5408 Assignment 3

Ruize Nie

B00739439

A part Semantic Analysis of Twitter
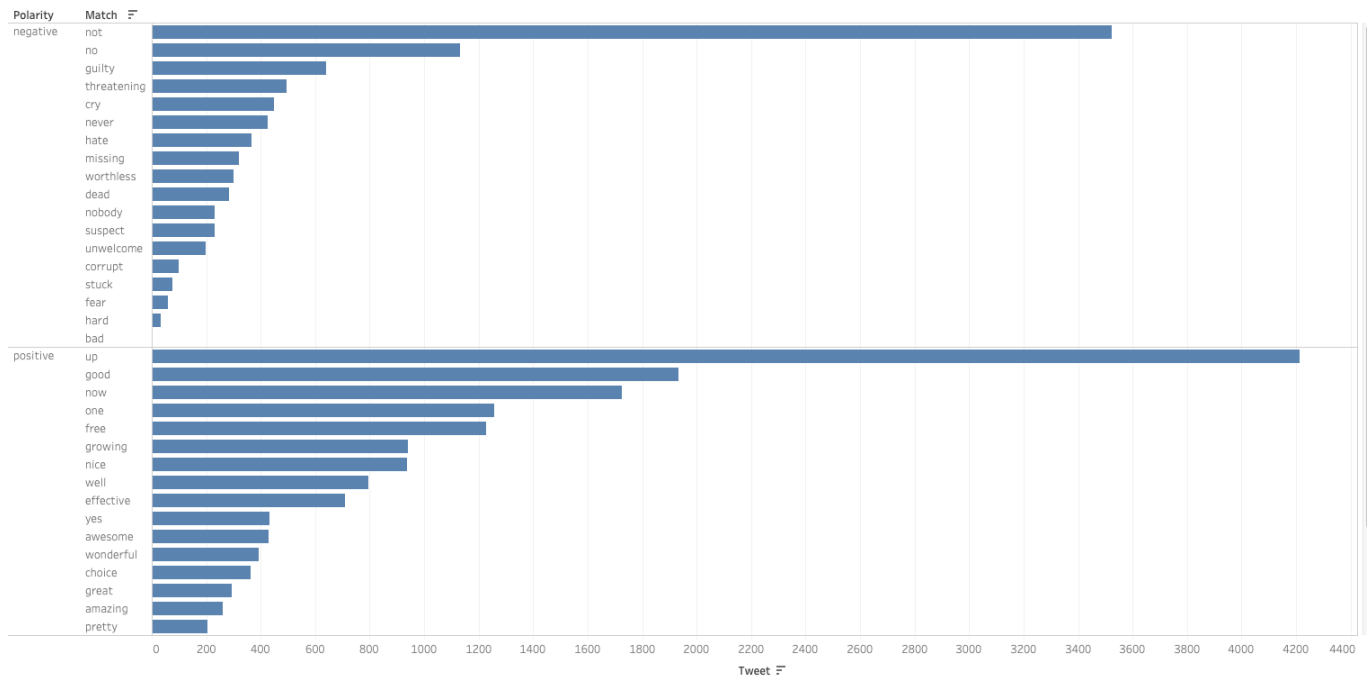
1. Tweets capture using Tweepy. (show only top 20, rest attach on folder)

1	RT More brought to us by Palmater Reconciliation childwelfare indigenous CHRT
2	RT peryer Yale and Harvard students have taken to the field to protest university holdings in fossil fuel companies and Puerto Ric
3	Want sims discover university so bad
4	RT Students at Grand Valley State University in Michigan voted last week to stop reciting the Pledge of Allegiance at student
5	The only test they re interested in is this is your money in fact Green Green dipl
6	are represented by the University of Calgary Love seeing your skills
7	RT Blessed to receive an offer from Charleston Southern University Coach Toomer Huggins80
8	RT Tv3 news clip on the conferment of Her Majesty Seri Paduka Baginda The Raja Permaisuri Agong
9	the high rise aparteme
10	RT Canada votes against Israel for the first time in 14 years and our Ambassador boasts that Canada has regained its voice No
11	Canada s medicine is still cheaper than the US and besides medicine costing money has nothing to
12	RT Geekhome Canada shed 22 600 jobs in October according to ADP report via
13	RT akel Next week concerts the artist Elissa in Canada Toronto 29th November at Friday night Ottawa 30th November at S
14	PLQ Inc pr sent depuis 1867 Qu on les l che Ils ont toujours agi en collabos Pour qu on soit pour toujours da
15	RT Trump cut his teeth on the scorched earth approach of Roy Cohn but like all bullies when push comes to shove he does fold
16	I had the opportunity to speak to several police officers I thanked them for protecting the Jewis
17	RT Handsworth Market Soho Road 9th March 1968 University of Birmingham Phyllis Nicklin
18	ASS You are beyond amazing
19	Huge congrats to Canada s Pospisil and Shapovapov for making history today and advancing Canada to the Fi
20	RT Maharaja Rameshwar Singh of Raj Darbhanga Donated 50 lakhs of rupees for the foundation of the BHU Played leading role

2. Tweets semantic analysis [1] (show only top 20, rest attach on folder)

	Tweet	Message	Match	Polarity
1				
2	1	RT More brou	null	neutral
3	2	RT peryer Yal	null	neutral
4	3	Want sims di	bad	negative
5	4	RT Students a	null	neutral
6	5	The only test	null	neutral
7	6	are represent	null	neutral
8	7	RT Blessed to	null	neutral
9	8	RT Tv3 news	null	neutral
10	9	the high rise a	null	neutral
11	10	RT Canada vc	null	neutral
12	11	Canada s me	null	neutral
13	12	RT Geekhom	null	neutral
14	13	RT akel Next	null	neutral
15	14	PLQ Inc pr se	null	neutral
16	15	RT Trump cut	null	neutral
17	16	I had the opp	null	neutral
18	17	RT Handswor	null	neutral
19	18	ASS You are k	amazing	positive
20	19	Huge congrat	null	neutral

3. Visualize the most frequently occurring words in the positive and negative tweets

B part Semantic Analysis of News Article

1. News capture using News API. (show only top 20, rest attach on folder)

1	Title	Desc	Content
2	London Bridg	Saskia Jones	Jeremy Corbyn the leader of the opposition Labour Party told Sky News that he believed that the criminal justice system needed to be
3	Racial Slurs a	How a proter	White fraternity members approached on campus declined to comment on the protests The Syracuse Intrafraternity Council said it wa
4	University of	University of	The University of Alberta campus is pictured in Edmonton Alta in an Aug 26 2016 file photo CODIE MCLACHLAN The Globe and Mail So
5	Canadian Th	Activists witl	Activists with The Satanic Temple say a now former therapist in Canada has been pushing conspiracies about globe spanning mind cor
6	Ron MacLea	Longtime na	Longtime national sports broadcaster Ron MacLean said he hopes to continue his friendship with former co host Don Cherry after his c
7	Cal State Cha	Thursday A c	Ive had a chance to work with the board of trustees and appoint many new presidents I think Ive been on 21 searches and weve totally
8	Student pove	A student wh	Image copyrightAFPImage caption Protesters gathered at the Lyon university attended by the student Students have marched through
9	Silicon Valle	Singularity U	The Fine Print The following comments are owned by whoever posted them We are not responsible for them in any way
10	Brigham You	The Church o	The lowest income students at the Brigham Young University Idaho face a momentous choice pay for private health insurance or drop
11	Millions of S	A massive da	A massive database storing tens of millions of SMS text messages most of which were sent by businesses to potential customers has
12	Brigham You	Brigham You	The lowest income students at the Brigham Young University Idaho face a momentous choice pay for private health insurance or drop
13	Report predi	Report predi	VANCOUVER Superbugs are likely to kill nearly 400 000 Canadians and cost the economy about 400 billion in gross domestic product o
14	Report predi	Report predi	Superbugs are likely to kill nearly 400 000 Canadians and cost the economy about 400 billion in gross domestic product over the next 3
15	Irving Burgie	He and Mr B	He added The audience had a very good time it appeared They understood or had previously encountered and mastered Lord Burgesss
16	Texas Is Taki	A takeover o	Around the country there have been more than 100 state takeovers of school districts since 1989 and those with predominantly nonwh
17	Here s what	The history o	The history of higher education in America is older than the nation s founding The oldest known college to continuously exist on US so
18	Buttigieg ris	Pete Buttigie	WASHINGTON Reuters Pete Buttigieg has pulled to the top of the pack of Democratic presidential candidates in Iowa for the first time
19	Early release	Early release	Usman Khan was convicted on terrorism charges but let out of prison early He attended a Learning Together conference for ex offende
20	Dating com a	Following th	Following the recent news about the Badoo and Bumble and Badoo exit there is more consolidation in the dating app space It seems

2. Term frequency-inverse document frequency

1	Search Query	Document containing term(df)	Total Documents(N)/ number of documents term appeared (df)	Log10(N/df)
2	Canada	7	0.14	-1.17113577
3	University	49	0.98	-113.9740856
4	Dalhousie University	0	0	0
5	Halifax	0	0	0
6	Canada Education	20	0.4	-2.512941595

3. Highest occurrence of the word "Canada" (show only top 20, rest attach on folder)

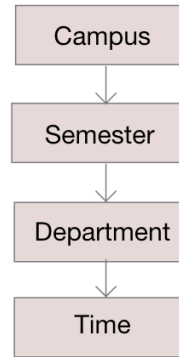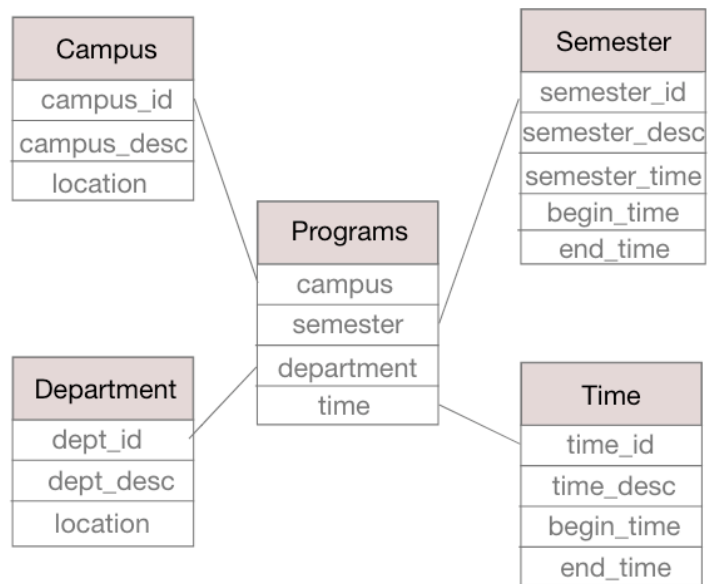|	Term	Canada		
1	Term	Canada		
2	Canada appeared in number_articles documents	Total Words	Frequency (f)	Computing (f/m)
3	London Bridge Attack Victims Were Advocates for Prisoner Rehabilitation	73	0	0
4	Racial Slurs and 15 Days That Shook Syracuse	61	0	0
5	University of Alberta students call on school to fire assistant lecturer who denies the Holodomor The Globe and Mail	102	0	0
6	Canadian Therapist Gives Up License After Satanists Expose Her Mind Control Talks	94	2	0.021276596
7	Ron MacLean addresses Don Cherry controversy as he scores honorary degree from University of Alberta Edmonton Journal	88	2	0.022727273
8	Cal State Chancellor Talks Public Education	72	0	0
9	Student poverty French march in protest after suicide bid	65	0	0
10	Silicon Valley s Singularity University Is Cutting Staff CEO Exits	67	0	0
11	Brigham Young University is forcing poor students to choose between free health care and a college education	102	0	0
12	Millions of SMS messages exposed in database security lapse	93	0	0
13	Brigham Young University Idaho is forcing poor students to choose between free health care and a college education edited	105	0	0
14	Report predicts drug resistance likely to kill 400 000 Canadians by 2050 CTV News	99	0	0
15	Report predicts drug resistance likely to kill 400 000 Canadians by 2050 North Island Gazette	100	0	0
16	Irving Burgie Who Wrote Calypso Hits for Harry Belafonte Dies at 95	80	0	0
17	Texas Is Taking Over Houston s Schools Prompting Charges of Racism	94	0	0
18	Here s what the oldest colleges in America used to look like and what they look like today	104	0	0
19	Buttigieg rises in Iowa to lead Democratic White House pack poll	89	0	0
20	Early release of convicted terrorist under scrutiny after London Bridge attack Global News	98	1	0.010204082

4. Program output

```
Run:    News_5408a3 ×
        /Users/nieruize/PycharmProjects/Twitter/venv/bin/python /Users/nieruize/PycharmProjects/Twitter/News_5408a3.py
        highest relative frequency of artile is: Ron MacLean addresses Don Cherry controversy as he scores honorary degree from University of Alberta Edmonton Journal

        Process finished with exit code 0
```

C part Business Intelligence of Dalhousie University Data Model

1. For the Dalhousie University data model, the main facts are the sum of the number of programs within each department by times.

2. In the Dalhousie University data model, the possible dimensions which can be used in the dimension table are time, semester, campus & department. Each of the dimension provide source to the total number of programs in the fact table.

3. The attributes of the different dimensions of Dalhousie University data model are as:
   a. Time dimension: time_id, time_description, beginning_time, ending_time
   b. Semester dimension: semester_id, semester_description, semester_time, ending_date, begin_date.
   c. Campus dimension: campus_id, campus_description, campus_locaton
   d. Department dimension: department_id, department_description, department_location

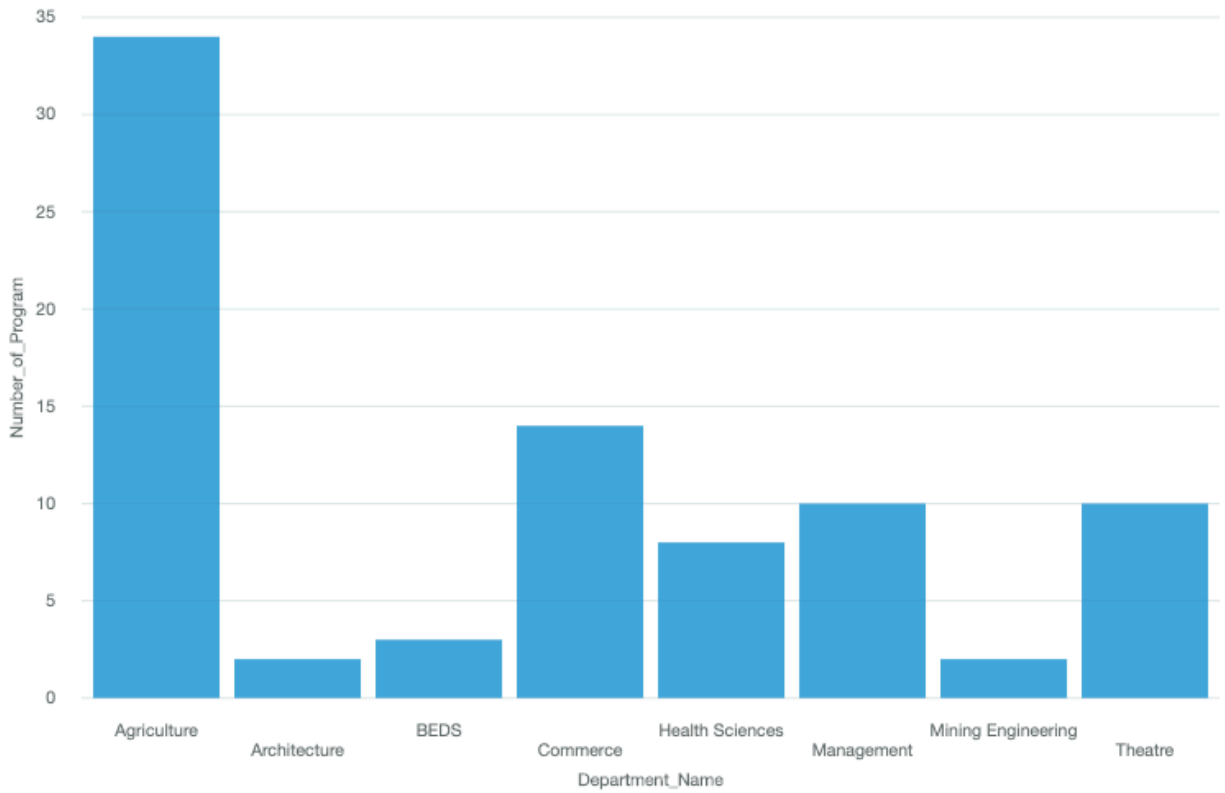4. For the Dalhousie University data model, the attribute hierarchy from top to down is as:
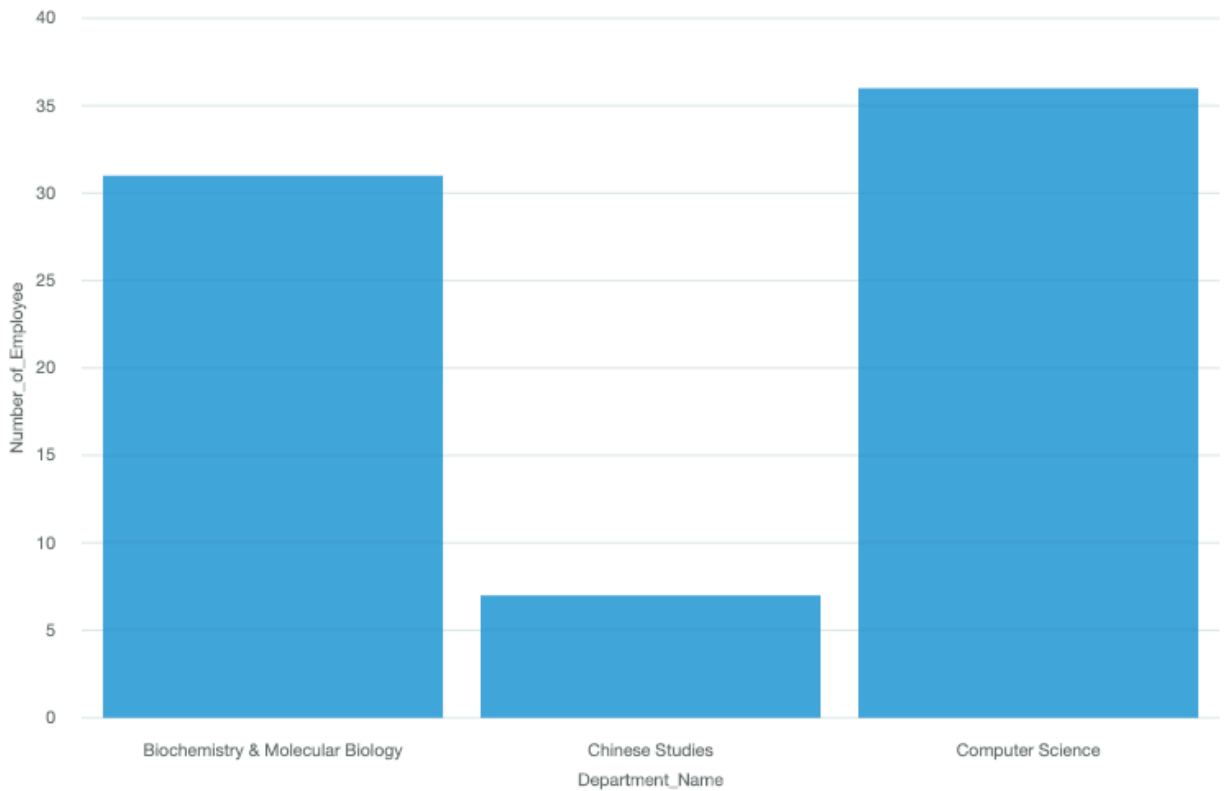
5. Data warehouse design

