QTM-340: Data Science with Text
Homework 6
Due Tuesday, 11/10 4:20 PM via Canvas

You've created your datasheet and submitted your project proposal. Now it's time to dig in! For your next homework, your assignment is as follows:

Complete an initial test of your method, as set forth in your project proposal.

Remember that **you do not need to have your data perfectly clean** before you perform this initial analysis. It just needs to be clean enough for the method to be run. (You can do additional cleaning later).

Note, also, that **you *will* need to pre-process your data in some ways** (segmenting corpus or document, tokenizing documents or sentences) to run the method. This will likely take you some time!

Submission details:

Your homework should consist of a Jupyter notebook with the following sections (not necessarily in this order):*

1. A **text cell** that states the question you're exploring, including any personal, project-specific definitions of terms requiring clarification (e.g. "change").
2. A **code cell (or cells)** with any cleaning required for the method to run. (As above, this does not need to be the only cleaning of the data you ever perform; only what is needed for this initial analysis).
3. A **code cell (or cells)** with any corpus pre-processing and document segmentation.
4. A **code cell (or cells)** with any required tokenization and stop-word removal
5. A **code cell (or cells)** that implements the method you have chosen to answer the question
6. A **text cell** that briefly (in a paragraph) interprets the results, and identifies the next steps for your project that are suggested by the results.

* Sections 2-5 are intended as a guide. If your approach requires a different format, you can diverge from the specified cells. But you must include cells 1 and 6 (the research question and the analysis) regardless of the structure of your project code.

Upload your Notebook (the file ending in .ipynb) to Canvas by the date specified above.