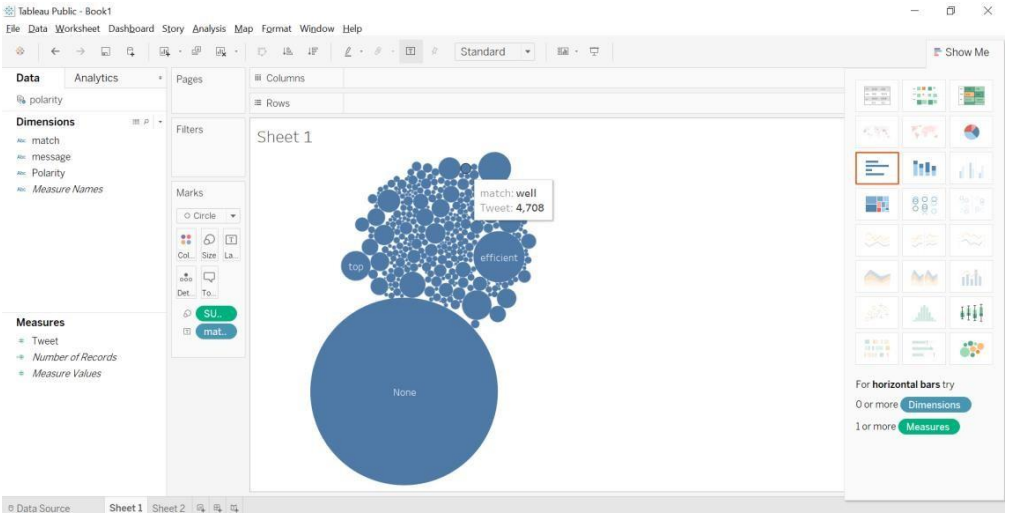
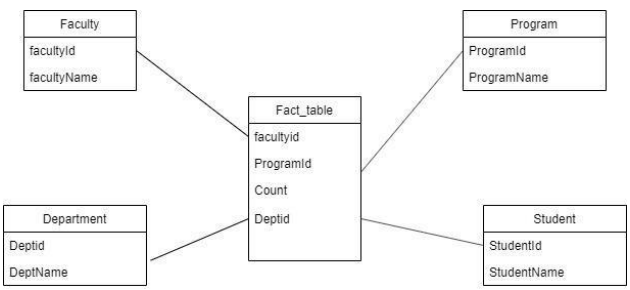
Report

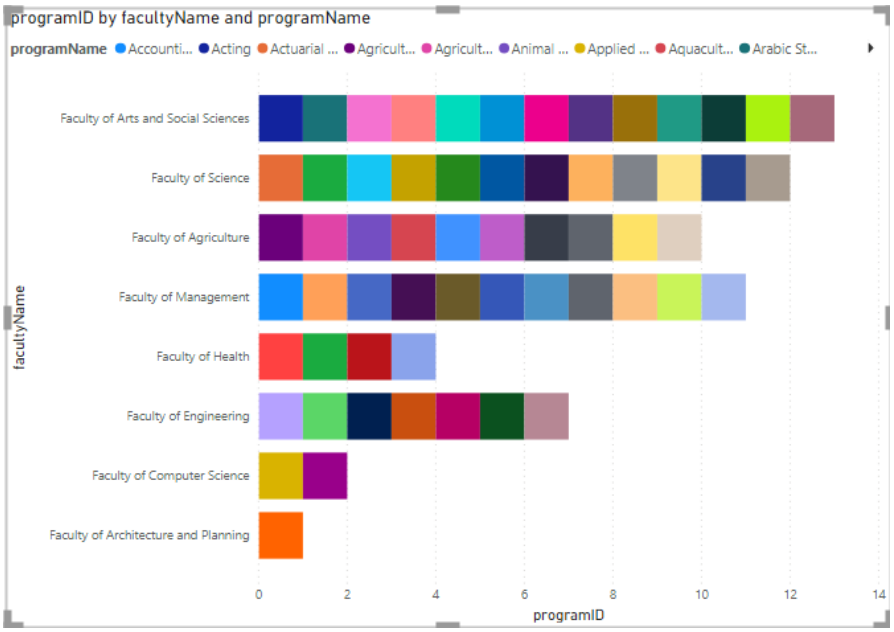
A) The tweets were extracted in previous assignments. They were cleaned using regular expression. The cleaned tweets was stored into a text file. The text file of positive and negative words was downloaded from online sources. Each tweets were stored in a list and its bag of words was generated using dictionary. The list of positive and negative words was also generated. If the positive word count was greater than negative than positive polarity was assigned, else negative was assigned. If count were same for both of them neutral polarity was assigned. Then after looping through all the tweets it was written in a "polarity" named csv file. Then all the positive , negative and neutral tweets most frequent word was visualized using tableau. It is shown in the figure below.[1-3]



B) The news article were extracted in the previous assignment and were cleaned using regular expression. Each new text file was generated for news article containing "title", "description", and news "content". Now each 500 documents were opened and count was increased if "Canada", "University", "Dalhousie University", "Halifax", "Canada Education" appeared in the document. Then the division was done of "Total documents and in number of documents the term appeared and logarithmic function was applied to it. Then the result was stored in the "tfidf.csv" file. Now, the frequency count of word "Canada" was done in each file through similar procedure and was stored in "tfifdf2.csv" file.[1-3]

C) The star schema and the report generated through the powerBI is shown in the figures below. It can be seen from the report that faculty of computer science doesnot have maximum departments. The number of courses in each department can be seen from the figure below.




programID by facultyName and programName
programName
Accounti...
Acting
Actuarial ...
Agricult...
Agricult...
Animal ...
Applied ...
Aquacult...
Arabic St...
Faculty of Arts and Social Sciences
Faculty of Science
Faculty of Agriculture
Faculty of Management
Faculty of Health
Faculty of Engineering
Faculty of Computer Science
Faculty of Architecture and Planning
facultyName
0
2
4
6
8
10
12
14
programID

REFERENCES:

[1] Get and Work With Twitter Data in Python Using Tweepy. (2018, February 5). Retrieved November 25, 2019, from https://www.earthdatascience.org/courses/earth-analytics-python/using-apis-natural-language-processing-twitter/get-and-use-twitter-data-in-python/.

[2] Python client library. (n.d.). Retrieved November 25, 2019, from https://newsapi.org/docs/client-libraries/python.

[3] Python 3.8.0 documentation. (n.d.). Retrieved December 2, 2019, from https://docs.python.org/3/.