Flights dataset

Each row of the dataset contains the number of flights made in August 2020 by specific airport and carrier and all different reasons of delay for these flights.

Our dataset contains the number of cancelled flights and 6 different delay reasons for each one of 17 different carriers and 355 airports.

Flights delay time :

arr_delay variable represent all the delay (in minutes) that was caused by the five reasons (carrier-weather-NAS-security-late aircraft).

Delay time is directly proportional to the total number of flights. So, when we study all delay variables we should take into consideration the number of flights.

Security delay is very low and almost not effective to the total delay. The remaining four reasons are effective and their distributions are close to each other with slight increase in carrier delay.

Carriers, their delay and cancelled flights :

One of the things we can notice that the most number of cancelled flights lie with HA carrier although HA is the least one existed in our data with very low carrier delay. So, it is clear that the carrier delay is not the reason behind cancelling the flights. Another case that ensures our conclusion is OO carrier that existed the most in our data with slight large median and maximum carrier delay but a very low cancelled flights.

All plots ensures our earlier conclusions that carrier delay is directly proportional to carrier_ct but cancelled flight in each carrier does not depend on the two variables.