

# MET CS 688 Assignment 5

---

**Please follow the submission requirements at the end of the assignment!**

Objectives: Demonstrate your ability to use the rtweet and RedditExtractor libraries. Show that you can generate bigrams from a corpus or tidy text object. Show that you can compute the sentiment score using the function provided in the lecture.

Part A (65 points) Using the rtweet library, do the following 6 tasks. You do not have to do them in the order listed below if the order listed below does not make sense for your data or processing choices. However, please call out all 6 tasks clearly in your assignment.

- 1) Pick two countries and search for the tweets associated with these two terms. (*use a search tool, do not retrieve tweets from specific usernames*)
- 2) Process each set of tweets into tidy text or corpus objects.
- 3) Use some of the pre-processing transformations described in the lecture.
- 4) Get a list of the most frequent terms from each country's tweets. Compare them. Do the results make sense?
- 5) Show top word pairs (bigrams) for each country as described in the lecture.
- 6) Compute the sentiment score (as described in the lecture) for all the tweets for each country. Compare the sentiments for the two countries. Do the results make sense?

Part B (35 points) Use the RedditExtractor library, do the following 5 tasks. You do not have to do them in the order listed below if the order listed below does not make sense for your data or processing choices. However, please call out all 5 tasks clearly in your assignment.

- 1) Select an article on Reddit that has at least 5 comments (more is better). Retrieve the comments. You may choose any article from any subreddit. If you don't know which subreddit to choose, you can use the "World News" subreddit and search for one of the countries you studied in Part A (Twitter).
- 2) Process the comments into tidy text or corpus objects..
- 3) Use some of the pre-processing transformations described in the lecture.
- 4) Get a list of the most frequent terms.
- 5) Compute the sentiment score for the comments.

## **SUBMISSION REQUIREMENTS:**

- **Create a Word, PDF, or Rmd document. If you use Rmd you will need to make sure to save the output as a PDF.**

- For each question, state the question you are answering. Then answer the question by explaining in sentences (in English, not in R or other languages) what you did to get to the answer. You may include screenshots and/or copy-paste of key lines of code and the corresponding output in your answer. (If you are using Rmd, this means you must generally use `echo=FALSE` and/or `include=FALSE` for the body of the document.)
- Full code should be included as an Appendix to your Word or PDF document. Coding must be in R. Do NOT include full code in the main part of your document.
- Please ensure that a Word or PDF file as the first file in your submission.
- You may also separately upload your R and/or Rmd code to Blackboard.
- If your facilitator tells you to submit the files differently than the above guidelines, you are expected to respect your facilitator's wishes starting on the next assignment.
- Facilitators can deduct up to 20% if you fail to follow these requirements (more if the questions are not actually answered).
- Facilitators can deduct 5% for each day the assignment is late. You may submit one (and only one) of the six assignments up to three days late with no penalty but all other assignments will be penalized.
- Unless your facilitator or the professor agrees, your assignment will not be graded if it is more than 3 days late (e.g., no credit will be given after Friday at 6 AM Boston time). The professor will usually ask the facilitator to make the decision but in rare cases (<1% of the time) has overridden a facilitator. Do not expect the professor to override in most cases.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder