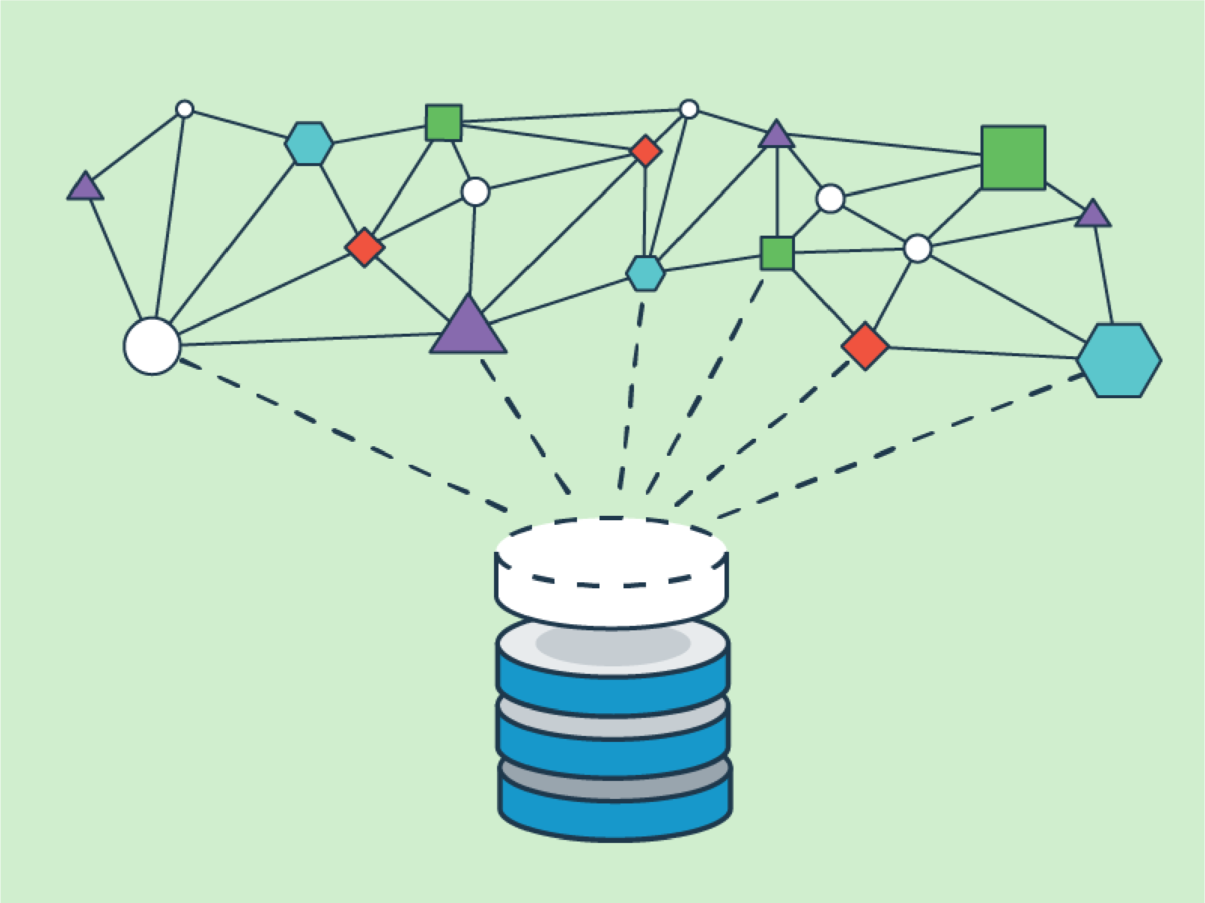
ASSIGNMENT 4 REPORT

CSCI 5408 Data Management, Warehousing and Analytics



ANAND BHADANIA

B00837967

INTRODUCTION

Assignment 4 was given with several tasks which include the following steps:

