Programming Assignment 01

Similar to the lecture on Tuesday when we examined the language used for Data Mining and Data Science, you are going to examine the language used when talking about Data Mining and Machine Learning.

To do this you will do the following:
1. Google "What is Data Mining" and "What is Machine Learning" separately. Be sure to record any conditions that you feel will impact the results you will get.

2. Pick the first fifteen sites that are returned as results for **BOTH** "What is Data Mining" and "What is Machine Learning" searches.

3. From the fifteen selected example sites, record a selected block of text that started with something akin to "Data Mining is …" or "Machine Learning is …".

4. For each example record at least the following attributes: category (Data Mining/ Machine Learning), the date the data was pulled, the human readable name of the website the data was pulled from, the exact url to the page the data was taken from, and the selected block of text.
   a. Be sure to detail each attribute you record and the level of measurement for each attribute.

5. Import the data into either RMarkdown or Jupyter Notebooks (your choice).

6. Treat the selected block of texts as a bag of words and get frequency counts for each word.

7. Explore your data by getting a bar chart of the frequency counts for the words used for each search phrase and generate a word cloud for each search phrase.

8. Interpret the differences between the word clouds.

Be sure to provide a reference to anything you may have used to complete the homework: tutorial, article, stack overflow post,...

Please follow the programming homework submission template, which I will attach to the assignment in Canvas.

Submit the assignment as a pdf or word document through Canvas.