Overview

Jun Xiang

Muskaan

So, according to our thought process as a team, we wanted to figure out what factors affect the election results in Singapore and what is the extent of that effect. We broke this problem down into a few parts. One of them was to analyse the speeches given by the Prime Minister and Members of Parliament before and after elections. The main purpose of doing this was to understand what they talk about in their speeches and how they try to influence voters. So, for this, I took a sample speech transcript and used tokenizing, removing stop words, parts of speech tagging, word lemmatizing and then finally got a word cloud frequency distribution for the final list of words used during the speech.

This process helped us identify a few commonly used words, but it didn’t help us identify the context in which they were used. However, it did help us solidify our previous hypothesis with the household data. So, one of my next few steps would be to find the context in which these words were used and to scale the word cloud to multiple transcripts. Now, the problem that I encountered while looking for transcripts was that they’re not many of those available in text format. However, there’s lots of YouTube videos with speeches. So, to scale it further I would like to investigate converting YouTube speeches to text format for further analysis.

Sherman

Zexel

I swapped the key value pairs of the dictionary to match it with our other data.

Next step would be figuring out how to combine the datasets, form our hypotheses regarding the household races and use linear regression as the first form of analysis.