DALHOUSIE UNIVERSITY

CSCI 5408 : Data Management, Warehousing & Analytics

Assignment 3: Sentiment & Semantic Analysis

Sharan Sudhir – B00833694

Sentiment Analysis:

1) Twitter Data Extracted in previous assignment is used for this assignment[1].
2) Twitter Data in JSON format is extracted, extracted data is cleaned.
3) After the data is cleaned, bag of words is created for each twitter text and stored in a list.
4) Positive and Negative words list is taken from online website and is loaded.
5) The created bag of words list is compared with the loaded Positive and Negative word List, based on which the final data is segregated.
6) After the final data has been segregated PrettyTable Library is used to create table.
7) For Tableau, the most occurred words in positive and negative words list is stored in a csv format file.
8) This csv format file is uploaded to the Tableau account for visualization[2].

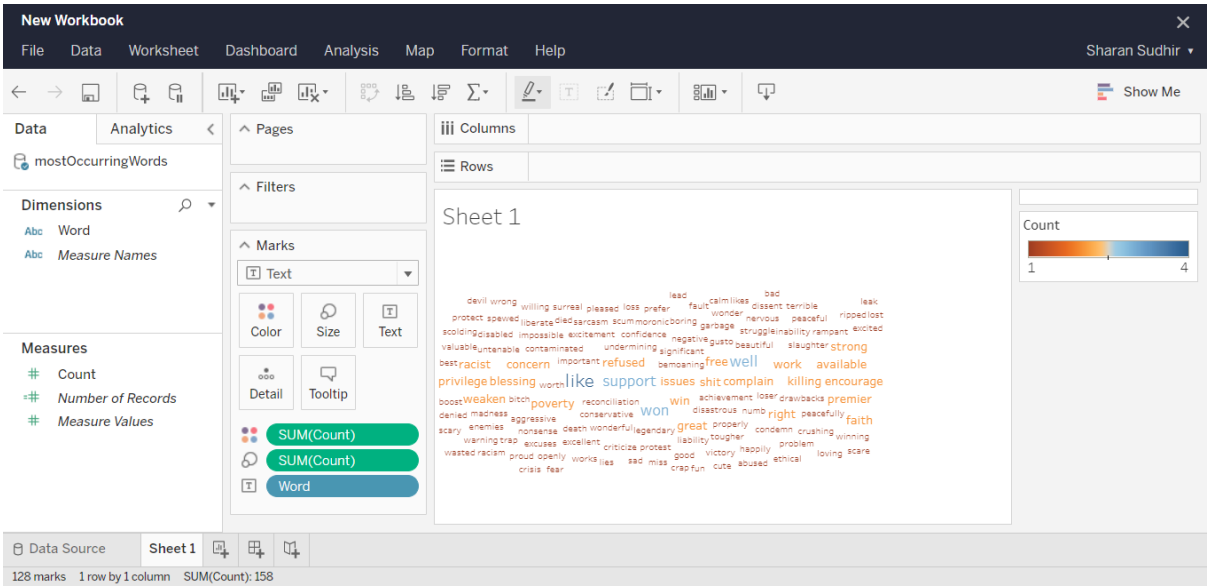

Figure 1Tableau Dashboard

Figure 2 Output

Semantic Analysis:

1) News Articles extracted during last assignment is used for this assignment.
2) News articles are stored in JSON format and loaded in the program.
3) After the data has been loaded, each tweet word is split and appended to a list.
4) Each word of the list is then compared with the search query and TFIDF table is generated.
5) The data is also used to calculate frequency of word Canada with each article's total words, using which the relative frequency is calculated.

Business Intelligence:

1)Fact and Dimension Table was created for Analytics-5408.

