In this portion of the assignment will talk about the observations we witnessed while annotating the five different texts from the dataset. The annotation was definitely one of the most interesting tasks of this assignment. Initially our group thought it would be a simple task, however, as we were going through the texts it was much more complicated. Even though it did take a significant amount of time, the less challenging parts of the task were figuring out all the names of the people and their pronouns. Figuring out the names and the pronouns were fairly easy because we are aware of certain background knowledge which leads us to believe that those entities where persons and the corresponding pronouns were it's coreference. The requirement of our assignment was to label all entities that are people. The most challenging part was definitely understanding other entities whether it may be considered as people which includes Canadians, public officials, titles etc. Furthermore, we faced various challenges with the five different texts. Below, we jotted some of our experiences with each of the texts.

**Document 1: 5c5d3e251e67d78e275e54b5.txt**

The text from document 1 was very exciting to annotate because there were many names visible. All the names included the unique first and last name of the person. Highlighting the entities were quite easy since we have a background knowledge about Canadian politics. The text mostly talked about allegations against Canadian Prime Minister Justin Trudeau. We know that his pronouns are he/him, so it was less complicated to find his coreference. Most challenging part was figuring out whether his title would be considered as a part of his coreference chain. The question arose because throughout the text he was refereed as prime minister so, it was a bit confusing whether we have to highlight it or not. Nonetheless, we did consider the title as part of the cluster because we know that it is a person, and the title is referring back to a person. This similar challenge kept on popping up as we move on with the text for example minister of justice and attorney general. We know that this title refers back to a person, but does that mean it considered a coreference? Another difficult part was understanding whether entities like “public officials” or “office staff” is considered under people or not. We highlighted those entities as people because we are aware of who they might be despite of specific identity. Lastly, there were some coreferee but they did not tie back to any names in the text. We highlighted those entities assuming it refers to the author.

**Document 2: 5c2a60611e67d78e27b8feef.txt**

Observations from document 2 were quite similar to document 1. The topic is about US politics and the government which we are familiar about. Due to the background knowledge, it was quite simple to figure the entities that are people. The author mentions the names of the politicians once or twice throughout the text and later just refers back to them using their title. We highlighted the pronouns and titles since we thought it relates back to the specific person for example the text indicated Donald Trump as the President of United States so, during that time and in the text, we know the president is referring to Donald Trump. Furthermore, there were entities like “federal workers”, which we highlighted since we know they are people but, unsure whether they are considered as people entity.

**Document 3: 5c5d532a795bd2d5c282a094.txt**

Document 3 had very little annotation, because there were few entities that were people. This led us to believe that the type of text and context matters when observing specific entities. This document was mainly talking about an organization and business so, there were more numbers and statistics involved. One of the challenges that we faced with this text was that there were quotes from the organization which was personalized, “We are obligated to respect the confidentiality of our policy”. This was a tricky situation because we are familiar that ‘We’ and ‘our’ is representing the people of the organization so, we highlighted those as a coreference of the organization Suncor Energy.

**Notes from Lovely:**

* For Document 3, I highlighted Suncor / EDC as even though they were not people, the document talked enough about them both, taking up most of the topics of the document. The challenge was whether it would be practical to leave it since they’re not people or include them. Through the lens of a machine, however, especially one trying to summarize a document, I thought that excluding these entities may leave out the most important parts of what to summarize in the document.

**Document 4: 5c5da7aa1e67d78e275f8a3c.txt & Document 5: 5c5e50711e67d78e27616b23.txt**

There was not any new observations or challenges when annotating documents 4 and 5. The outcome was quite like document 1.

In terms of writing an annotation manual to explain the task for somebody else, I would give them advice that when they find their first entity, try to skim through the rest of the document to see if they would appear throughout and highlight them. Then, move on to the next named entity and do the same thing. This way, it’s more systemic and they won’t be trying to move in between different entities throughout the document. Once all the entities are found, read through the document again and ensure that all the pronouns or referents are highlighted. It is easy to miss those pronouns since they can easily appear in between everything.