Raw table creation query in Hive:

Create external table json\_table\_2 (line, string) ;

Loading data from twitter.json to raw table:

load data local inpath 'Desktop/Twitter.json' overwrite into table json\_table\_2;

Main Table creation query in Hive:

CREATE EXTERNAL TABLE json\_assign7(

retweet\_count int,

created\_at timestamp,

text string,

id bigint,

source string,

in\_reply\_to\_screen\_name string,

user struct<location: string, id: bigint, id\_str: bigint, name:string, screen\_name: string, geo\_enabled: boolean, lang: string,protected: boolean, verified: boolean, followers\_count:int,friends\_count: int, listed\_count: int, favourites\_count:int,statuses\_count: int, profile\_background\_color: int>,

contributors string,

is\_quote\_status boolean,

entities struct <user\_mentions:ARRAY<struct <screen\_name : string,name : string, id : bigint >>>);

Inseting data into main table from raw table:

insert into table json\_assign7 select cast(get\_json\_object(line,'$.retweet\_count') as int) as retweet\_count,

from\_unixtime(unix\_timestamp(get\_json\_object(line, "$.created\_at"),"EEE MMM d HH:mm:ss Z yyyy")) as created\_at,

get\_json\_object(line,'$.text') as text,

cast(get\_json\_object(line,'$.id') as bigint) as id,

get\_json\_object(line,'$.source') as source,

get\_json\_object(line,'$.in\_reply\_to\_screen\_name') as in\_reply\_to\_screen\_name,

named\_struct('location',get\_json\_object(line,'$.user.location'),'id',cast(get\_json\_object(line,'$.user.id') as bigint),'id\_str',cast(get\_json\_object(line,'$.user.id\_str') as bigint),'name',get\_json\_object(line,'$.user.name'),'screen\_name',get\_json\_object(line,'$.user.screen\_name'),'geo\_enabled',cast(get\_json\_object(line,'$.user.geo\_enabled') as boolean),'lang',get\_json\_object(line,'$.user.lang'),'protected',cast(get\_json\_object(line,'$.user.protected') as boolean),'verified',cast(get\_json\_object(line,'$.user.verified') as boolean),'followers\_count',cast(get\_json\_object(line,'$.user.followers\_count') as int),'friends\_count',cast(get\_json\_object(line,'$.user.friends\_count')as int),'listed\_count',cast(get\_json\_object(line,'$.user.listed\_count') as int),'favourites\_count',cast(get\_json\_object(line,'$.user.favourites\_count') as int),'statuses\_count',cast(get\_json\_object(line,'$.user.statuses\_count') as int),'profile\_background\_color',cast(get\_json\_object(line,'$.user.profile\_background\_color') as int)) as user,

get\_json\_object(line,'$.contributors') as contributors,

cast(get\_json\_object(line,'$.is\_quote\_status') as boolean) as is\_quote\_status,

named\_struct('user\_mentions',ARRAY(named\_struct('screen\_name',get\_json\_object(line,'$.entities.user\_mentions[0].screen\_name'),'name',get\_json\_object(line,'$.entities.user\_mentions[0].name'),'id',cast(get\_json\_object(line,'$.entities.user\_mentions[0].id') as bigint)))) as entities from json\_table\_2

Answers :

1. What are the hashtags used and how many times each are used?





Query:

select hashtag,count(\*) as No\_of\_hashtag from json\_assign7 lateral view explode(split(text, ' ')) b as hashtag where hashtag rlike "^#[a-zA-Z0-9]+$" group by hashtag order by No\_of\_hashtag;

Answer:

#1 1

#smallbiz 1

#WakeUpAmerica 1

#WCS16 1

#VoterFraud 1

#USA 1

#UPUPUP 1

#Trumpforlife 1

#TrumpMovement 1

#TeamTrump 1

#TBT 1

#SCOTUS 1

#Rio2016 1

#RiggedSystem 1

#POTUS 1

#Obamacare 1

#NotoTrump 1

#NeverTrump 1

#NATO 1

#MSM 1

#MAGA3pm 1

#LeadRight2016 1

#Hillaryforlife 1

#Hate 1

#GOPConvention 1

#Frozen 1

#DemsInPhilly 1

#DefendtheSecond 1

#DNC 1

#AmericasMerkel 1

#AlwaysTrump 1

#60Minutes 1

#TrumpTrain 2

#MAGATickets 2

#2A 2

#tcdisrupt 2

#ThrowbackThursday 2

#LawandOrder 2

#LESM 3

#Hillary 3

#Trump 4

#MakeAmericaSafeAgain 4

#NeverHillary 5

#RNCinCLE 10

#Trump2016 12

#AmericaFirst 13

#MakeAmericaGreatAgain 14

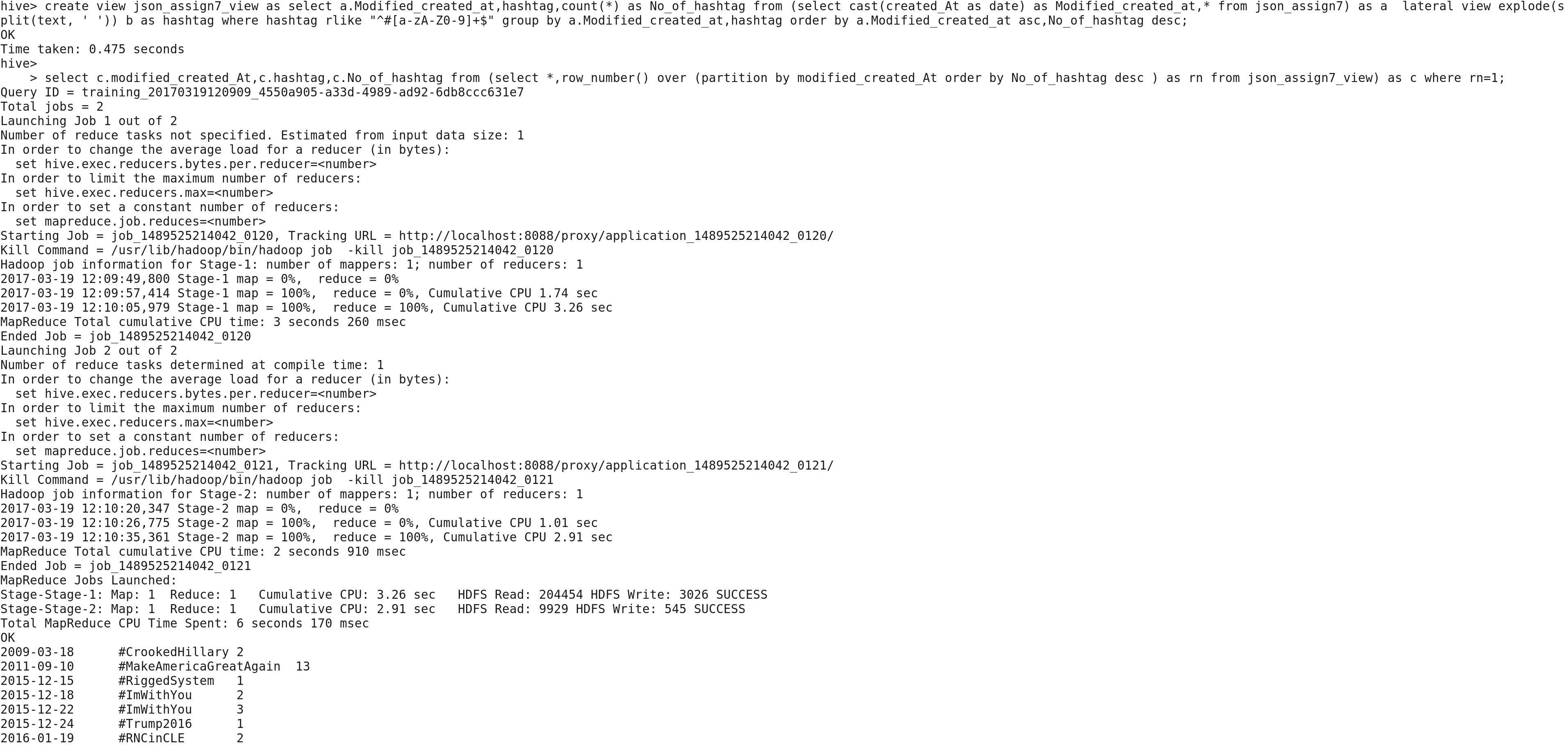
#CrookedHillary 16

#TrumpPence16 16

#MAGA 18

#ImWithYou 23

1. What is the most trending hashtag in a day and how many times is it tweeted? [Note: day should be in the format ‘yyyy-mm-dd’]





Query:

* create view json\_assign7\_view as select a.Modified\_created\_at,hashtag,count(\*) as No\_of\_hashtag from (select cast(created\_At as date) as Modified\_created\_at,\* from json\_assign7) as a lateral view explode(split(text, ' ')) b as hashtag where hashtag rlike "^#[a-zA-Z0-9]+$" group by a.Modified\_created\_at,hashtag order by a.Modified\_created\_at asc,No\_of\_hashtag desc;
* select c.modified\_created\_At,c.hashtag,c.No\_of\_hashtag from (select \*,row\_number() over (partition by modified\_created\_At order by No\_of\_hashtag desc ) as rn from json\_assign7\_view) as c where rn=1;

Answer:

2009-03-18 #CrookedHillary 2

2011-09-10 #MakeAmericaGreatAgain 13

2015-12-15 #RiggedSystem 1

2015-12-18 #ImWithYou 2

2015-12-22 #ImWithYou 3

2015-12-24 #Trump2016 1

2016-01-19 #RNCinCLE 2

2016-01-25 #Hillary 1

2016-01-29 #ImWithYou 2

2016-02-20 #MakeAmericaGreatAgain 1

2016-03-20 #CrookedHillary 2

2016-04-04 #DNC 1

2016-04-26 #Trumpforlife 1

2016-05-21 #MAGA 2

2016-05-24 #ImWithYou 2

2016-06-07 #MAGA 1

2016-07-08 #NeverHillary 2

2016-07-11 #ThrowbackThursday 1

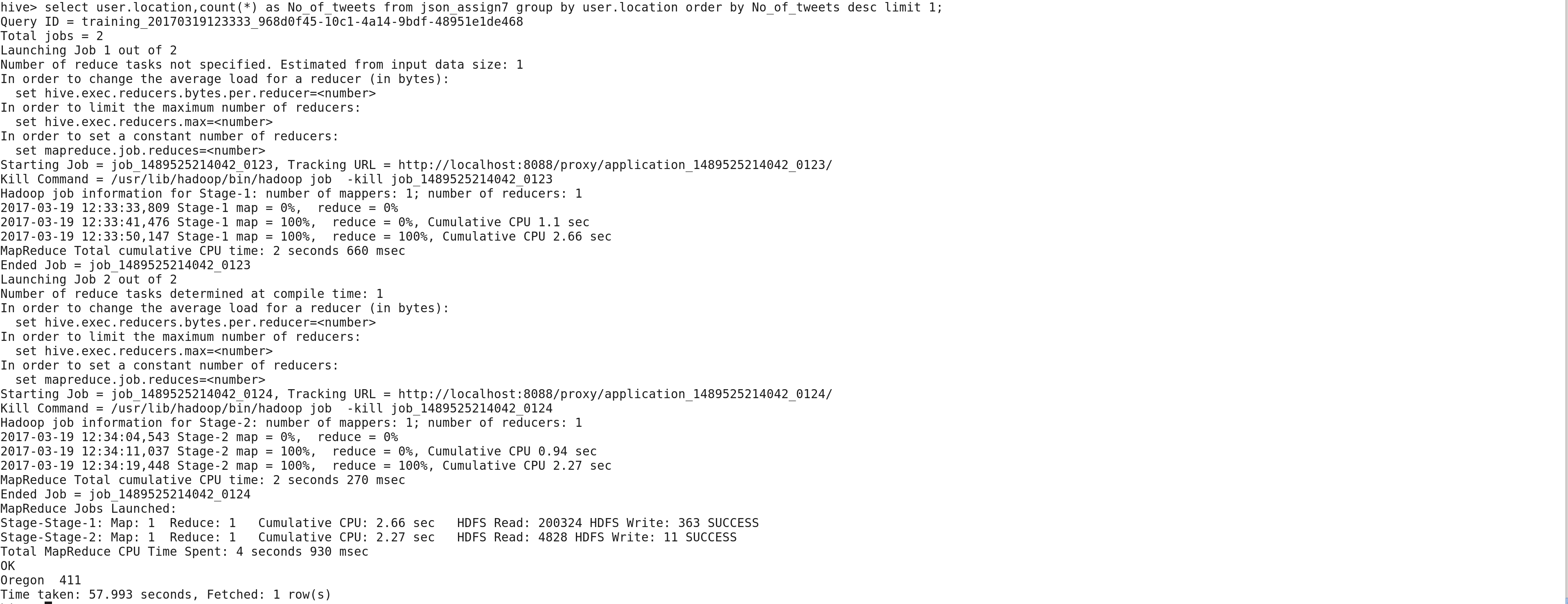
2016-07-24 #CrookedHillary 2

2016-07-29 #TrumpPence16 2

2016-08-23 #tcdisrupt 2

1. Which state users are most active and how many tweets are posted by State?

State having most active users:



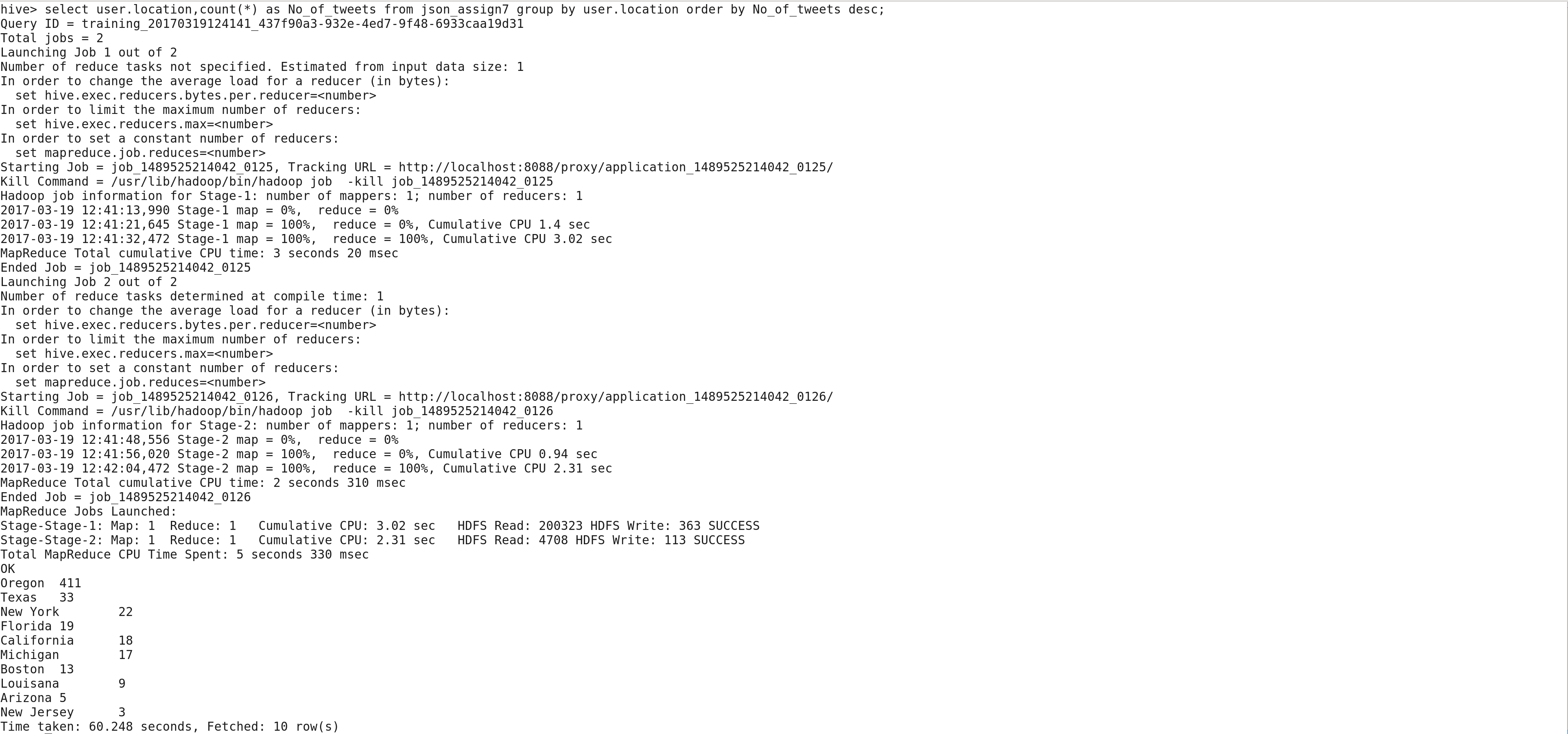
Query:

select user.location,count(\*) as No\_of\_tweets from json\_assign7 group by user.location order by No\_of\_tweets desc limit 1;

Answer:

Oregon 411

Tweets posted by state:



Query:

select user.location,count(\*) as No\_of\_tweets from json\_assign7 group by user.location order by No\_of\_tweets desc;

Answer:

Oregon 411

Texas 33

New York 22

Florida 19

California 18

Michigan 17

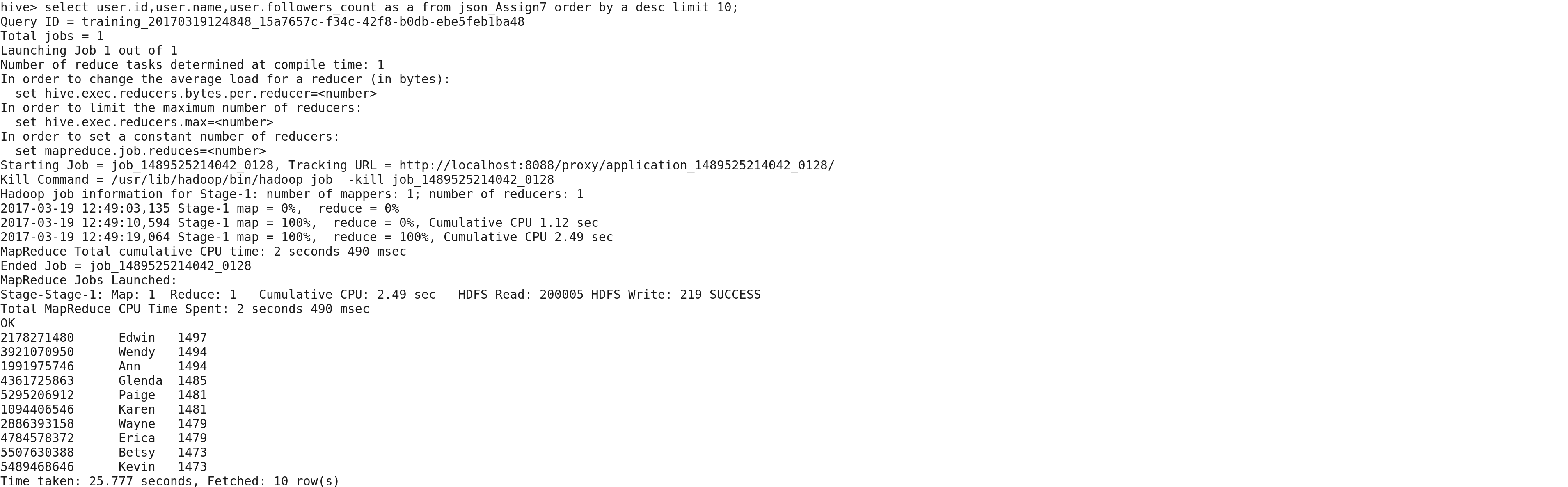
Boston 13

Louisana 9

Arizona 5

New Jersey 3

1. Based on the user’s followers count, who are the top ten users who have tweeted?



Query:

select user.id,user.name,user.followers\_count as a from json\_Assign7 order by a desc limit 10;

Answer:

2178271480 Edwin 1497

3921070950 Wendy 1494

1991975746 Ann 1494

4361725863 Glenda 1485

5295206912 Paige 1481

1094406546 Karen 1481

2886393158 Wayne 1479

4784578372 Erica 1479

5507630388 Betsy 1473

5489468646 Kevin 1473

1. What is the score for each tweet that was posted? Does the tweet have a positive or negative sentiment?

Query :

Creation of dict table :

create or replace external table dict(word string,score int)

row format delimited

fields terminated by '\t';

Loading of data from Dictionary.txt file

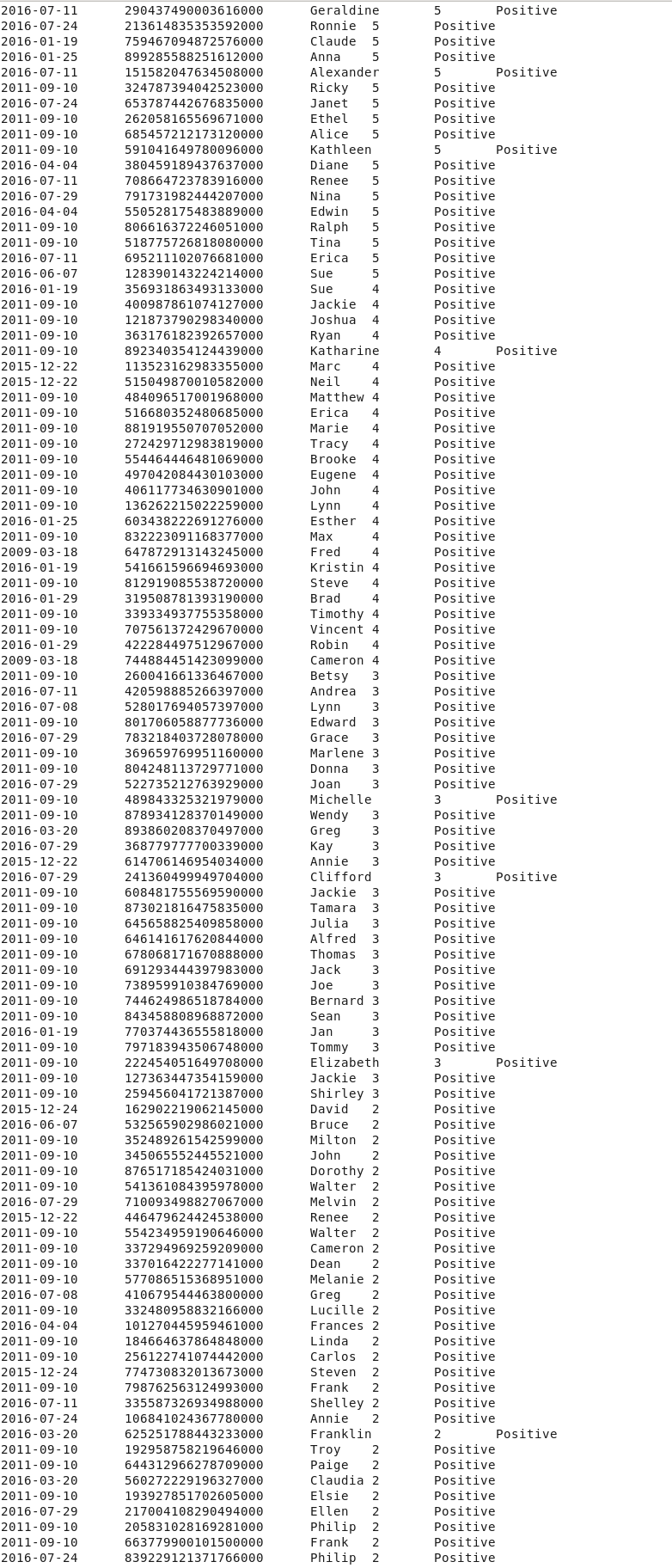
load data local inpath 'Desktop/Dictionary.txt' overwrite into table dict;

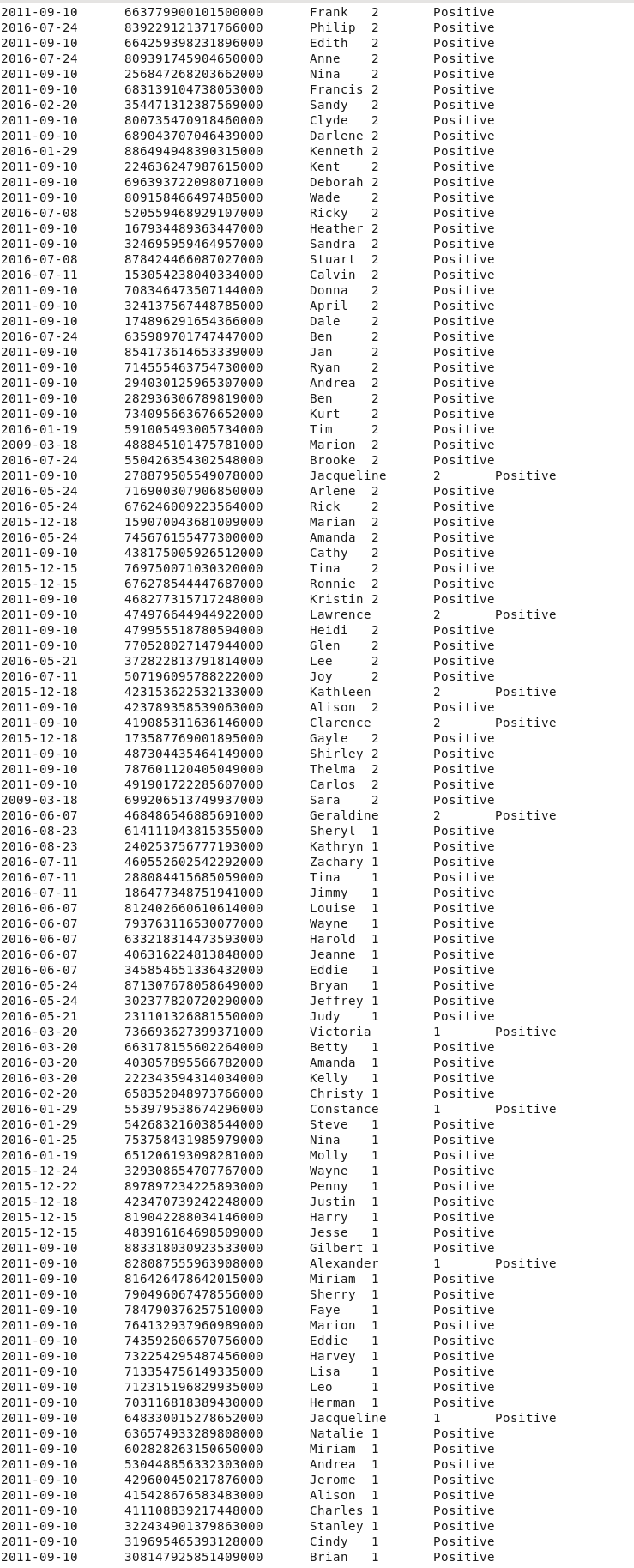
query :

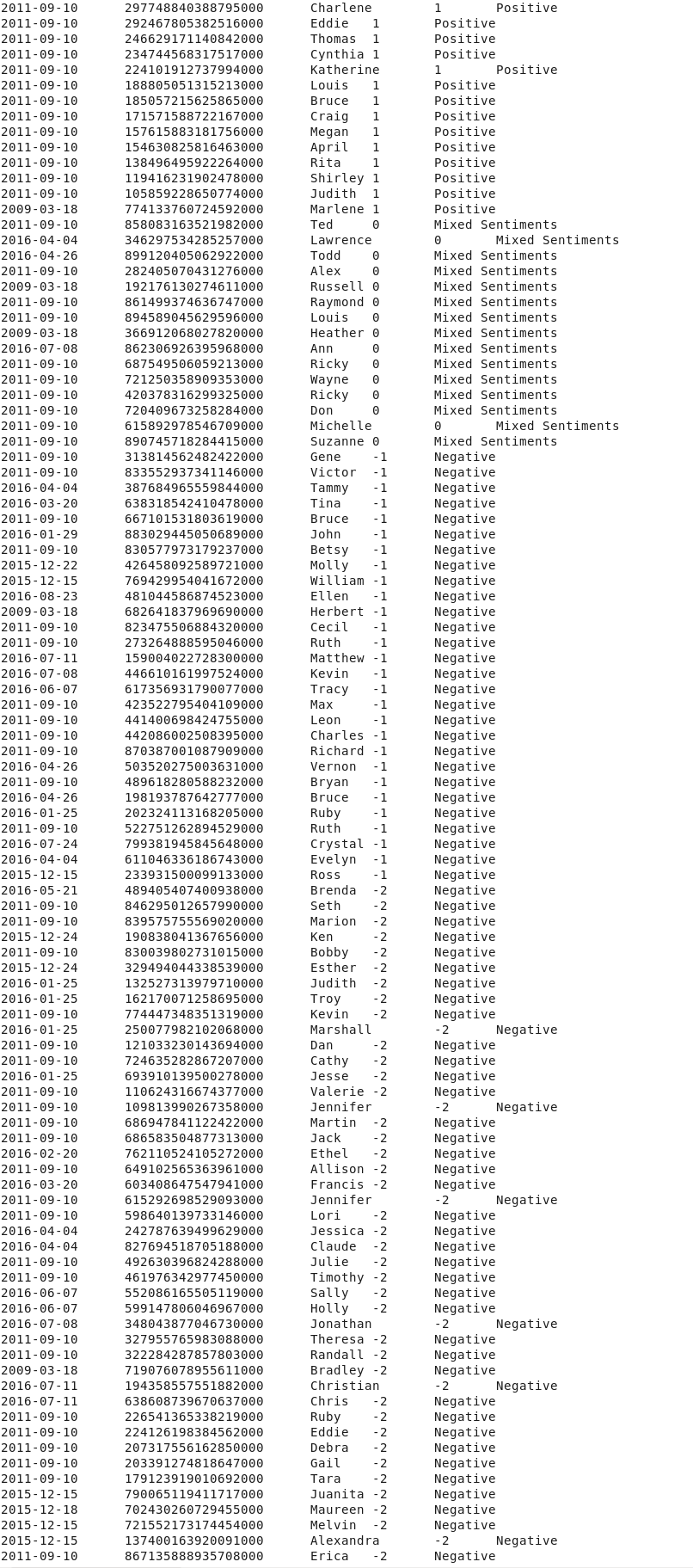
select temp.Modified\_created\_at,temp.id,temp.user\_name,temp.total\_score,case when temp.total\_score>0 then 'Positive' when temp.total\_score<0 then 'Negative' when temp.total\_score=0 then 'Mixed Sentiments' else 'No Sentiments' end from (select b.Modified\_created\_at,b.id,b.user\_name,sum(a.score) as total\_Score from dict as a right outer join (select date(created\_at) as Modified\_created\_at,id,user.name as user\_name, word from json\_assign7 lateral view explode(split(text,'\\W')) b as word) as b on a.word=lower(b.word) group by b.Modified\_created\_at,b.id,b.user\_name order by total\_Score desc ) as temp;

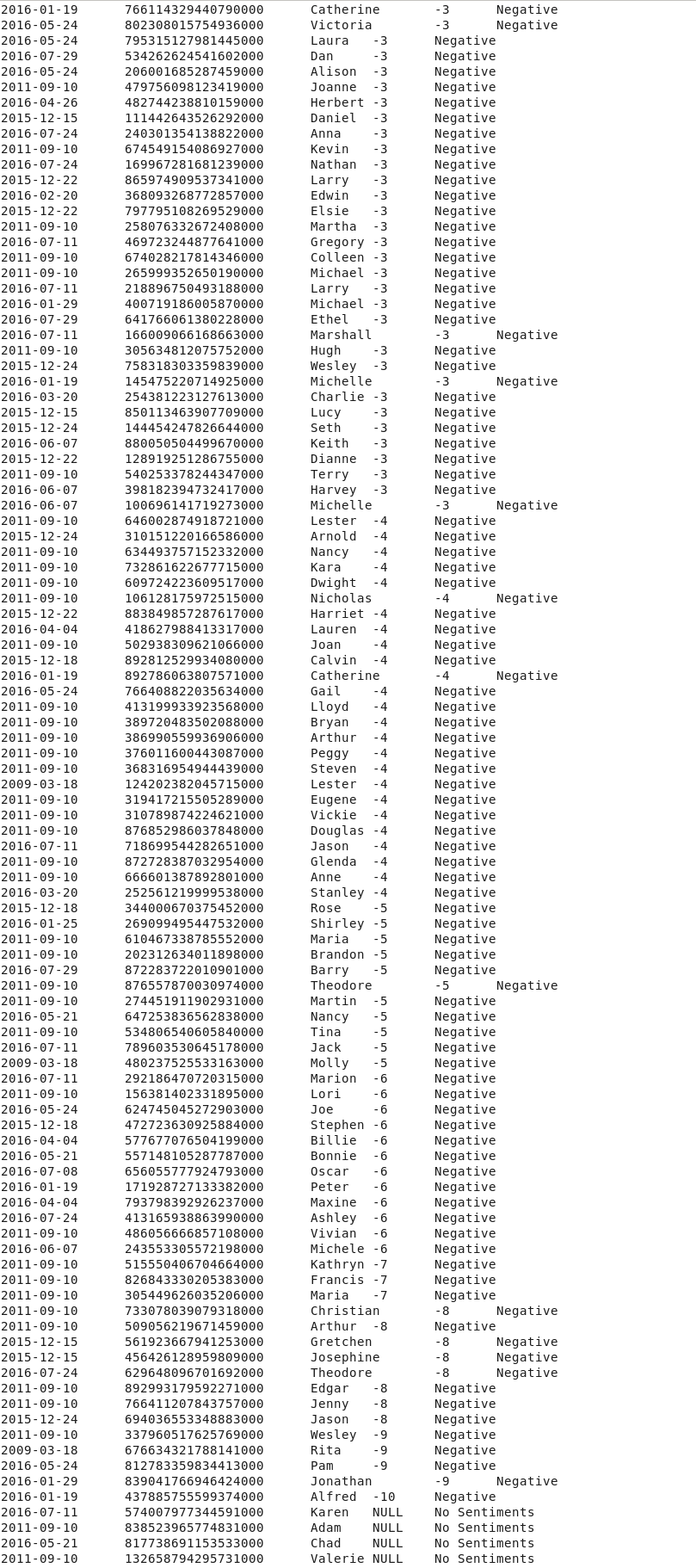
Answer:

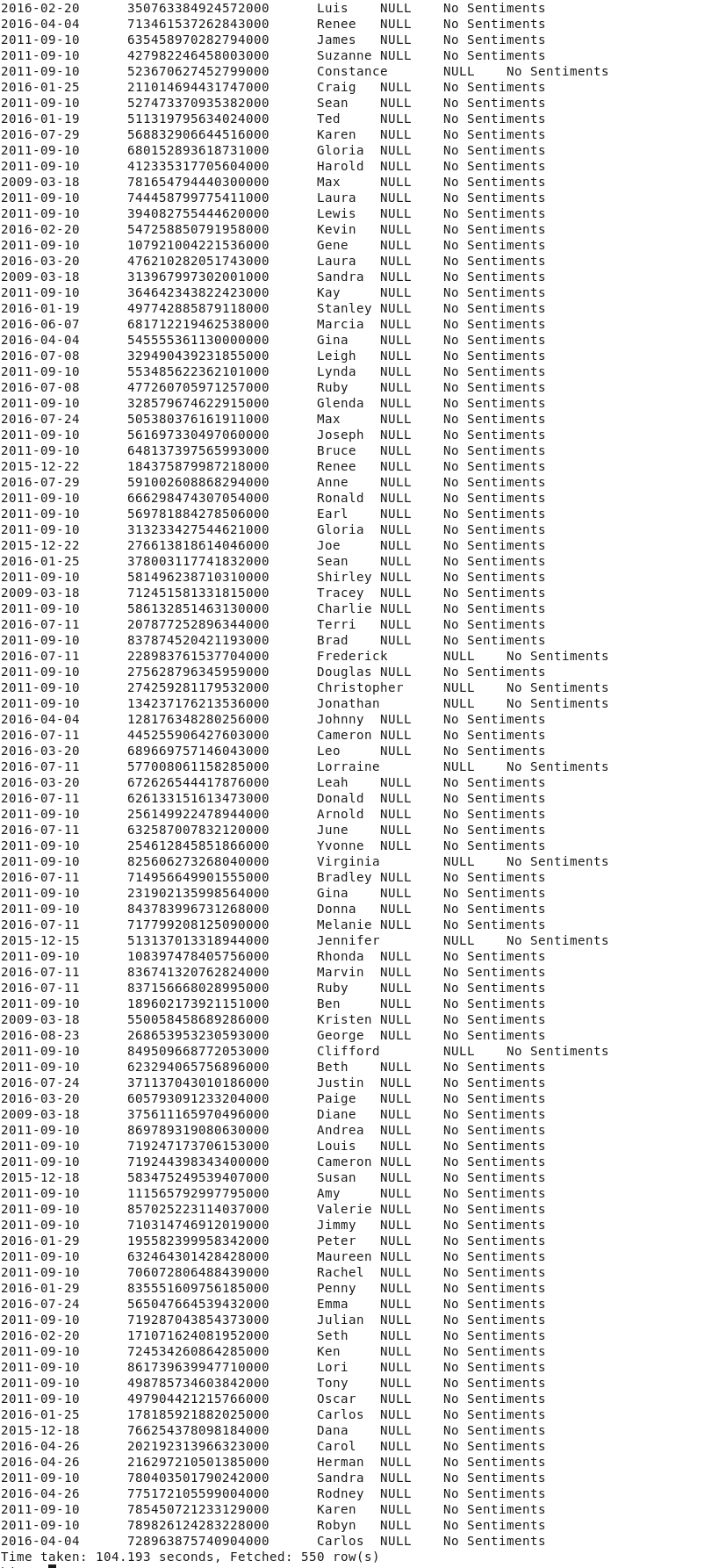












Answer:

2011-09-10 605834921020583000 Sharon 15 Positive

2016-05-21 845311701170392000 Eva 11 Positive

2011-09-10 599474773831196000 Gayle 10 Positive

2011-09-10 243288421754373000 Gretchen 10 Positive

2011-09-10 498961330469141000 Ethel 9 Positive

2011-09-10 431074808378375000 Wanda 9 Positive

2011-09-10 103386304747066000 Marion 9 Positive

2011-09-10 898147844068158000 Alexandra 8 Positive

2011-09-10 428691317581544000 Marlene 8 Positive

2011-09-10 138226334124559000 Kristin 8 Positive

2011-09-10 626124027133657000 Robin 8 Positive

2011-09-10 777009459272216000 Mary 8 Positive

2011-09-10 751600583089572000 Erika 8 Positive

2011-09-10 449403860041669000 Benjamin 8 Positive

2016-01-25 745597655932217000 Ben 8 Positive

2011-09-10 300969903297498000 Steve 7 Positive

2016-07-24 215124999743772000 Theresa 7 Positive

2009-03-18 477123889205614000 Judith 7 Positive

2015-12-24 483358381242562000 Jackie 7 Positive

2011-09-10 545893443885142000 Dorothy 7 Positive

2011-09-10 526352956630677000 Kyle 7 Positive

2011-09-10 508665556481819000 Emma 7 Positive

2011-09-10 486991358380482000 Sean 7 Positive

2011-09-10 484175757542947000 Amy 7 Positive

2015-12-15 785031848181625000 Hazel 7 Positive

2011-09-10 164894051801320000 Clyde 7 Positive

2016-07-08 562903027578790000 Bob 7 Positive

2011-09-10 796414064963404000 Marvin 7 Positive

2011-09-10 822954972084547000 Brian 6 Positive

2011-09-10 838602765409998000 Sheryl 6 Positive

2011-09-10 872586346272913000 Jeanette 6 Positive

2016-07-24 320144721281132000 Patrick 6 Positive

2011-09-10 241361331445150000 Terri 6 Positive

2016-07-11 560940286305638000 Lee 6 Positive

2011-09-10 262429684879633000 Marguerite 6 Positive

2011-09-10 276423954974353000 Rose 6 Positive

2011-09-10 330899242233600000 Herbert 6 Positive

2011-09-10 339186342910956000 Terry 6 Positive

2016-06-07 329223296619955000 Ronald 6 Positive

2016-05-21 394544985816104000 Jesse 6 Positive

2016-04-04 650998480861539000 Penny 6 Positive

2011-09-10 605024365728698000 Paul 6 Positive

2011-09-10 704777688247274000 Kelly 6 Positive

2016-01-29 178834874759199000 Thelma 6 Positive

2016-01-25 647254983184067000 Sherry 6 Positive

2011-09-10 767632381389183000 Glenn 6 Positive

2016-04-26 500296478047702000 Glen 5 Positive

2011-09-10 425237326343322000 Louis 5 Positive

2011-09-10 105743242835042000 Alex 5 Positive

2016-01-19 683815678105403000 Gail 5 Positive

2016-07-11 290437490003616000 Geraldine 5 Positive

2016-07-24 213614835353592000 Ronnie 5 Positive

2016-01-19 759467094872576000 Claude 5 Positive

2016-01-25 899285588251612000 Anna 5 Positive

2016-07-11 151582047634508000 Alexander 5 Positive

2011-09-10 324787394042523000 Ricky 5 Positive

2016-07-24 653787442676835000 Janet 5 Positive

2011-09-10 262058165569671000 Ethel 5 Positive

2011-09-10 685457212173120000 Alice 5 Positive

2011-09-10 591041649780096000 Kathleen 5 Positive

2016-04-04 380459189437637000 Diane 5 Positive

2016-07-11 708664723783916000 Renee 5 Positive

2016-07-29 791731982444207000 Nina 5 Positive

2016-04-04 550528175483889000 Edwin 5 Positive

2011-09-10 806616372246051000 Ralph 5 Positive

2011-09-10 518775726818080000 Tina 5 Positive

2016-07-11 695211102076681000 Erica 5 Positive

2016-06-07 128390143224214000 Sue 5 Positive

2016-01-19 356931863493133000 Sue 4 Positive

2011-09-10 400987861074127000 Jackie 4 Positive

2011-09-10 121873790298340000 Joshua 4 Positive

2011-09-10 363176182392657000 Ryan 4 Positive

2011-09-10 892340354124439000 Katharine 4 Positive

2015-12-22 113523162983355000 Marc 4 Positive

2015-12-22 515049870010582000 Neil 4 Positive

2011-09-10 484096517001968000 Matthew 4 Positive

2011-09-10 516680352480685000 Erica 4 Positive

2011-09-10 881919550707052000 Marie 4 Positive

2011-09-10 272429712983819000 Tracy 4 Positive

2011-09-10 554464446481069000 Brooke 4 Positive

2011-09-10 497042084430103000 Eugene 4 Positive

2011-09-10 406117734630901000 John 4 Positive

2011-09-10 136262215022259000 Lynn 4 Positive

2016-01-25 603438222691276000 Esther 4 Positive

2011-09-10 832223091168377000 Max 4 Positive

2009-03-18 647872913143245000 Fred 4 Positive

2016-01-19 541661596694693000 Kristin 4 Positive

2011-09-10 812919085538720000 Steve 4 Positive

2016-01-29 319508781393190000 Brad 4 Positive

2011-09-10 339334937755358000 Timothy 4 Positive

2011-09-10 707561372429670000 Vincent 4 Positive

2016-01-29 422284497512967000 Robin 4 Positive

2009-03-18 744884451423099000 Cameron 4 Positive

2011-09-10 260041661336467000 Betsy 3 Positive

2016-07-11 420598885266397000 Andrea 3 Positive

2016-07-08 528017694057397000 Lynn 3 Positive

2011-09-10 801706058877736000 Edward 3 Positive

2016-07-29 783218403728078000 Grace 3 Positive

2011-09-10 369659769951160000 Marlene 3 Positive

2011-09-10 804248113729771000 Donna 3 Positive

2016-07-29 522735212763929000 Joan 3 Positive

2011-09-10 489843325321979000 Michelle 3 Positive

2011-09-10 878934128370149000 Wendy 3 Positive

2016-03-20 893860208370497000 Greg 3 Positive

2016-07-29 368779777700339000 Kay 3 Positive

2015-12-22 614706146954034000 Annie 3 Positive

2016-07-29 241360499949704000 Clifford 3 Positive

2011-09-10 608481755569590000 Jackie 3 Positive

2011-09-10 873021816475835000 Tamara 3 Positive

2011-09-10 645658825409858000 Julia 3 Positive

2011-09-10 646141617620844000 Alfred 3 Positive

2011-09-10 678068171670888000 Thomas 3 Positive

2011-09-10 691293444397983000 Jack 3 Positive

2011-09-10 738959910384769000 Joe 3 Positive

2011-09-10 744624986518784000 Bernard 3 Positive

2011-09-10 843458808968872000 Sean 3 Positive

2016-01-19 770374436555818000 Jan 3 Positive

2011-09-10 797183943506748000 Tommy 3 Positive

2011-09-10 222454051649708000 Elizabeth 3 Positive

2011-09-10 127363447354159000 Jackie 3 Positive

2011-09-10 259456041721387000 Shirley 3 Positive

2015-12-24 162902219062145000 David 2 Positive

2016-06-07 532565902986021000 Bruce 2 Positive

2011-09-10 352489261542599000 Milton 2 Positive

2011-09-10 345065552445521000 John 2 Positive

2011-09-10 876517185424031000 Dorothy 2 Positive

2011-09-10 541361084395978000 Walter 2 Positive

2016-07-29 710093498827067000 Melvin 2 Positive

2015-12-22 446479624424538000 Renee 2 Positive

2011-09-10 554234959190646000 Walter 2 Positive

2011-09-10 337294969259209000 Cameron 2 Positive

2011-09-10 337016422277141000 Dean 2 Positive

2011-09-10 577086515368951000 Melanie 2 Positive

2016-07-08 410679544463800000 Greg 2 Positive

2011-09-10 332480958832166000 Lucille 2 Positive

2016-04-04 101270445959461000 Frances 2 Positive

2011-09-10 184664637864848000 Linda 2 Positive

2011-09-10 256122741074442000 Carlos 2 Positive

2015-12-24 774730832013673000 Steven 2 Positive

2011-09-10 798762563124993000 Frank 2 Positive

2016-07-11 335587326934988000 Shelley 2 Positive

2016-07-24 106841024367780000 Annie 2 Positive

2016-03-20 625251788443233000 Franklin 2 Positive

2011-09-10 192958758219646000 Troy 2 Positive

2011-09-10 644312966278709000 Paige 2 Positive

2016-03-20 560272229196327000 Claudia 2 Positive

2011-09-10 193927851702605000 Elsie 2 Positive

2016-07-29 217004108290494000 Ellen 2 Positive

2011-09-10 205831028169281000 Philip 2 Positive

2011-09-10 663779900101500000 Frank 2 Positive

2016-07-24 839229121371766000 Philip 2 Positive

2011-09-10 664259398231896000 Edith 2 Positive

2016-07-24 809391745904650000 Anne 2 Positive

2011-09-10 256847268203662000 Nina 2 Positive

2011-09-10 683139104738053000 Francis 2 Positive

2016-02-20 354471312387569000 Sandy 2 Positive

2011-09-10 800735470918460000 Clyde 2 Positive

2011-09-10 689043707046439000 Darlene 2 Positive

2016-01-29 886494948390315000 Kenneth 2 Positive

2011-09-10 224636247987615000 Kent 2 Positive

2011-09-10 696393722098071000 Deborah 2 Positive

2011-09-10 809158466497485000 Wade 2 Positive

2016-07-08 520559468929107000 Ricky 2 Positive

2011-09-10 167934489363447000 Heather 2 Positive

2011-09-10 324695959464957000 Sandra 2 Positive

2016-07-08 878424466087027000 Stuart 2 Positive

2016-07-11 153054238040334000 Calvin 2 Positive

2011-09-10 708346473507144000 Donna 2 Positive

2011-09-10 324137567448785000 April 2 Positive

