Textual description: Did the team include a textual description of their solution? Is it understandable? Is the described process correct?

Code quality: Is the code readable and concise? Does it use the powerful features of the libraries, or does it re-implement everything from scratch?

Results: Is the final result correct? Are all the assumptions well justified? Are there textual comments/visualizations to convince you of the final result?

Textual description: - Too short, if it needs to be this short there should be some comments in the code that give me the information about what is actually done.

+ Task D and E were a lot better.

Code quality: - Task B, to many long code lines without any form of comments, they look nice, but are hard to read.

+ Very good use of the libraries functionality

Results: - Task C, is scatterplot really the best visualization?

+ Task C, good that they checked the correlation without the two “outliers”.

Review

The textual description throughout the notebook are short and descriptive, but for task B they could have been a little longer and described more of what was going to be done and not just commenting on the result that was achieved. Here part D and E where very good at communicating the actions taken for getting the results and what the results told you.

For every task the different functionalities of the libraries are used in an effective way which makes the code a lot cleaner. For task B there are a few lines of code that are very long and without any comments in the code it is hard to understand exactly what they do, comments in the code or a better textual description could help on solving this. Task C and D where the cells of code are longer are better at this which makes them more understandable.

Your results are good, and the plots used are good to show of the result. But the plots should have been bigger so that it would have been easier to see the differences on the log axis. The same goes for the box plots. In task C it is very good that you checked the correlation both with the two data point with a high ViewCount and Score and with out them. It makes it very clear why the correlation was as high as it was for all data points.

All in all this was a very well done exercise, keep it up!

Grades:

Textual description: 5

Code quality: 5,25

Results: 5,5