CSCI5408- ASSIGNMENT 3

A. SENTIMENT ANALYSIS:

Data Pre-processing-

The data pre-processing and scraping is done with the help of the Assignment 2. URLs, emoticons, emojjis and special characters are removed using regex. Lemmatization or stemming is not done so as to preserve the original text. Also I've used langdetect[1] library to remove other language tweets and only English language is considered.

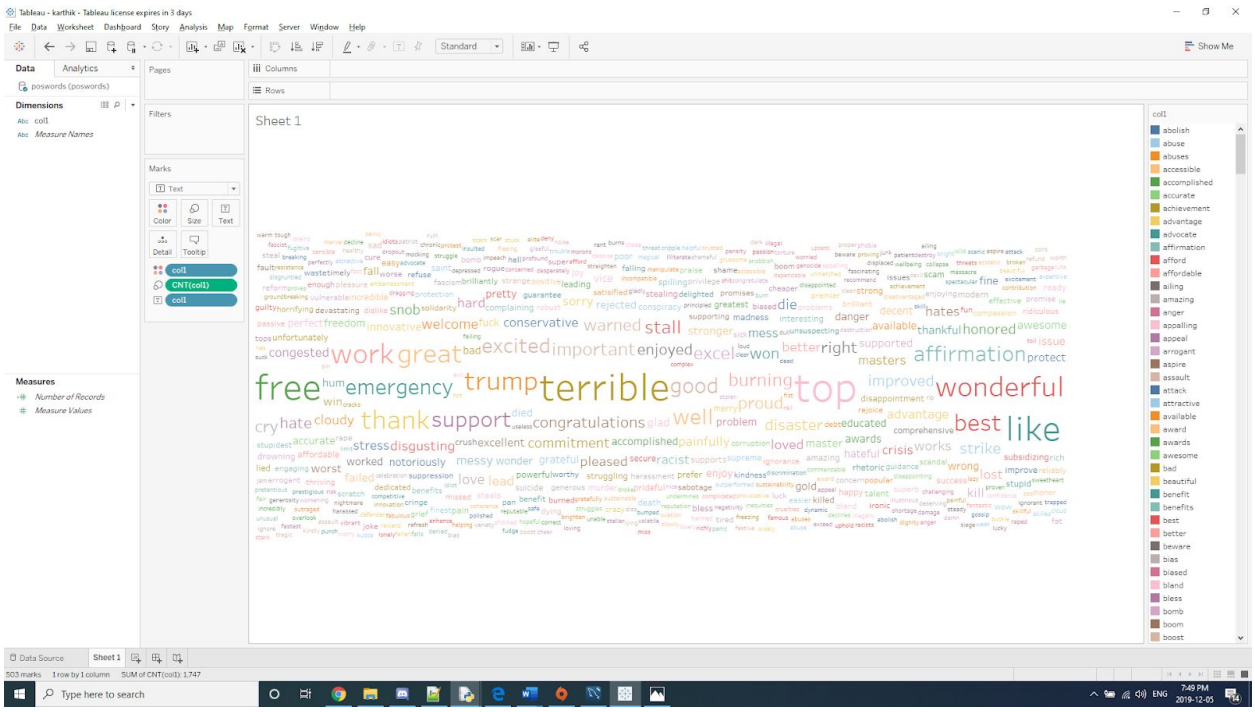
Polarity-

For finding polarity, tweet is considered and the number of positive and negative words in the tweets are found. If positive words are greater than the negative words, then it is termed as Positive and for the inverse it is termed Negative. If both the numbers are the same, then it is termed Neutral polarity.

A word is tagged as positive, if the word is available in the positive word list[2] and the previous 2 words does not contain a negation word[3].

Similarly, a word is tagged a negative, if the word is available in the negative word list[4] and the previous 2 words does not contain a negation word[3].

Positive and negative words available in the tweets are taken and word cloud is formed using Tableau.



CSCI5408- ASSIGNMENT 3

B. SEMANTIC ANALYSIS:

Similarly, the data pre-processing and scraping is done with the help of the Assignment 2. URLs, emoticons, emojjis and special characters are removed using regex. Lemmatization or stemming is not done so as to preserve the original text.

Each news article from the mongodb is considered and are not stored as files and is used directly. TF-IDF are found respectively and the result is stored as csv.

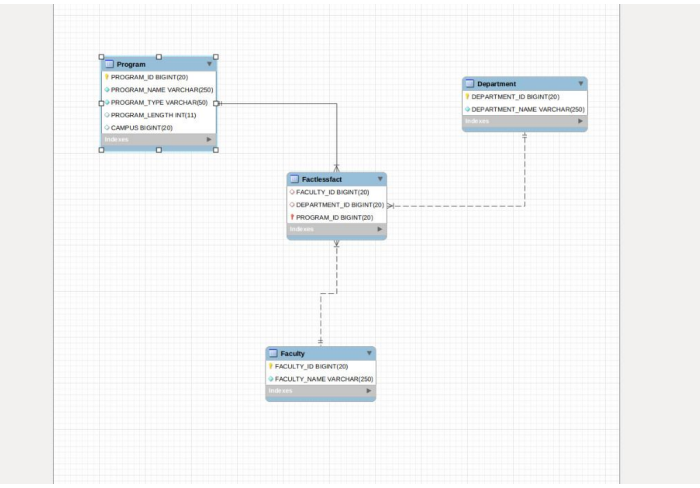
Refer attached files.