2011-09-10 174896291654366000 Dale 2 Positive

2016-07-24 635989701747447000 Ben 2 Positive

2011-09-10 854173614653339000 Jan 2 Positive

2011-09-10 714555463754730000 Ryan 2 Positive

2011-09-10 294030125965307000 Andrea 2 Positive

2011-09-10 282936306789819000 Ben 2 Positive

2011-09-10 734095663676652000 Kurt 2 Positive

2016-01-19 591005493005734000 Tim 2 Positive

2009-03-18 488845101475781000 Marion 2 Positive

2016-07-24 550426354302548000 Brooke 2 Positive

2011-09-10 278879505549078000 Jacqueline 2 Positive

2016-05-24 716900307906850000 Arlene 2 Positive

2016-05-24 676246009223564000 Rick 2 Positive

2015-12-18 159070043681009000 Marian 2 Positive

2016-05-24 745676155477300000 Amanda 2 Positive

2011-09-10 438175005926512000 Cathy 2 Positive

2015-12-15 769750071030320000 Tina 2 Positive

2015-12-15 676278544447687000 Ronnie 2 Positive

2011-09-10 468277315717248000 Kristin 2 Positive

2011-09-10 474976644944922000 Lawrence 2 Positive

2011-09-10 479955518780594000 Heidi 2 Positive

2011-09-10 770528027147944000 Glen 2 Positive

2016-05-21 372822813791814000 Lee 2 Positive

2016-07-11 507196095788222000 Joy 2 Positive

2015-12-18 423153622532133000 Kathleen 2 Positive

2011-09-10 423789358539063000 Alison 2 Positive

2011-09-10 419085311636146000 Clarence 2 Positive

2015-12-18 173587769001895000 Gayle 2 Positive

2011-09-10 487304435464149000 Shirley 2 Positive

2011-09-10 787601120405049000 Thelma 2 Positive

2011-09-10 491901722285607000 Carlos 2 Positive

2009-03-18 699206513749937000 Sara 2 Positive

2016-06-07 468486546885691000 Geraldine 2 Positive

2016-08-23 614111043815355000 Sheryl 1 Positive

2016-08-23 240253756777193000 Kathryn 1 Positive

2016-07-11 460552602542292000 Zachary 1 Positive

2016-07-11 288084415685059000 Tina 1 Positive

2016-07-11 186477348751941000 Jimmy 1 Positive

2016-06-07 812402660610614000 Louise 1 Positive

2016-06-07 793763116530077000 Wayne 1 Positive

2016-06-07 633218314473593000 Harold 1 Positive

2016-06-07 406316224813848000 Jeanne 1 Positive

2016-06-07 345854651336432000 Eddie 1 Positive

2016-05-24 871307678058649000 Bryan 1 Positive

2016-05-24 302377820720290000 Jeffrey 1 Positive

2016-05-21 231101326881550000 Judy 1 Positive

2016-03-20 736693627399371000 Victoria 1 Positive

2016-03-20 663178155602264000 Betty 1 Positive

2016-03-20 403057895566782000 Amanda 1 Positive

2016-03-20 222343594314034000 Kelly 1 Positive

2016-02-20 658352048973766000 Christy 1 Positive

2016-01-29 553979538674296000 Constance 1 Positive

2016-01-29 542683216038544000 Steve 1 Positive

2016-01-25 753758431985979000 Nina 1 Positive

2016-01-19 651206193098281000 Molly 1 Positive

2015-12-24 329308654707767000 Wayne 1 Positive

2015-12-22 897897234225893000 Penny 1 Positive

2015-12-18 423470739242248000 Justin 1 Positive

2015-12-15 819042288034146000 Harry 1 Positive

2015-12-15 483916164698509000 Jesse 1 Positive

2011-09-10 883318030923533000 Gilbert 1 Positive

2011-09-10 828087555963908000 Alexander 1 Positive

2011-09-10 816426478642015000 Miriam 1 Positive

2011-09-10 790496067478556000 Sherry 1 Positive

2011-09-10 784790376257510000 Faye 1 Positive

2011-09-10 764132937960989000 Marion 1 Positive

2011-09-10 743592606570756000 Eddie 1 Positive

2011-09-10 732254295487456000 Harvey 1 Positive

2011-09-10 713354756149335000 Lisa 1 Positive

2011-09-10 712315196829935000 Leo 1 Positive

2011-09-10 703116818389430000 Herman 1 Positive

2011-09-10 648330015278652000 Jacqueline 1 Positive

2011-09-10 636574933289808000 Natalie 1 Positive

2011-09-10 602828263150650000 Miriam 1 Positive

2011-09-10 530448856332303000 Andrea 1 Positive

2011-09-10 429600450217876000 Jerome 1 Positive

2011-09-10 415428676583483000 Alison 1 Positive

2011-09-10 411108839217448000 Charles 1 Positive

2011-09-10 322434901379863000 Stanley 1 Positive

2011-09-10 319695465393128000 Cindy 1 Positive

2011-09-10 308147925851409000 Brian 1 Positive

2011-09-10 297748840388795000 Charlene 1 Positive

2011-09-10 292467805382516000 Eddie 1 Positive

2011-09-10 246629171140842000 Thomas 1 Positive

2011-09-10 234744568317517000 Cynthia 1 Positive

2011-09-10 224101912737994000 Katherine 1 Positive

2011-09-10 188805051315213000 Louis 1 Positive

2011-09-10 185057215625865000 Bruce 1 Positive

2011-09-10 171571588722167000 Craig 1 Positive

2011-09-10 157615883181756000 Megan 1 Positive

2011-09-10 154630825816463000 April 1 Positive

2011-09-10 138496495922264000 Rita 1 Positive

2011-09-10 119416231902478000 Shirley 1 Positive

2011-09-10 105859228650774000 Judith 1 Positive

2009-03-18 774133760724592000 Marlene 1 Positive

2011-09-10 858083163521982000 Ted 0 Mixed Sentiments

2016-04-04 346297534285257000 Lawrence 0 Mixed Sentiments

2016-04-26 899120405062922000 Todd 0 Mixed Sentiments

2011-09-10 282405070431276000 Alex 0 Mixed Sentiments

2009-03-18 192176130274611000 Russell 0 Mixed Sentiments

2011-09-10 861499374636747000 Raymond 0 Mixed Sentiments

2011-09-10 894589045629596000 Louis 0 Mixed Sentiments

2009-03-18 366912068027820000 Heather 0 Mixed Sentiments

2016-07-08 862306926395968000 Ann 0 Mixed Sentiments

2011-09-10 687549506059213000 Ricky 0 Mixed Sentiments

2011-09-10 721250358909353000 Wayne 0 Mixed Sentiments

2011-09-10 420378316299325000 Ricky 0 Mixed Sentiments

2011-09-10 720409673258284000 Don 0 Mixed Sentiments

2011-09-10 615892978546709000 Michelle 0 Mixed Sentiments

2011-09-10 890745718284415000 Suzanne 0 Mixed Sentiments

2011-09-10 313814562482422000 Gene -1 Negative

2011-09-10 833552937341146000 Victor -1 Negative

2016-04-04 387684965559844000 Tammy -1 Negative

2016-03-20 638318542410478000 Tina -1 Negative

2011-09-10 667101531803619000 Bruce -1 Negative

2016-01-29 883029445050689000 John -1 Negative

2011-09-10 830577973179237000 Betsy -1 Negative

2015-12-22 426458092589721000 Molly -1 Negative

2015-12-15 769429954041672000 William -1 Negative

2016-08-23 481044586874523000 Ellen -1 Negative

2009-03-18 682641837969690000 Herbert -1 Negative

2011-09-10 823475506884320000 Cecil -1 Negative

2011-09-10 273264888595046000 Ruth -1 Negative

2016-07-11 159004022728300000 Matthew -1 Negative

2016-07-08 446610161997524000 Kevin -1 Negative

2016-06-07 617356931790077000 Tracy -1 Negative

2011-09-10 423522795404109000 Max -1 Negative

2011-09-10 441400698424755000 Leon -1 Negative

2011-09-10 442086002508395000 Charles -1 Negative

2011-09-10 870387001087909000 Richard -1 Negative

2016-04-26 503520275003631000 Vernon -1 Negative

2011-09-10 489618280588232000 Bryan -1 Negative

2016-04-26 198193787642777000 Bruce -1 Negative

2016-01-25 202324113168205000 Ruby -1 Negative

2011-09-10 522751262894529000 Ruth -1 Negative

2016-07-24 799381945845648000 Crystal -1 Negative

2016-04-04 611046336186743000 Evelyn -1 Negative

2015-12-15 233931500099133000 Ross -1 Negative

2016-05-21 489405407400938000 Brenda -2 Negative

2011-09-10 846295012657990000 Seth -2 Negative

2011-09-10 839575755569020000 Marion -2 Negative

2015-12-24 190838041367656000 Ken -2 Negative

2011-09-10 830039802731015000 Bobby -2 Negative

2015-12-24 329494044338539000 Esther -2 Negative

2016-01-25 132527313979710000 Judith -2 Negative

2016-01-25 162170071258695000 Troy -2 Negative

2011-09-10 774447348351319000 Kevin -2 Negative

2016-01-25 250077982102068000 Marshall -2 Negative

2011-09-10 121033230143694000 Dan -2 Negative

2011-09-10 724635282867207000 Cathy -2 Negative

2016-01-25 693910139500278000 Jesse -2 Negative

2011-09-10 110624316674377000 Valerie -2 Negative

2011-09-10 109813990267358000 Jennifer -2 Negative

2011-09-10 686947841122422000 Martin -2 Negative

2011-09-10 686583504877313000 Jack -2 Negative

2016-02-20 762110524105272000 Ethel -2 Negative

2011-09-10 649102565363961000 Allison -2 Negative

2016-03-20 603408647547941000 Francis -2 Negative

2011-09-10 615292698529093000 Jennifer -2 Negative

2011-09-10 598640139733146000 Lori -2 Negative

2016-04-04 242787639499629000 Jessica -2 Negative

2016-04-04 827694518705188000 Claude -2 Negative

2011-09-10 492630396824288000 Julie -2 Negative

2011-09-10 461976342977450000 Timothy -2 Negative

