**Overall Information**

* You must build this report in a Markdown or Jupyter-(like) notebook.
* You should include code chunks that are printed out, showing both the code and the output in the document.
* You should include headers and notes that are written in a markdown section.

**Code Sections**

**Load Spacy and Process Text**

* In order to show that you got your installation working, we will import a package that we will use a lot.
* In Python, you import packages by using import PACKAGENAME where PACKAGENAME is the name of the library you want to use.
* Import the spacy library.
* Next, you will want to load the spacy English language model. Use the following code to load the model and then process a text.
* Find a news website (like Fox News, HuffPo, etc.) and copy a paragraph/article from that site. Cut and paste it into the section that says “put in some text from a news website”.

nlp = spacy.load(**"en\_core\_web\_sm"**)

processed\_text = nlp(**"PUT IN SOME TEXT FROM A NEWS WEBSITE."**)

**Examine Spacy Output**

* Use the following code to show the output from the Spacy model. Be sure to include the indentation!
* This code loops over each word in the processed text, and prints out the word, the lemma, the part of speech, the part of speech part specific tag, the dependency parsing type, if the word is a number or not, and if the word is a stop word or not.

**for** token **in** processed\_text:

print(token.text, token.lemma\_, token.pos\_, token.tag\_, token.dep\_,

token.is\_alpha, token.is\_stop)

* Examine the output and answer the following questions in a markdown chunk. Note that you should thoroughly detail your answers and generally one or two phrases does not answer a question.
  + Did all the words break apart in a way that you expected it to – we asked spacy to identify all the separate words in a sentence … is that what it did?
  + Does there appear to be a common part of speech from your words? What does that imply?
  + Examining the last column (is\_stop) what kinds of words do you expect stop words to be? Do you think they will be useful for text analysis?

**Examine Spacy Output 2**

* Use the following code to show the output from the Spacy model. Be sure to include the indentation!
* This code examines the named entities within the text:

**for** ent **in** processed\_text.ents:

print(ent.text, ent.start\_char, ent.end\_char, ent.label\_)

* Examine the output and answer the following questions in a markdown chunk. Note that you should thoroughly detail your answers and generally one or two phrases does not answer a question.
  + Given the output, what do you think named entities are?
  + How does the output differ from the text provided by printing the parts of speech?
  + What might you guess spacy tends to label in named entity recognition?