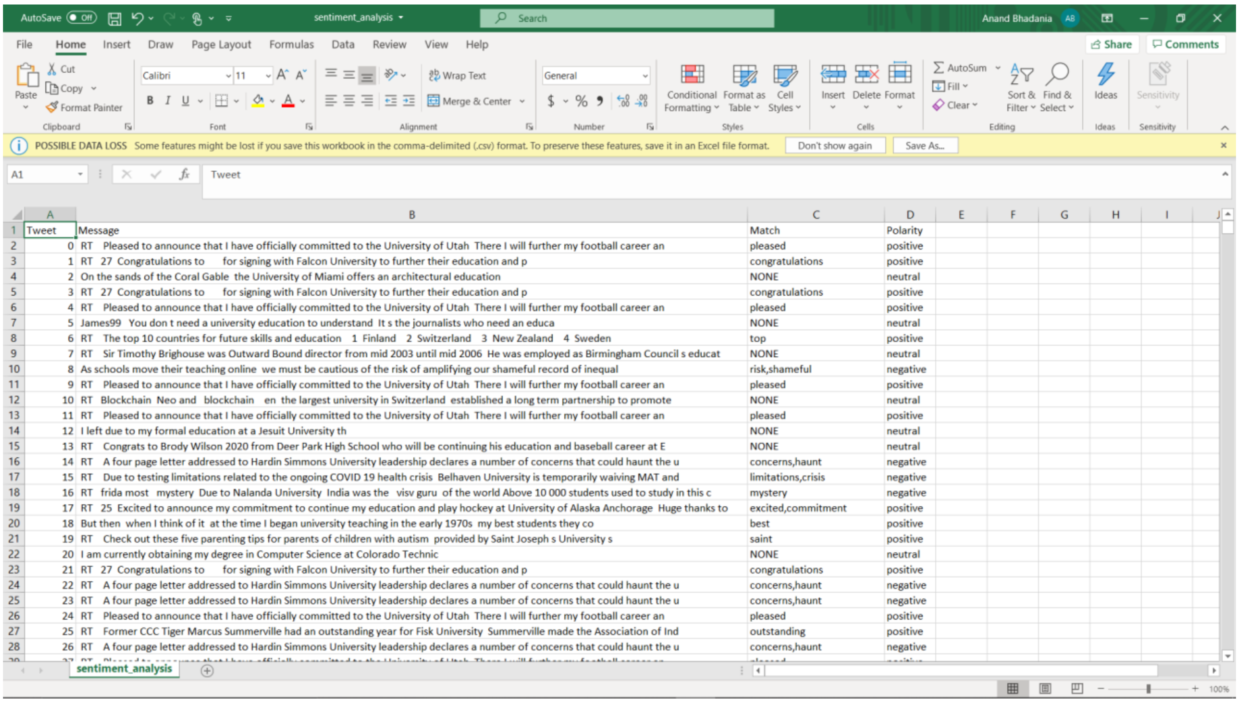
1. Sentiment Analysis
2. Semantic Analysis

SENTIMENT ANALYSIS

To perform this task, I've used the python script which was used to extract tweet from twitter api in Assignment 3. I modified that script to collect only the tweet's message and ignored the metadata. The extracted tweets are in the clean format which I cleaned using Regular Expression (RE) in the Python script. After that, to obtain the polarity for Positive and Negative words, I downloaded files from online sources to get different positive and negative words [1][2].

Then I wrote another python script to classify each tweet as either "positive", "negative", or "neutral". The script is named "Sentiment_Analysis_Script" inside the folder named "sentiment analysis". The script generates 3 excel files which are also inside the folder mentioned above. File named "sentiment_analysis" records the polarity of each tweet along with tweet's message and the word with which the content of the tweet matched.

Other 2 files named "Positive_words_count" "Negative_words_count" contains the positive and negative words along with their count of occurrences in the tweets respectively.



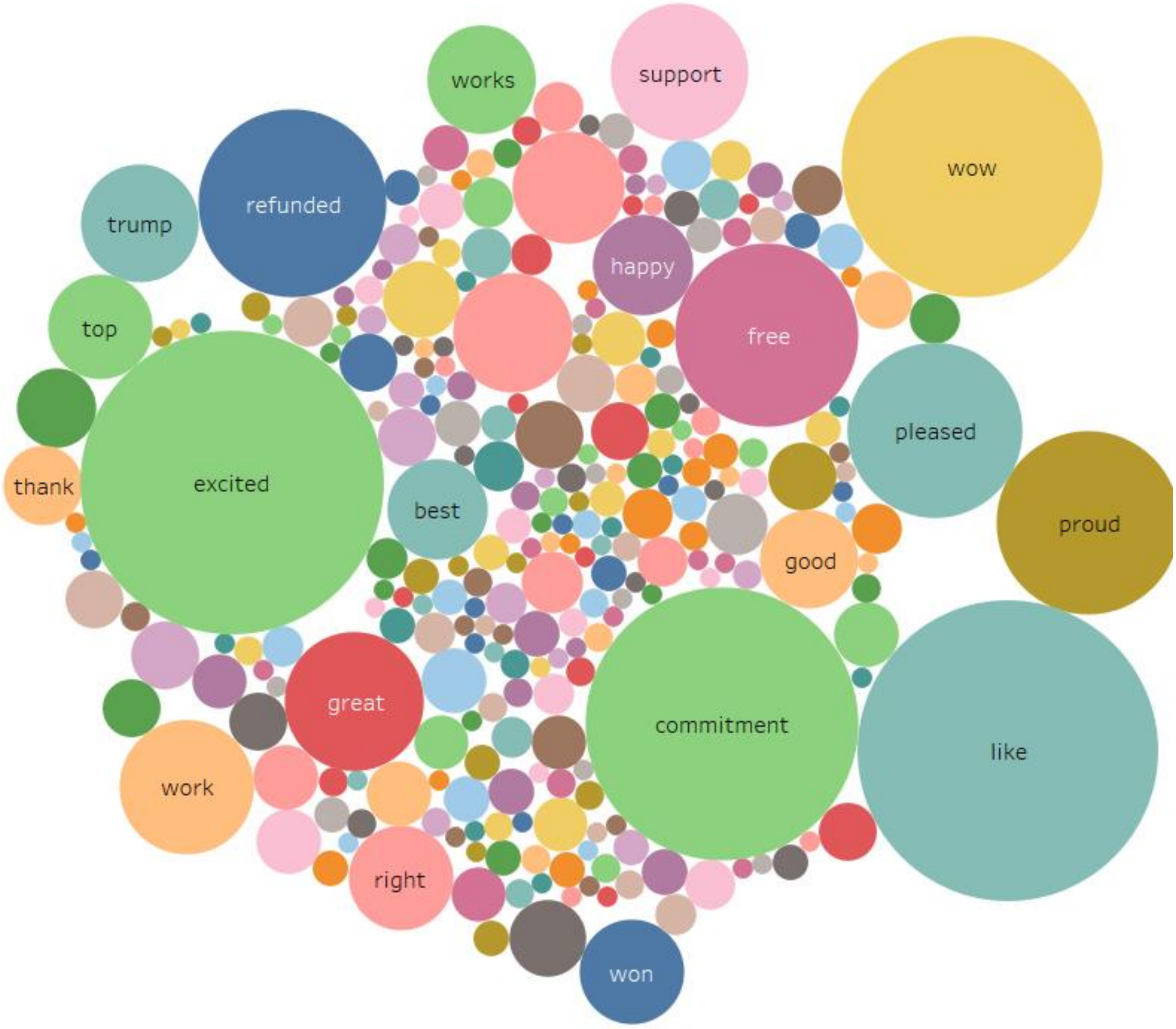
POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.

Tweet	Message	Match	Polarity
0	RT Pleased to announce that I have officially committed to the University of Utah There I will further my football career an	pleased	positive
1	RT 27 Congratulations to for signing with Falcon University to further their education and p	congratulations	positive
2	On the sands of the Coral Gable the University of Miami offers an architectural education	NONE	neutral
3	RT 27 Congratulations to for signing with Falcon University to further their education and p	congratulations	positive
4	RT Pleased to announce that I have officially committed to the University of Utah There I will further my football career an	pleased	positive
5	James99 You don t need a university education to understand It s the journalists who need an educa	NONE	neutral
6	RT The top 10 countries for future skills and education 1 Finland 2 Switzerland 3 New Zealand 4 Sweden	top	positive
7	RT Sir Timothy Brighouse was Outward Bound director from mid 2003 until mid 2006 He was employed as Birmingham Council s educat	NONE	neutral
8	As schools move their teaching online we must be cautious of the risk of amplifying our shameful record of inequal	risk,shameful	negative
9	RT Pleased to announce that I have officially committed to the University of Utah There I will further my football career an	pleased	positive
10	RT Blockchain Neo and blockchain en the largest university in Switzerland established a long term partnership to promote	NONE	neutral
11	RT Pleased to announce that I have officially committed to the University of Utah There I will further my football career an	pleased	positive
12	I left due to my formal education at a Jesuit University th	NONE	neutral
13	RT Congrats to Brody Wilson 2020 from Deer Park High School who will be continuing his education and baseball career at E	NONE	neutral
14	RT A four page letter addressed to Hardin Simmons University leadership declares a number of concerns that could haunt the u	concerns,haunt	negative
15	RT Due to testing limitations related to the ongoing COVID 19 health crisis Belhaven University is temporarily waiving MAT and	limitations,crisis	negative
16	RT frida most mystery Due to Nalanda University India was the visv guru of the world Above 10 000 students used to study in this c	mystery	negative
17	RT 25 Excited to announce my commitment to continue my education and play hockey at University of Alaska Anchorage Huge thanks to	excited,commitment	positive
18	But then when I think of it at the time I began university teaching in the early 1970s my best students they co	best	positive
19	RT Check out these five parenting tips for parents of children with autism provided by Saint Joseph s University s	saint	positive
20	I am currently obtaining my degree in Computer Science at Colorado Technic	NONE	neutral
21	RT 27 Congratulations to for signing with Falcon University to further their education and p	congratulations	positive
22	RT A four page letter addressed to Hardin Simmons University leadership declares a number of concerns that could haunt the u	concerns,haunt	negative
23	RT A four page letter addressed to Hardin Simmons University leadership declares a number of concerns that could haunt the u	concerns,haunt	negative
24	RT Pleased to announce that I have officially committed to the University of Utah There I will further my football career an	pleased	positive
25	RT Former CCC Tiger Marcus Summerville had an outstanding year for Fisk University Summerville made the Association of Ind	outstanding	positive
26	RT A four page letter addressed to Hardin Simmons University leadership declares a number of concerns that could haunt the u	concerns,haunt	negative

Figure 1 sentiment_analysis CSV file | Source: Author

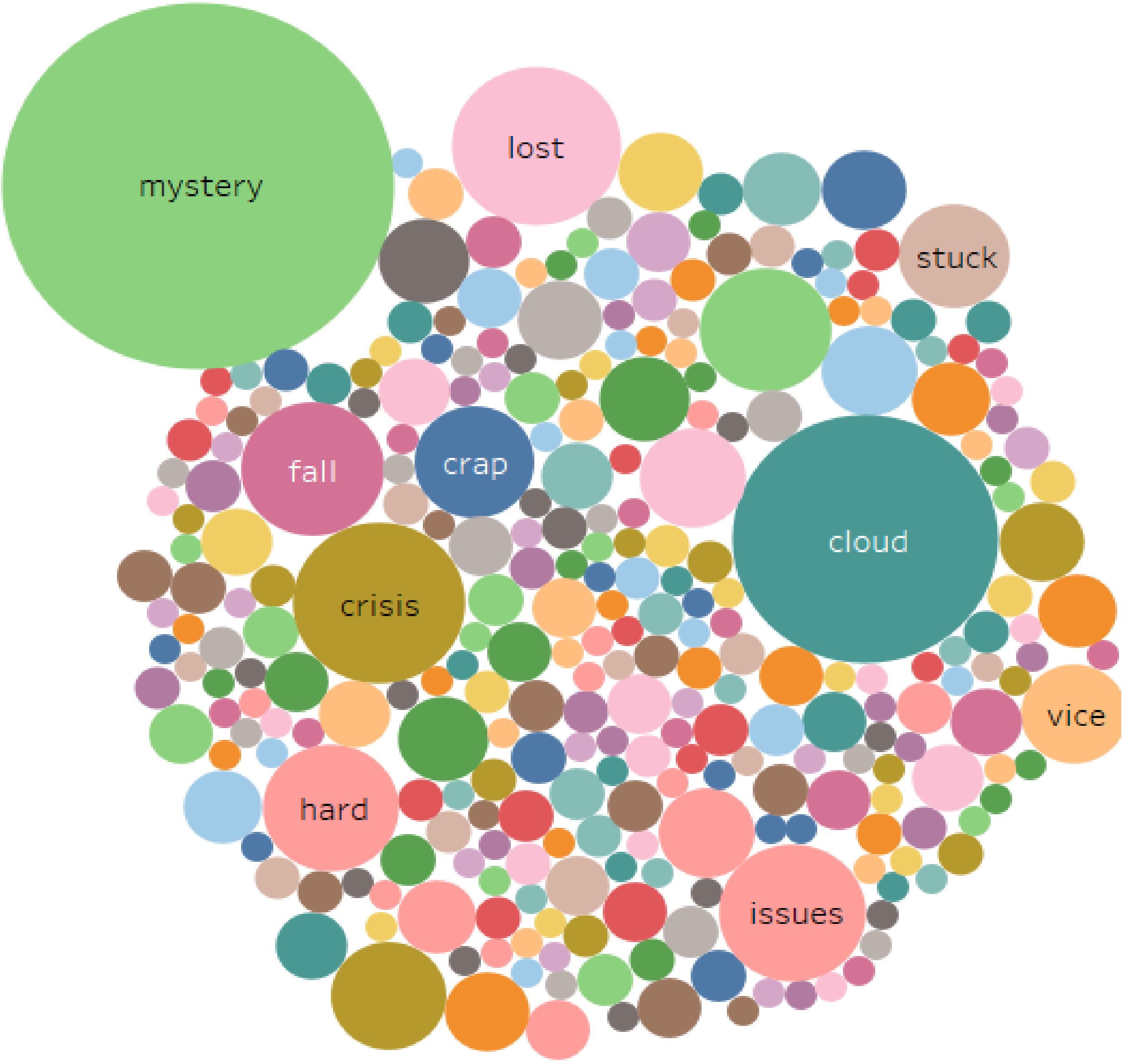
For visualization, I installed tableau from the link mentioned in the lab7 under Lab-Tutorial section under course's content on brightspace [3][4]. I imported excel files named

“Positive_words_count” and “Negative_words_count” to create visualization for the most frequently occurring words in the positive and negative tweets respectively [5].



Words. Color shows details about words. Size shows sum of Count. The marks are labeled by words.

Figure 2 Positive Words Visualisation| Source: Author



Words. Color shows details about words. Size shows sum of Count. The marks are labeled by words.

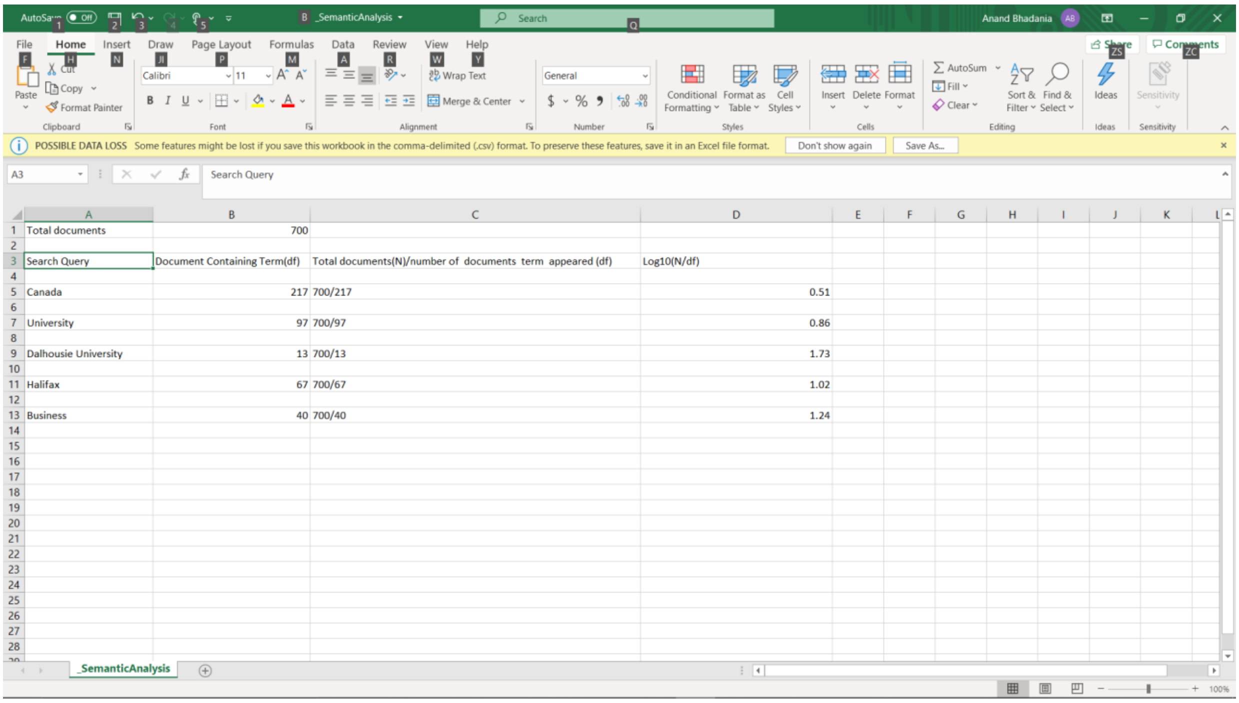
Figure 3 Negative Words Visualisation| Source: Author

The visualisation for both positive and negative words and attached in pictures folders.

