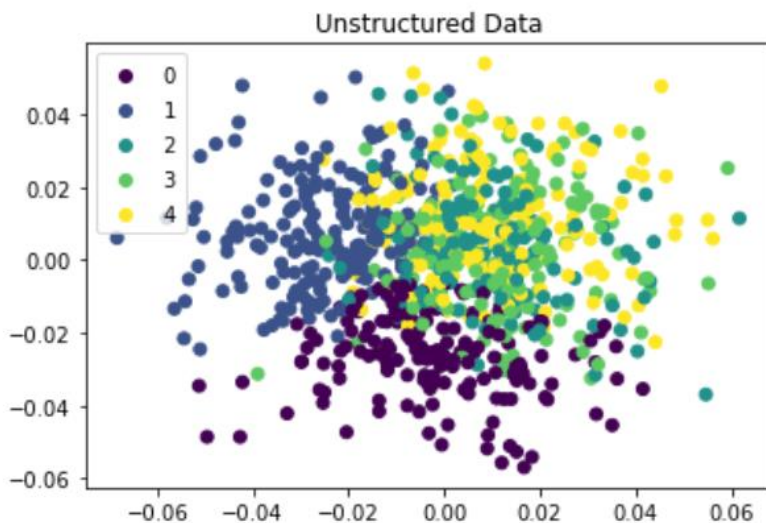
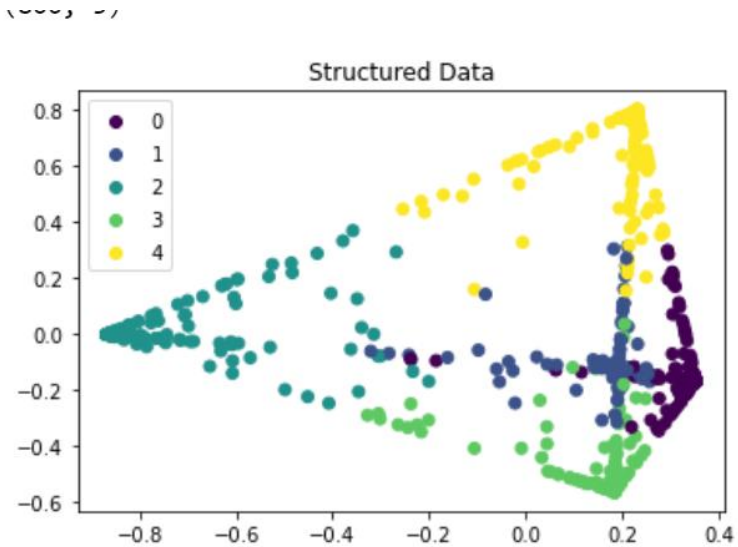


Kobe Rico
Assignment 2

Task 1: Your list of stop words

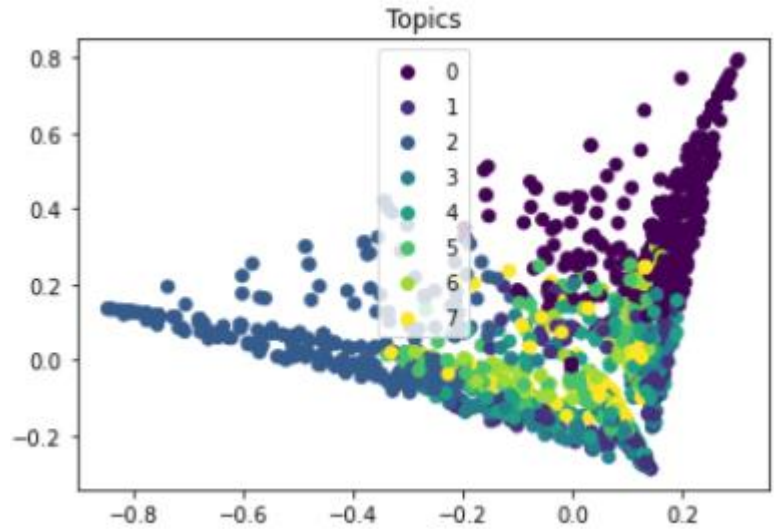
```
my_stops = text.ENGLISH_STOP_WORDS.union(  
    ['this', 'a', 'the', 'every', 'https', 'com', 'COVID-19', 'coronavirus',  
    'almost', 'an', 'that', 'at', 'covid', 'corona', 'pandemic', 'crisis', 'www', '2020',  
    'virus', 'coivd_19', 'http', 'said', 'say', 'read', 'got',  
    'going']  
)
```