6. Connect database with Cognos BI [2]
   a. Check how many programs offer by each department
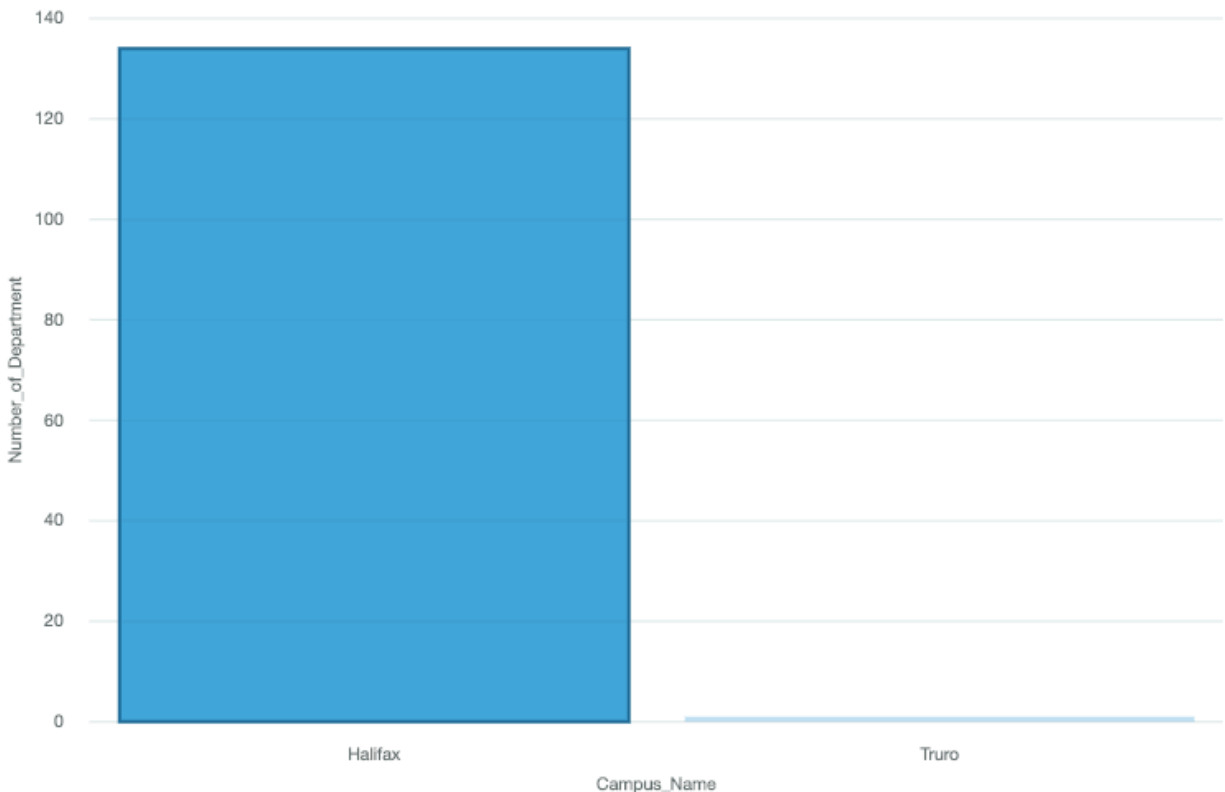   b. Only show those departments which offer more than one program

7. Multidimensional data analysis
    a. Analysis Number of Department each Campus [2]

b. Analysis Number of Employee each Department [2]

8. Using my BI Framework, I can only answer the question related to Which Department offer highest number of Program. And from the report it is Agriculture Department, and Computer Science are not offer the highest number of programs. However, Since I don't have enough data related to courses which department offered. So I can give the report for that.

Reference:
[1] A. Navlani, "Latent Semantic Analysis using Python," 2018,
https://www.datacamp.com/community/tutorials/discovering-hidden-topics-python.

[2] A. Hussain, "Cognos Analytics 11: Smarter Self-Service BI," 2018,
https://www.royalcyber.com/blog/business-intelligence/cognos-analytics-11-smarter-self-service-bi/.