**Assignment 4**

**Due 7/16/2020 + one week grace pero**

**Please finish the avgwordlength\_student.ipynb nookbook that can be found in the assignment 4 folder. Please also download the two files “shakespeare” (which is your data file) and “test\_helper.py” from the assignment 4 folder and add them to the same folder where you store avgwordlength\_student.ipynb.**

**You need to use the Spark VM to do this assignment. Week 7 video A includes instructions for accessing the VM. The password for windows login is Password1!**

**You need to replace <Fill IN> in the notebook with your code. Before you work on this notebook, please watch the videos week 7 A and B. They are very relevant to the assignment. It would be very difficult for you to do the assignment without watching the videos.**

**Sometimes, if you see an error that is probably related the Spark system, you can re-start pyspark and see if the problem persists. Please save your code often as the VM can crash/hang. Also, please back up your code by email yourself the code.**

**Things to submit: just the notebook avgwordlength\_*yourname*.ipynb**