

Big Data Science - Homework 3

Jude Naveen Raj Ilango - jni215

Prince Cluster Output

```
(base) [jni215@c38-05 jni215]$ ls -l
total 339735404
drwxr-x--- 3 jni215 jni215 4096 Dec 11 12:49 cityscapes
drwxr-x--- 5 jni215 jni215 4096 Dec 11 21:29 cityscapesValImages
-rw-r----- 1 jni215 jni215 170 Dec 3 15:20 cookies.txt
drwxr-x--- 4 jni215 jni215 4096 Dec 15 00:39 gan_dataset
drwxr-x--- 3 jni215 jni215 4096 Dec 3 15:25 gtFine
-rw-r----- 1 jni215 jni215 252567705 Mar 21 2016 gtFine_trainvaltest.zip
-rw-r----- 1 jni215 jni215 347397537544 Mar 23 2016 leftImg8bit_sequence_trainvaltest.zip
drwxr-x---+ 2 jni215 jni215 4096 Dec 11 14:41 outputs
-rw-r----- 1 jni215 jni215 57 Nov 26 17:48 post.txt
drwxr-x--- 3 jni215 jni215 4096 Dec 15 12:38 probunet_models
-rw-r----- 1 jni215 jni215 238803811 Mar 15 16:18 sentiment140.csv
(base) [jni215@c38-05 jni215]$ wc -l sentiment140.csv
1600000 sentiment140.csv
(base) [jni215@c38-05 jni215]$ head -n 5 sentiment140.csv
"0","1467810369","Mon Apr 06 22:19:45 PDT 2009","NO_QUERY","_TheSpecialOne_","@switchfoot http://twitpic.com/2y1zl - Awww, that's a bummer. You shoulda got David Carr of Third Day to do it. ;D"
"0","1467810672","Mon Apr 06 22:19:49 PDT 2009","NO_QUERY","scotthamilton","is upset that he can't update his Facebook by texting it... and might cry as a result School today also. Blah!"
"0","1467810917","Mon Apr 06 22:19:53 PDT 2009","NO_QUERY","mattycus","@Kenichan I dived many times for the ball. Managed to save 50% The rest go out of bounds"
"0","1467811184","Mon Apr 06 22:19:57 PDT 2009","NO_QUERY","ElleCTF","my whole body feels itchy and like its on fire "
"0","1467811193","Mon Apr 06 22:19:57 PDT 2009","NO_QUERY","Karoli","@nationwideclass no, it's not behaving at all. i'm mad. why am i here? because I can't see you all over there. "
```

Top 20 Hashtags	Top 20 Mentions
#fb	@mileycyrus
#followfriday	@tommcfly
#squarespace	@ddlovato
#iranelection	@jonasbrothers
#sebdays	@jonathanknight
#ff	@davidarchie
#musicmonday	@
#bgt	@the
#fail	@donniewahlberg
#marsiscoming	@taylorswift
#bsb	@jordanknight
#iphone	@mitchelmusso
#andyhurleyday	@selenagomez
#myweakness	@doughiemcfly
#mcflyforgermany	@work
#iremember	@aplusk
#f	@peterfacinelli
#asot	@dannymcfly
#.	@joeymcintyre
#trackle	@youngq

Observation:

Hashtags: The hashtags, besides generic social influencer tags such as those representing days of the week, bring out the current trends such as movies, technology (iphone), services (squarespace) and events.

Mentions: We observe that most of the mentions are about Singers and TV Actors. A conclusion that can be made from this is that most of the audience consists of a younger age group as a lot of the actors mentioned belong to popular Disney shows. It also brings out the most talked-about singers of the given time period.

N-Gram Analysis:

Top 20 1-Gram	Top 20 2-Gram	Top 20 3-Gram	Top 20 4-Gram
i you my have im get go so do just be its day now up good work like out all	i_have i_be i_dont i_think i_love i_cant i_miss i_get i_want gon_na i_just i_know wish_i i_can i_need wan_na i_feel i_wish i_hate i_do	i_wish_i i_wan_na wish_i_could i_think_i im_gon_na i_love_you i_miss_you i_dont_know i_miss_my i_dont_have wan_na_go i_dont_want i_feel_like i_know_i i_dont_think i_cant_wait just_wake_up i_be_so i_hope_you dont_wan_na	i_wish_i_could i_wan_na_go i_dont_wan_na just_get_back_from i_wish_i_have day_use_once_you get_follower_day_use follower_day_use_once everyone_you_train_pay add_everyone_you_train you_train_pay_vip once_you_add_everyone use_once_you_add you_add_everyone_you i_dont_know_what just_get_home_from wish_i_could_go im_gon_na_miss hope_you_feel_better dont_wan_na_go

Nouns-Only N-Gram Analysis:

