
TASK 1

In this task I will calculate the average flight delays in the dataset.

Note 1: The code was written in Jupiter's Notebook and executed through there and the console. All figures are from the console. Configuring the max memory in 14g was necessary for the project to run successfully with Jupiter's Notebook.

Note 2: In order for the code to run successfully with any given path a variable path was created in which you must declare the path of the file in your local PC.

CALCULATION

For the calculation of the average delays all available delays were used. There are 7 types of delays; delay on the departure time, arrival time, delay due to carrier or weather, NAS delays and security delay. As it turned out the average delay is 223.79334577647123. SparkSQL was used and the result in the console can be seen in task1.png.