Hi Gaoshaung,

I think you did a good job with this assignment, but I ran into some errors in the later parts and it wasn’t clear to me how to correct them from your code. With some tweaks, I think your code is close to running properly. The main errors seemed to involve running the python filters. One general suggestion I have is to use the markdown cells instead of comments when explaining your code. Detailed comments are below.

Problem 1

For Problem 1, we mostly got similar answers but used slightly different methods in some cases. When counting the four names, I included lower-case instances of the names, though I now think it makes more sense to count names only when the first letter is capitalized, like you did.

Problem 2

Our answers were mostly the same, but on the last question (top 10 bikes at top destination station), I think you should have used column 7 instead of column 5 because you essentially repeated what you did for the previous question.

Problem 3

I think there might be an error in the order of your commands in this part. I’m not sure what the purpose is of copying simplefilter.py to lower and split. In my code, I just used chmod on each of my python script files individually and that seemed to work. Your commands appeared to be correct, but when running the cells in order, I ran into issues setting up the python filters properly.

Extra Credit

This seemed to be incomplete. I found that I had issues running this command when including the ‘time’ function. I also recommend taking the parallel commands out of quotations. For this part, there should also be a command to remove stop words from the documents.

Let me know if you have any questions about my comments.

Amit