Top 20 1-Gram	Top 20 2-Gram	Top 20 3-Gram	Top 20 4-Gram
im day today time lol work night home thanks morning tomorrow u twitter amp week fun haha thing love friend	im_im today_day day_today mother_day ice_cream im_home day_work im_work day_im day_everyone movie_award lol_im im_bed mtv_movie day_school father_day today_im follower_day work_today day_day	mtv_movie_award follower_day_everyone train_pay_vip day_everyone_train everyone_train_pay jon_amp_kate sebdays_sebdays_sebdays mother_day_mom work_work_work air_france_flight new_moon_trailer limit_story_ch father_day_dad air_france_plane got_ta_love ha_ha_ha mother_day_mother story_ch_stress mcflyforgermany_mcflyforg ermany_mcflyforgermany sun_jun_pdt	day_everyone_train_pay everyone_train_pay_vip follower_day_everyone_train sebdays_sebdays_sebdays_sebdays limit_story_ch_stress mcflyforgermany_mcflyforgermany_mcflyforgermany_mcflyforgermany rd_place_word_vote place_word_vote_time quothappy_birthday_roni_amp amp_mickeyquot_plz_plz hes_rd_place_word mickeyquot_plz_plz_plz plz_quothappy_birthday_roni birthday_roni_amp_mickeyquot roni_amp_mickeyquot_plz didnt_command_tag_tag tomfelton_hes_rd_place mtv_movie_award_tonight noone_question_minute_thinking question_minute_thinking_cap

POS Analysis:

Top 20 JJ-NN	Top 20 RB-VB	Top 20 JJ-VB	Top 20 VB-NN
last_night good_morning cant_wait good_luck good_night good_day happy_birthday last_day next_week great_day first_time next_time cant_sleep beautiful_day much_fun long_time last_week long_day happy_mothers few_days	just_got just_woke just_had just_finished not_going ive_been not_feeling still_have really_want just_saw now_im ill_be not_be just_found not_working just_want just_watched not_looking really_need just_made	cant_believe dont_want dont_know cant_find doesnt_want wont_be cant_get ive_got back_hurts ive_been ill_be cant_be cant_seem cant_afford im_going havent_been cant_go havent_seen much_needed wont_let	got_ta have_fun think_im got_home go_home get_followers add_everyone was_fun had_fun thank_god get_home take_care do_anything know_im be_fun do_something cant_sleep be_home going_home work_tomorrow

Dependency Parser Analysis:

Subj - Root - Obj
i_miss_you i_love_you i_love_it i_have_headache i_hate_it i_miss_it i_have_idea i_wan_na i_miss_he i_have_day i_think_im i_want_it i_have_dream i_do_what i_have_time i_make_it i_have_one i_cant_sleep i_see_you i_miss_friend

Entity Analysis:

Entities
day one night morning i today first last_night summer sunday monday two friday saturday weekend this_morning lol b june afternoon

BEST KEYWORD EXTRACTION METHOD : I believe that **POS ANALYSIS** gives the best keywords as it expresses the feelings felt by the tweeters and that is a good indicator of tweet sentiment.

Negative Sentences - POS Analysis:

Top 20 JJ-NN	Top 20 RB-VB	Top 20 JJ-VB	Top 20 VB-NN
last_night last_day next_week cant_wait good_morning cant_sleep good_luck last_week good_day long_day last_time sore_throat first_time next_time few_days long_time next_year bad_day good_night sad_day	just_got not_going not_feeling just_woke ive_been still_have just_had really_want not_working not_looking not_be just_want just_found now_im just_finished just_saw really_need ill_be only_have just_watched	cant_believe dont_want cant_find dont_know doesnt_want wont_be back_hurts cant_get cant_seem cant_afford ive_got ive_been cant_go cant_be wont_let havent_been cant_see havent_seen ill_be doesnt_know	got_ta think_im go_home got_home have_fun cant_sleep get_home do_anything work_tomorrow work_today do_something going_home come_home know_im have_time miss_u have_nothing be_home do_today was_fun

Positive Sentences - POS Analysis:

Top 20 JJ-NN	Top 20 RB-VB	Top 20 JJ-VB	Top 20 VB-NN
good_morning cant_wait last_night good_luck good_night happy_birthday good_day great_day happy_mothers beautiful_day next_week first_time much_fun next_time great_time nice_day good_times good_time long_time good_thing	just_got just_finished just_had just_woke ive_been ill_be just_saw just_made just_watched now_im finally_got still_have just_came well_done just_bought just_found just_wanted just_have not_be not_going	cant_believe dont_know much_needed ill_be dont_forget im_going much_appreciated dont_want ive_been ive_got sure_is ive_seen better_get short_stack super_excited wont_be havent_seen im_gon cant_get good_looking	have_fun got_ta get_followers add_everyone got_home think_im had_fun was_fun thank_god be_fun take_care having_fun know_im get_home be_home are_u do_something think_ill is_fun do_today

Observations:

We obtain pretty good results that talk about positive and negative feelings of the tweeter respectively. POS tagging using Verbs especially helps bring out the sentiment of the text.