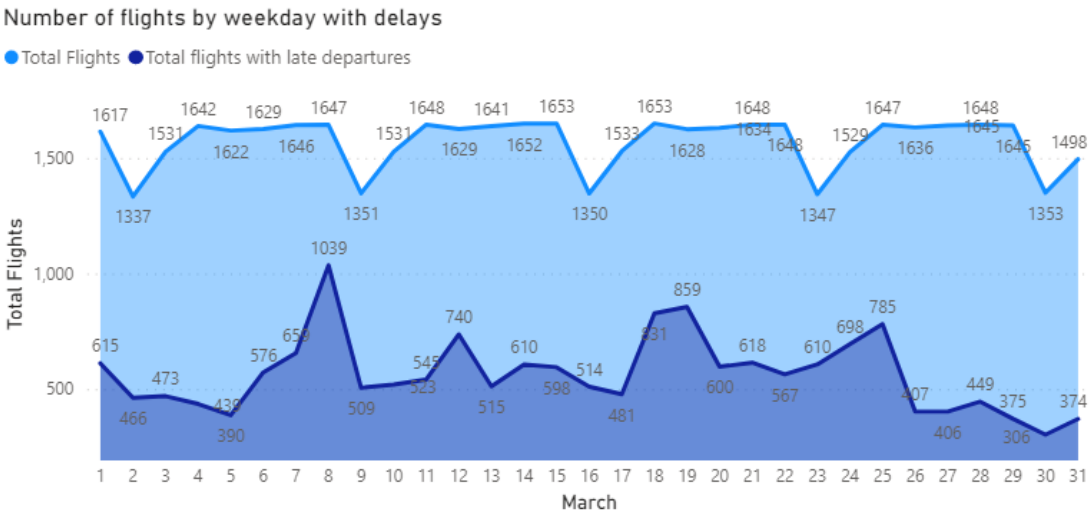


This means that Saturdays are the weakest day in terms of flights traffic.

Another interesting insight is that almost 75% of flights are operated by just 2 airports in New York – JFK and LaGuardia



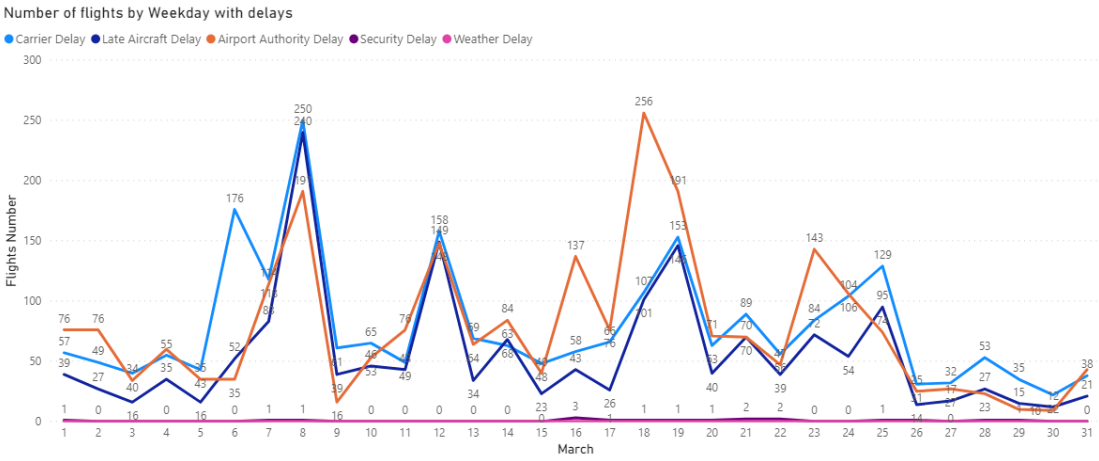
Let's focus on them as they hold the majority of the outgoing flights.



This graph shows us the total number of flights operated in March (upper blue line) with the total of delayed flights by late departure (lower blue line).