C. BUSINESS INTELLIGENCE:

For the given problem, a factless fact table is considered. The dimensions of the fact table are Faculty, Program, Department.

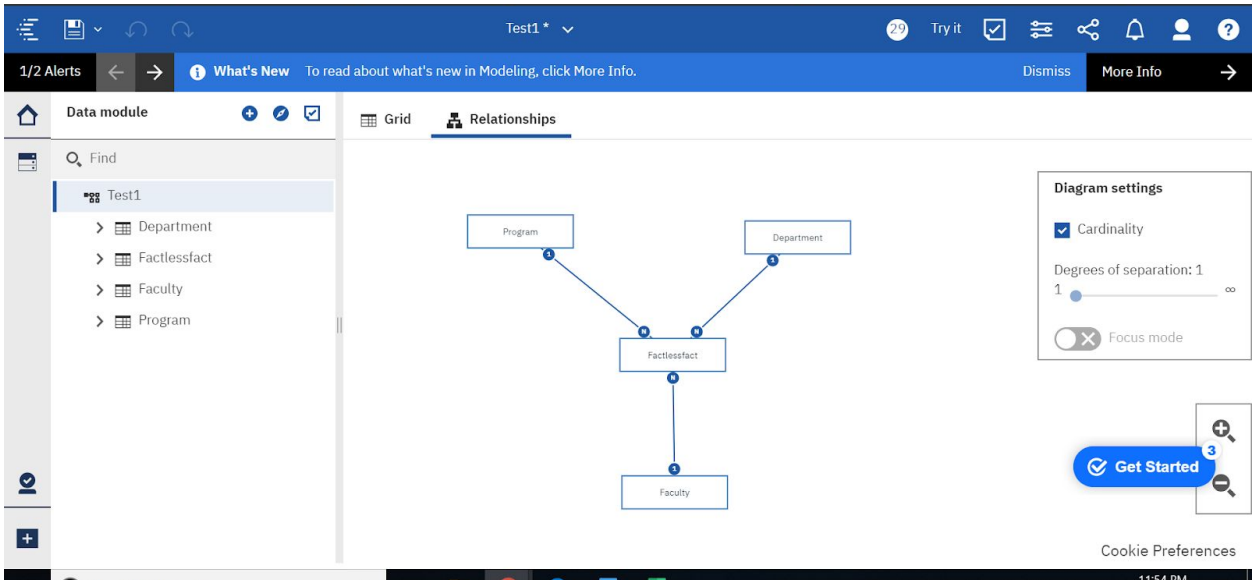
Fact(PROGRAM_ID, DEPARTMENT_ID, FACULTY_ID)
Program(PROGRAM_ID, PROGRAM_NAME, PROGRAM_TYPE, PROGRAM_LENGTH)
Faculty(FACULTY_ID, FACULTY_NAME)
Department(DEPARTMENT_ID, DEPARTMENT_NAME)

Data model in MySQL:

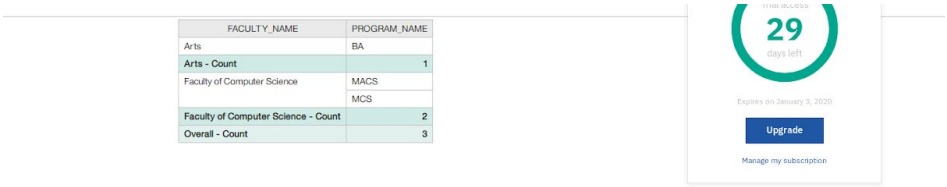


CSCI5408- ASSIGNMENT 3

Star Schema for the given problem[5]:



Query:



FACULTY_NAME	PROGRAM_NAME
Arts	BA
Arts - Count	1
Faculty of Computer Science	MACS
	MCS
Faculty of Computer Science - Count	2
Overall - Count	3

References:

[1] “langdetect,” PyPI. [Online]. Available: https://pypi.org/project/langdetect/. [Accessed: 05-Dec-2019].
[2] Index of /~liub/FBS. [Online]. Available: http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar. [Accessed: 05-Dec-2019].
[3] “Negatives and Negation–Grammar Rules,” Grammarly, 21-May-2019. [Online]. Available: https://www.grammarly.com/blog/negatives/. [Accessed: 05-Dec-2019].
[4] Index of /~liub/FBS. [Online]. Available: http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar. [Accessed: 05-Dec-2019].
[5] R. Gupta, H. Pamnani and M. Bhanderi, "Cognos BI", Dalhousie University, 2019.