Tasks 2: Here's the resulting graphs from Task 2

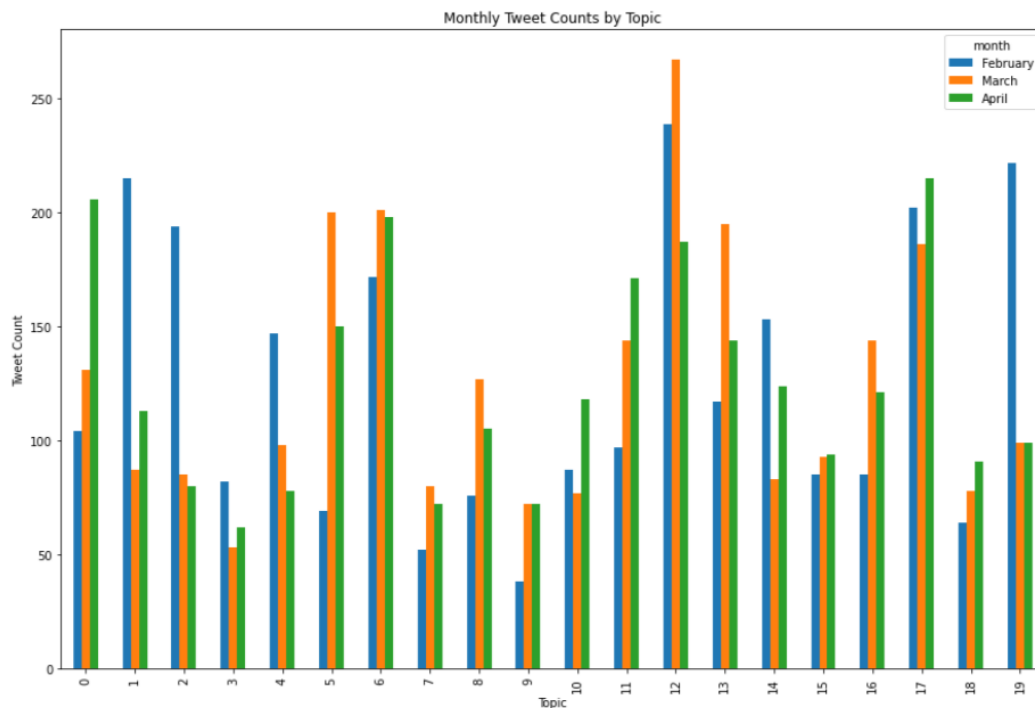


Task 3:

The topics appear to be structured which means that the tweets are grouped tightly around a series of distinct topics. This means that our model did a relatively good job at sorting the topics. Though, we see that while each are clustered, there is a lot of overlapping in the middle. This might be either just from common words we need to consider removing that do not give enough meaning to tweets or common words that are necessary in describing different topics.



Task 4:



Task 5:

Topic 6 was present relatively equally in all months and included:

['just', 'white', 'response', 'countries', 'trump', 'emergency', 'president', 'hope', 'sure', 'god', 'p m', 'donald']

We can conclude that this implies that coronavirus is a national emergency that President Trump has responded to.

Topic 6 implies that the bar graph is accurate in terms of news because it has been constantly stated that we are in a national emergency, with many announcements from our President each month.

19 was present a lot in February including keywords: ['19', 'hands', 'just', 'uk', 'watch', 'lockdown', 'economy', 'months', 'post', 'let', 'follow', 'save']

This topic is about the economy and being on lockdown and the importance of washing hands.

Topic 19 somewhat compliments what was happening in February, especially the emphasis on washing hands the possibility of lockdown. So this implies the bar graph was good.

12 was present a lot in March: ['19', 'amp', 'use', 'spread', 'disease', 'doing', 'realdonaldtrump', 'experts', 'really', 'stop', 'doctors', 'come']

This topic is about stopping the spread of disease and what experts/doctors advise

March was a month where a lot of states went on lockdown and social distancing so this is pretty accurate for what was being spread on social media.

17 was present a lot in April: ['fight', 'chinese', 'years', 'news', 'state', 'schools', 'coming', 'breaking', 'gov', 'minister', 'prime', 'spain']

This topic was about the spread of the disease from China and the closing of schools

This was kind of a repeated topic except, especially since we knew from the beginning of the origin of the disease and how schools started closing in April.

Overall the topics support the bar graph but there are still some flaws such as for topic 17 that did not seem very prevalent in April. Stop words need to be improved the bar graph to hold more meaning but overall it is good.