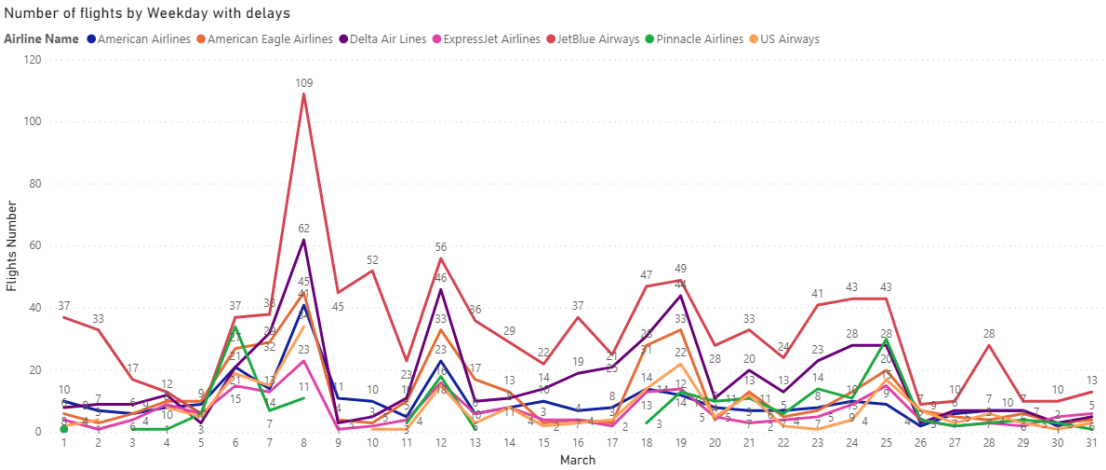
We can see that there is a spike of delayed flights on the second Friday of the month (day 8), that is slowly being improved by the end of the month (from day 26).

If we drill down to the reasons, we find that the spike on the 8th was created mostly due to Carrier & Late Aircraft delays, while later in the month it was due to Airport Authority delays:



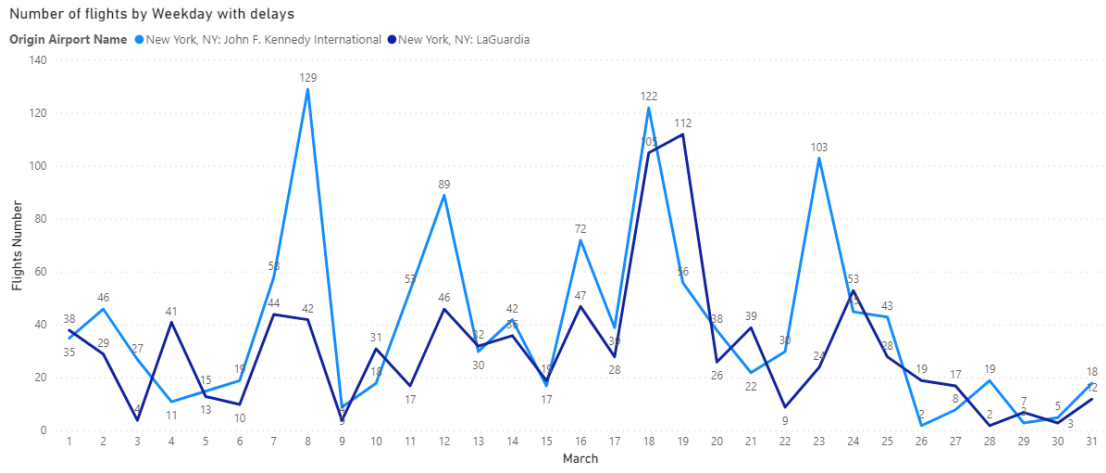
Security and Weather delays are almost non-existent in comparison to the rest.

Let's try to see if there is an indication if one of the airlines was causing the spike more than the others.



Indeed, we see that JetBlue (our highest flights operator) was inflicting the spike more than Delta (the second highest). This does show that JetBlue fixed the delays problem during the rest of the month, and they have managed to lower the numbers to be like the other companies which indicated of most likely a specific problem that they had on this day.

The other spike we saw was on the 18th, where we had an abnormal number of flight delays caused by Airport Authority. Finally, let's drill down on this one and see a graph with our airports and delays caused by Airport Authority:



It does seem that on the 18th, both airports have had many delays caused by their authorities, but as it seems over the month, JFK airport had more flights delays.

Note that going closer to the end of the month, the numbers descended drastically in both airports, and on some occasions, JFK had less delays than LaGuardia.