

**Dublin Business School**

**Continuous Assessment – Programming for Big data**

**CA4- READ FILE IN PYTHON**

**Rafael Pugliesi 10345243**

From a dataset which contained 422 comments about changing some code, my task was filter that data to extract all the valuable information and then analyze that data with my personal observations.

To complete this specific task I choose to analyze some specifics points. I decided for this case extract only time (day, month, year and time), day of the week (e.g. Monday), author of those comments, also I decided to extract number of lines in each comment.

I chose to work with classes In python, I used some code already created in class, this code was a python language code, and after some changes was able to read a file line by line, extract any information I want, and output this extracted information.

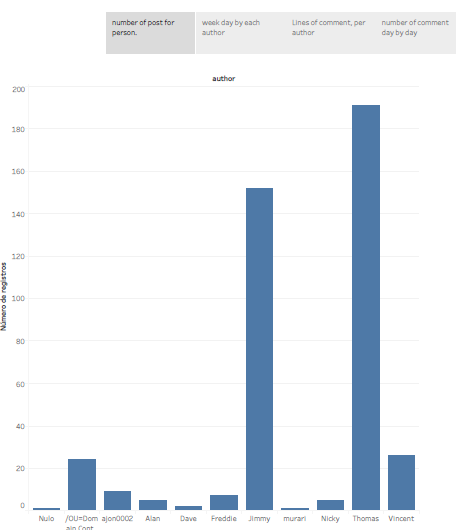
So I set some parameters in the code, and ran it. As mentioned an output file in CSV was created, in this case called “ARQUIVOOUTPUT.CSV”.

With this new data extract I decided to plot some graphs to make an easier visualization of this information.

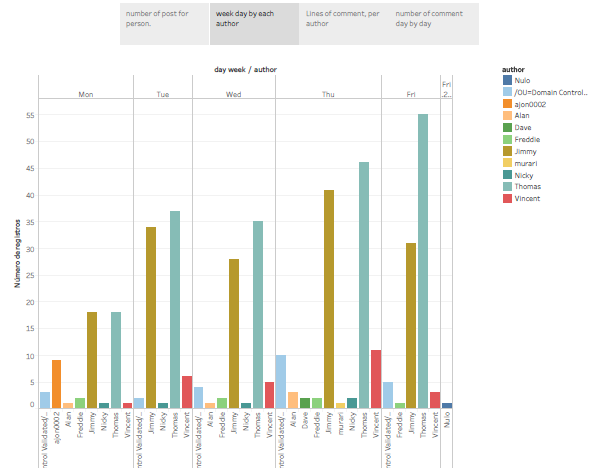
To plot this data I choose to use a tool called TABLEAU, which is a powerful tool to business intelligence and data and big data analytics.

Follow below some graph about information extracted.

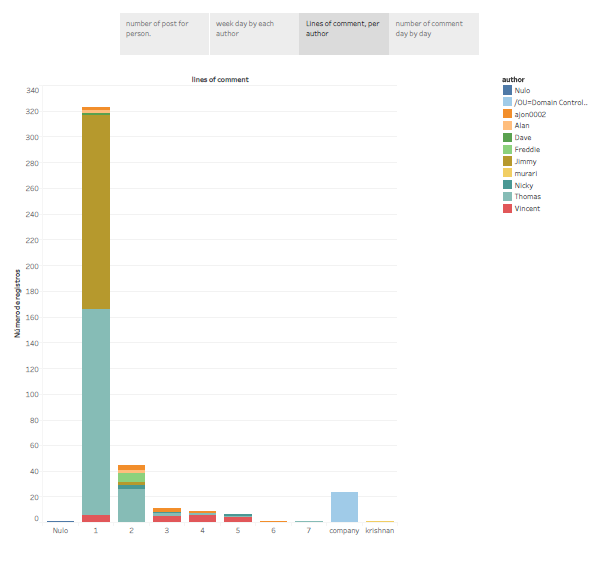
\*Number of post per person.



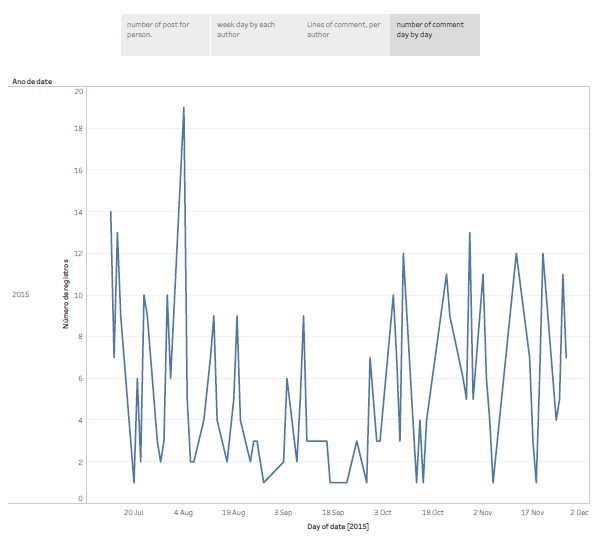
\*Post per week day by each author



\*Number of lines in each comment by authors



\*Number of comments day by day during the process



This is just a demonstration of what I could plot with this data. If I wanted I could go deeper and check comments by hour or minutes and compare with author, for example.

Unfortunately I was not able to create a unitest code, since my code was not working whit classes.