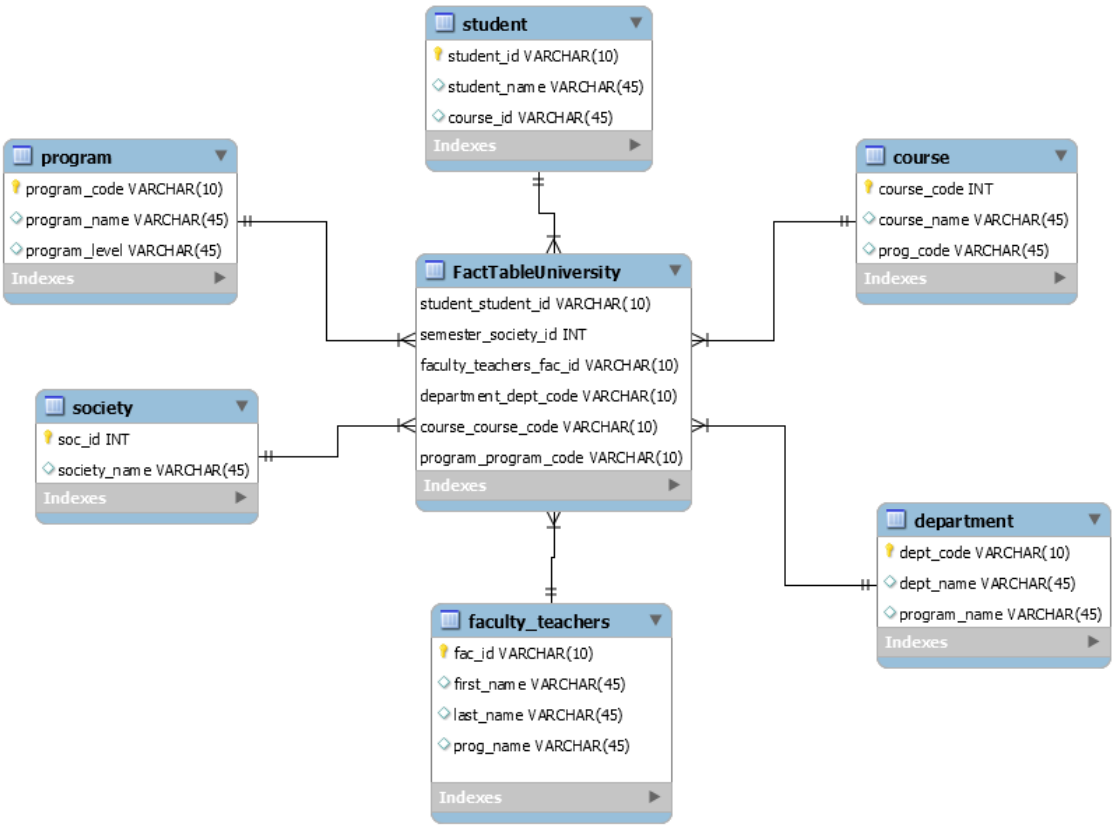

Figure 3 Fact & Dimenion table

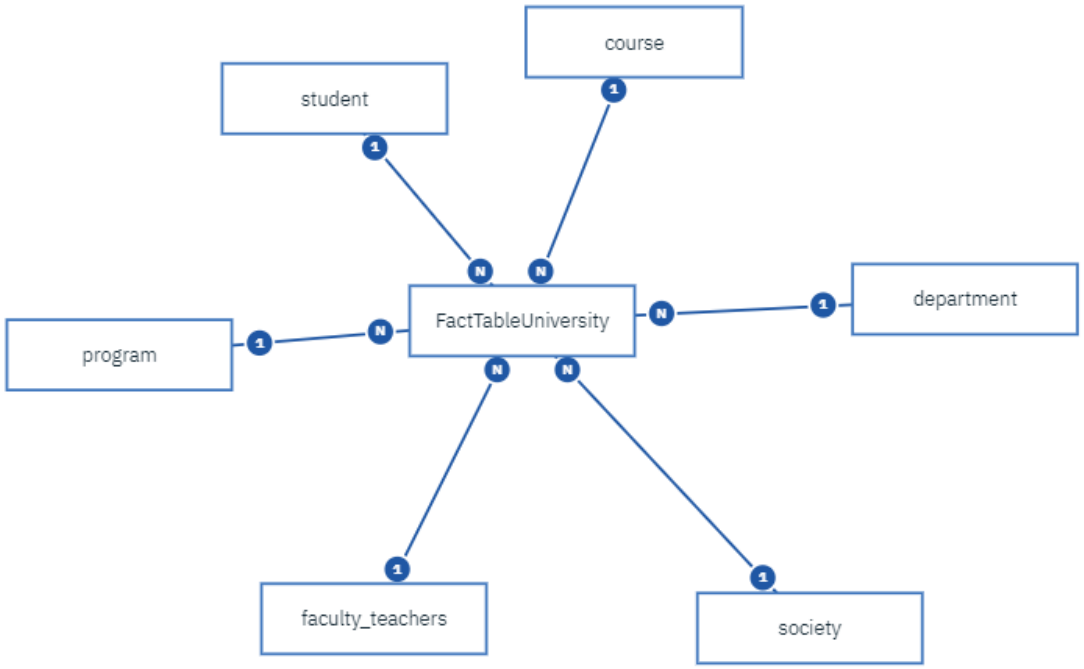

Figure 4 Star Schema

Star Schema Representation on Cognos BI

a. Does Computer Science offer highest number of programs?
Ans :Yes, screenshot attached below – Complete report generated using Cognos BI is kept in .zip folder

program_code	program_name	program_level
CS	Bachelor of Applied Computer Science	
	Bachelor of Computer Science	
	Master of Applied Computer Science	
	Master of Computer Science and PhD in Computer Science	
	Master of Electronic Commerce	
	Master of Health Informatics	
CS - Count	6	
HSCE	Diagnostic Medical Ultrasound	
	Medical Laboratory Technology (Post Diploma Only)	
	Nuclear Medicine Technology	
	Radiological Technology	
	Respiratory Therapy	
HSCE - Count	5	
PHYT	Joint MScPT-Rehabilitation Research	
	MSc (Rehabilitation Research)	
	MSc Physiotherapy	
	PhD in Health	
	PhD Interdisciplinary Research	
PHYT - Count	5	
POLI	BA Double Major (120 credit hours)	
	BA Honours Program in Political Science	
	BA Major in Political Science (120 credit hour)	
	BA Minor in Political Science (90 credit hours)	
	Combined Honours	
POLI - Count	5	

b. How many courses are there in each department or faculty?
Ans: Report generated using Cognos BI is placed in .zip folder.

course_code	course_name	prog_code
1105	Intro to Computer Programming	CS
1107	Social Computing	
1108	Experimental Robotics	