SEMANTIC ANALYSIS

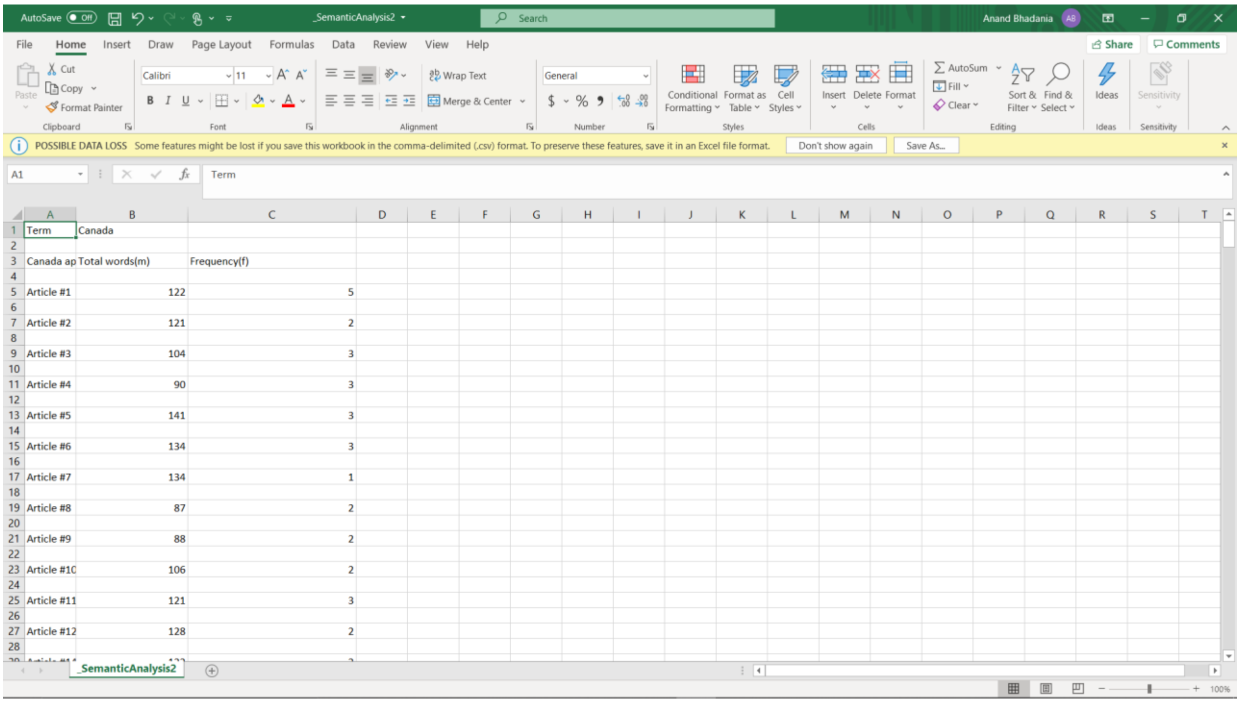
To perform Semantic analysis, I have used the python script, to extract news articles from the news API, used in Assignment 3. I modified that python script to generate different files for different news articles. Total files generated are 700 and all files are inside the folder named "semantic analysis". I created a python script named "Semantic_Analysis_Script" under folder "semantic analysis" which creates 2 excel files and print the article on the console which had highest relative frequency which is explained in the assignment 4 pdf file. One file is named

"_SemanticAnalysis" which stores the frequency count of occurrence of different words(Canada, University, Dalhousie University, Halifax, Business) in different news articles along with additional information. Second file named "_SemanticAnalysis2" stores the information about different articles which contained the word "Canada" and stores the number of words in the article along with the frequency of the word "Canada".


