CMSC 201 Section 40
Spring 2022
Project 3 - Grading Rubric

Total value: 80 points

Program Structure: 10 points

- The overall project code is divided into clear and logical steps, with each step placed into a code cell.
- Note that use of functions is NOT required due to the use of a Jupyter notebook. However, if the program structure would logically call for a function it should be there. That means repeating the same or very similar code in multiple different places in the program will incur a deduction of a number of points that depends on how egregious this is.

Clear and consistent use of markdown cells: 20 points

- 5 points for the initial markdown cell with student name, section, date, project description etc. as describe in Step 1 of the assignment
- 2 points each for markdown cells describing each code cell. For each separate code cell, there should be a markdown cell ahead of it that provides a brief explanation of what the following code cell will do
- 2 points for each markdown cell after a code cell that produces output, clearly explaining what the output is and what it means.
- Total points available for this factor = 20, meaning that there must be at least 8 markdown cells after the initial one.

Python Quality and Correctness: 40 points

- Each of the 7 steps in the assignment must be completed. There will be a 5 point deduction for each step not correctly completed. Note: this factor addresses the code itself for the steps. Correctness of the output is assessed separately; see below. If a step is completed, but an error in reading in data or calculating values causes incorrect

output, that would graded there. E.g., if your length of words in the file were to be off by one, that error would not be deducted in this factor.
- Variable/constant/function naming and use, including consistency with class standards, is worth a maximum or 10 points.
- As noted above, the use of functions is NOT mandatory. However, if functions are used, proper use of them will be evaluated for a total of 5 points
- Ability to read in and correctly process the 10 data files and the "stopwords.txt" file is worth 5 points
- Clear and well-structured loops are worth 5 points
- The format of printed output will not be graded, as long as it can be understood. That is, there are no specific formatting requirements about how many decimal places to show; how much space to put between columns; etc. As long as it can be understood, it's okay.

Correct Output: 10 points

- Up to 10 points will be awarded for correct output.
    - The initial TF table is worth 4 points
    - The IDF table is worth 4 points
    - The final TF-IDF table is worth 2 points