1109	Practical Data Science	
1110	Intro to Computer Science	
1120	Intro to Computer Systems	
1170	Intro to Web Design & Devel.	
1801	Case Studies in Comp & Society	
2100	Comm Skills: Oral/Written	
2110	Data Structures & Algorithms	
2112	Discrete Structures I	
2113	Discrete Structures II	
2114	Mathematical App. in Computing	
2122	Systems Programming	
2134	Software Development	
2141	Intro to Database Systems	
2170	Intro to Server Side Scripting	
2201	Introduction to Info Security	
2202	Comp. Modeling for Scientists	
2203	Data Science for Everyone	
2690	Intro to Software Projects	
2691	Introductory Project	
3101	Soc/Ethi/Prof Issues in Csci	
3130	Software Engineering	
3136	Principles/Programming Languag	
3151	Foundations of Machine Learn.	
3160	Designing User Interfaces	
3162	Digital Media	
3171	Network Computing	
3172	Web-Centric Computing	

References:

[1] Datawarehouse Assignment, "*Assignment 2*" in Brightspace, [online portal], 06 November 2019. Available: brightspace.com Reference Online, https://dal.brightspace.com/d2l/lms/dropbox/user/folders_history.d2l?db=62303&grpid=0&isprv=0&bp=0&ou=100142 [Accessed: November 30, 2019].

[2] Parul Pandey, "*Word Clouds in Tableau: Quick & Easy*" in Towards Data Science, [online article], 20 February 2019. Available: dzone.com Reference Online, https://towardsdatascience.com/word-clouds-in-tableau-quick-easy-e71519cf507a [Accessed: November 02, 2019].