_SemanticAnalysis
Search
Anand Bhadania
File
Home
Insert
Draw
Page Layout
Formulas
Data
Review
View
Help
POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.
Don't show again
Save As...
A3
Search Query
_SemanticAnalysis

	A	B	C	D
1	Total documents	700		
2				
3	Search Query	Document Containing Term(df)	Total documents(N)/number of documents term appeared (df)	Log10(N/df)
4				
5	Canada	217	700/217	0.51
6				
7	University	97	700/97	0.86
8				
9	Dalhousie University	13	700/13	1.73
10				
11	Halifax	67	700/67	1.02
12				
13	Business	40	700/40	1.24

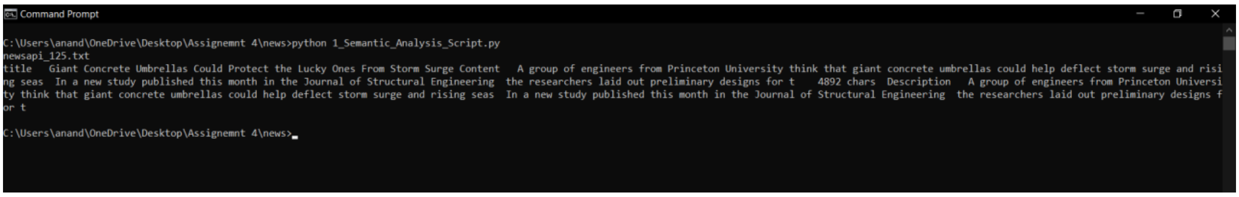
Figure 4 : _SemanticAnalysis CSV file| Source: Author


_SemanticAnalysis2
Search
Anand Bhadania
File
Home
Insert
Draw
Page Layout
Formulas
Data
Review
View
Help
POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.
Don't show again
Save As...
A1
Term
_SemanticAnalysis2

	A	B	C
1	Term	Canada	
2			
3	Canada ap	Total words(m)	Frequency(f)
4			
5	Article #1	122	5
6			
7	Article #2	121	2
8			
9	Article #3	104	3
10			
11	Article #4	90	3
12			
13	Article #5	141	3
14			
15	Article #6	134	3
16			
17	Article #7	134	1
18			
19	Article #8	87	2
20			
21	Article #9	88	2
22			
23	Article #10	106	2
24			
25	Article #11	121	3
26			
27	Article #12	128	2
28			

Figure 5 _SemanticAnalysis2 CSV file| Source: Author

The article with the highest relative frequency printed on the console is shown in the Figure 6.



```
C:\Users\anand\OneDrive\Desktop\Assignemnt 4\news>python 1_Semantic_Analysis_Script.py
newsapi_125.txt
title   Giant Concrete Umbrellas Could Protect the Lucky Ones From Storm Surge Content   A group of engineers from Princeton University think that giant concrete umbrellas could help deflect storm surge and risi
ng seas  In a new study published this month in the Journal of Structural Engineering  the researchers laid out preliminary designs for t    4892 chars  Description   A group of engineers from Princeton Universi
ty think that giant concrete umbrellas could help deflect storm surge and rising seas  In a new study published this month in the Journal of Structural Engineering  the researchers laid out preliminary designs f
or t

C:\Users\anand\OneDrive\Desktop\Assignemnt 4\news>
```

Figure 6 Highest relative frequency Article| Source: Author

* *For this assignment, all the scripts and images are available in the folder.*

REFERENCES

1. 262588213843476, “negative-words.txt,” *Gist*, 14-Dec-2012. [Online]. Available: https://gist.github.com/mkulakowski2/4289441. [Accessed: 08-Apr-2020].
2. 262588213843476, “positive-words.txt,” *Gist*, 14-Dec-2012. [Online]. Available: https://gist.github.com/mkulakowski2/4289437. [Accessed: 08-Apr-2020].
3. “BI Tools and Tableau - CSCI5408 - Data Mgmt Warhsng Analytics (Sec 1) - 2020 Winter,” *Brightspace.com*, 2020. [Online]. Available: https://dal.brightspace.com/d2l/le/content/110354/viewContent/1624572/View. [Accessed: 08-Apr-2020].
4. “Product Trials,” *Tableau Software*, 2012. [Online]. Available: https://www.tableau.com/products/trial. [Accessed: 08-Apr-2020].
5. Parul Pandey, “Word Clouds in Tableau: Quick & Easy. - Towards Data Science,” *Medium*, 20-Feb-2019. [Online]. Available: https://towardsdatascience.com/word-clouds-in-tableau-quick-easy-e71519cf507a. [Accessed: 08-Apr-2020].