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: - Missing in Task A, why use .createOrReplaceTempView() and . persist()?? Task B, the textual description should show the thought process and give an explanation of why you are doing as you are doing, not just the explanations of your results. There are actually no textual descriptions what so ever in this notebook. I can only find one note in the whole notebook, Task D.1.

+ Explanations to the results are short and concise and gets the message through.

Code quality: - Some places the SQL queries are written in just one line and other places they are written line by line, stick to one form (hopefully the line by line form) through out the whole exercise. In task D.3 you write “take the samples and apply the function” and “compute the confidence interval” maybe explain what this function is and how you calculate this (before explanation of the results)? Also, this task contains the smallest boxplot I have ever seen, impossible to read.

+ Uses built in functionality.

Results: - In task D when you remove the answers with a negative response time, it would still be possible to have an answer that looks like it is written in just some seconds, this should have been specified and done something with.

+ The results looks correct, but when things are on a log scale it is hard to differentiate between the different ticks on the axis, make the plots bigger!

Review:

The textual descriptions are missing throughout the whole exercise. There should be more to each task than an explanation of the results you got. Even if these explanations are short and communicate what you learned from the results, they do not help on understanding the process that made these results. It would be better if you before each cell in the start of a task tell us about what you are going to do and why you think this is the right thing to do, and also between some of the cells write what you have found out so far and if this findings are as expected or if they create a new challenge for you.

I only found one note in the notebook where it is explained why the negative time differences are removed, task D part 1. This decision creates the possibility for very low response times close to zero, here a small textual description would fit in nicely explaining what this might do to the results you might get.

Use of the features of the libraries are good and make the code short and sweet and the comments describe what it does. Instead of just writing “take the samples and apply the function” the function should be described before this cell instead of in the explanation of the result multiple cells below. Also, the box plot is way to small to be useful, I can barely see the different lines in it. It is the same for plots with log scaled axis, make them bigger so that it is possible to see the different ticks on the axis.

The results achieved looks to be correct.

All in all, despite some small hiccups, I think this is a good exercise.

Grades:

Textual description: 4,5

Code quality: 5,5

Results: 5,5