2016-06-07 552086165505119000 Sally -2 Negative

2016-06-07 599147806046967000 Holly -2 Negative

2016-07-08 348043877046730000 Jonathan -2 Negative

2011-09-10 327955765983088000 Theresa -2 Negative

2011-09-10 322284287857803000 Randall -2 Negative

2009-03-18 719076078955611000 Bradley -2 Negative

2016-07-11 194358557551882000 Christian -2 Negative

2016-07-11 638608739670637000 Chris -2 Negative

2011-09-10 226541365338219000 Ruby -2 Negative

2011-09-10 224126198384562000 Eddie -2 Negative

2011-09-10 207317556162850000 Debra -2 Negative

2011-09-10 203391274818647000 Gail -2 Negative

2011-09-10 179123919010692000 Tara -2 Negative

2015-12-15 790065119411717000 Juanita -2 Negative

2015-12-18 702430260729455000 Maureen -2 Negative

2015-12-15 721552173174454000 Melvin -2 Negative

2015-12-15 137400163920091000 Alexandra -2 Negative

2011-09-10 867135888935708000 Erica -2 Negative

2016-01-19 766114329440790000 Catherine -3 Negative

2016-05-24 802308015754936000 Victoria -3 Negative

2016-05-24 795315127981445000 Laura -3 Negative

2016-07-29 534262624541602000 Dan -3 Negative

2016-05-24 206001685287459000 Alison -3 Negative

2011-09-10 479756098123419000 Joanne -3 Negative

2016-04-26 482744238810159000 Herbert -3 Negative

2015-12-15 111442643526292000 Daniel -3 Negative

2016-07-24 240301354138822000 Anna -3 Negative

2011-09-10 674549154086927000 Kevin -3 Negative

2016-07-24 169967281681239000 Nathan -3 Negative

2015-12-22 865974909537341000 Larry -3 Negative

2016-02-20 368093268772857000 Edwin -3 Negative

2015-12-22 797795108269529000 Elsie -3 Negative

2011-09-10 258076332672408000 Martha -3 Negative

2016-07-11 469723244877641000 Gregory -3 Negative

2011-09-10 674028217814346000 Colleen -3 Negative

2011-09-10 265999352650190000 Michael -3 Negative

2016-07-11 218896750493188000 Larry -3 Negative

2016-01-29 400719186005870000 Michael -3 Negative

2016-07-29 641766061380228000 Ethel -3 Negative

2016-07-11 166009066168663000 Marshall -3 Negative

2011-09-10 305634812075752000 Hugh -3 Negative

2015-12-24 758318303359839000 Wesley -3 Negative

2016-01-19 145475220714925000 Michelle -3 Negative

2016-03-20 254381223127613000 Charlie -3 Negative

2015-12-15 850113463907709000 Lucy -3 Negative

2015-12-24 144454247826644000 Seth -3 Negative

2016-06-07 880050504499670000 Keith -3 Negative

2015-12-22 128919251286755000 Dianne -3 Negative

2011-09-10 540253378244347000 Terry -3 Negative

2016-06-07 398182394732417000 Harvey -3 Negative

2016-06-07 100696141719273000 Michelle -3 Negative

2011-09-10 646002874918721000 Lester -4 Negative

2015-12-24 310151220166586000 Arnold -4 Negative

2011-09-10 634493757152332000 Nancy -4 Negative

2011-09-10 732861622677715000 Kara -4 Negative

2011-09-10 609724223609517000 Dwight -4 Negative

2011-09-10 106128175972515000 Nicholas -4 Negative

2015-12-22 883849857287617000 Harriet -4 Negative

2016-04-04 418627988413317000 Lauren -4 Negative

2011-09-10 502938309621066000 Joan -4 Negative

2015-12-18 892812529934080000 Calvin -4 Negative

2016-01-19 892786063807571000 Catherine -4 Negative

2016-05-24 766408822035634000 Gail -4 Negative

2011-09-10 413199933923568000 Lloyd -4 Negative

2011-09-10 389720483502088000 Bryan -4 Negative

2011-09-10 386990559936906000 Arthur -4 Negative

2011-09-10 376011600443087000 Peggy -4 Negative

2011-09-10 368316954944439000 Steven -4 Negative

2009-03-18 124202382045715000 Lester -4 Negative

2011-09-10 319417215505289000 Eugene -4 Negative

2011-09-10 310789874224621000 Vickie -4 Negative

2011-09-10 876852986037848000 Douglas -4 Negative

2016-07-11 718699544282651000 Jason -4 Negative

2011-09-10 872728387032954000 Glenda -4 Negative

2011-09-10 666601387892801000 Anne -4 Negative

2016-03-20 252561219999538000 Stanley -4 Negative

2015-12-18 344000670375452000 Rose -5 Negative

2016-01-25 269099495447532000 Shirley -5 Negative

2011-09-10 610467338785552000 Maria -5 Negative

2011-09-10 202312634011898000 Brandon -5 Negative

2016-07-29 872283722010901000 Barry -5 Negative

2011-09-10 876557870030974000 Theodore -5 Negative

2011-09-10 274451911902931000 Martin -5 Negative

2016-05-21 647253836562838000 Nancy -5 Negative

2011-09-10 534806540605840000 Tina -5 Negative

2016-07-11 789603530645178000 Jack -5 Negative

2009-03-18 480237525533163000 Molly -5 Negative

2016-07-11 292186470720315000 Marion -6 Negative

2011-09-10 156381402331895000 Lori -6 Negative

2016-05-24 624745045272903000 Joe -6 Negative

2015-12-18 472723630925884000 Stephen -6 Negative

2016-04-04 577677076504199000 Billie -6 Negative

2016-05-21 557148105287787000 Bonnie -6 Negative

2016-07-08 656055777924793000 Oscar -6 Negative

2016-01-19 171928727133382000 Peter -6 Negative

2016-04-04 793798392926237000 Maxine -6 Negative

2016-07-24 413165938863990000 Ashley -6 Negative

2011-09-10 486056666857108000 Vivian -6 Negative

2016-06-07 243553305572198000 Michele -6 Negative

2011-09-10 515550406704664000 Kathryn -7 Negative

2011-09-10 826843330205383000 Francis -7 Negative

2011-09-10 305449626035206000 Maria -7 Negative

2011-09-10 733078039079318000 Christian -8 Negative

2011-09-10 509056219671459000 Arthur -8 Negative

2015-12-15 561923667941253000 Gretchen -8 Negative

2015-12-15 456426128959809000 Josephine -8 Negative

2016-07-24 629648096701692000 Theodore -8 Negative

2011-09-10 892993179592271000 Edgar -8 Negative

2011-09-10 766411207843757000 Jenny -8 Negative

2015-12-24 694036553348883000 Jason -8 Negative

2011-09-10 337960517625769000 Wesley -9 Negative

2009-03-18 676634321788141000 Rita -9 Negative

2016-05-24 812783359834413000 Pam -9 Negative

2016-01-29 839041766946424000 Jonathan -9 Negative

2016-01-19 437885755599374000 Alfred -10 Negative

2016-07-11 574007977344591000 Karen NULL No Sentiments

2011-09-10 838523965774831000 Adam NULL No Sentiments

2016-05-21 817738691153533000 Chad NULL No Sentiments

2011-09-10 132658794295731000 Valerie NULL No Sentiments

2016-02-20 350763384924572000 Luis NULL No Sentiments

2016-04-04 713461537262843000 Renee NULL No Sentiments

2011-09-10 635458970282794000 James NULL No Sentiments

2011-09-10 427982246458003000 Suzanne NULL No Sentiments

2011-09-10 523670627452799000 Constance NULL No Sentiments

2016-01-25 211014694431747000 Craig NULL No Sentiments

2011-09-10 527473370935382000 Sean NULL No Sentiments

2016-01-19 511319795634024000 Ted NULL No Sentiments

2016-07-29 568832906644516000 Karen NULL No Sentiments

2011-09-10 680152893618731000 Gloria NULL No Sentiments

2011-09-10 412335317705604000 Harold NULL No Sentiments

2009-03-18 781654794440300000 Max NULL No Sentiments

2011-09-10 744458799775411000 Laura NULL No Sentiments

2011-09-10 394082755444620000 Lewis NULL No Sentiments

2016-02-20 547258850791958000 Kevin NULL No Sentiments

2011-09-10 107921004221536000 Gene NULL No Sentiments

2016-03-20 476210282051743000 Laura NULL No Sentiments

2009-03-18 313967997302001000 Sandra NULL No Sentiments

2011-09-10 364642343822423000 Kay NULL No Sentiments

2016-01-19 497742885879118000 Stanley NULL No Sentiments

2016-06-07 681712219462538000 Marcia NULL No Sentiments

2016-04-04 545555361130000000 Gina NULL No Sentiments

2016-07-08 329490439231855000 Leigh NULL No Sentiments

2011-09-10 553485622362101000 Lynda NULL No Sentiments

2016-07-08 477260705971257000 Ruby NULL No Sentiments

2011-09-10 328579674622915000 Glenda NULL No Sentiments

2016-07-24 505380376161911000 Max NULL No Sentiments

2011-09-10 561697330497060000 Joseph NULL No Sentiments

2011-09-10 648137397565993000 Bruce NULL No Sentiments

2015-12-22 184375879987218000 Renee NULL No Sentiments

2016-07-29 591002608868294000 Anne NULL No Sentiments

2011-09-10 666298474307054000 Ronald NULL No Sentiments

2011-09-10 569781884278506000 Earl NULL No Sentiments

2011-09-10 313233427544621000 Gloria NULL No Sentiments

2015-12-22 276613818614046000 Joe NULL No Sentiments

2016-01-25 378003117741832000 Sean NULL No Sentiments

2011-09-10 581496238710310000 Shirley NULL No Sentiments

2009-03-18 712451581331815000 Tracey NULL No Sentiments

2011-09-10 586132851463130000 Charlie NULL No Sentiments

2016-07-11 207877252896344000 Terri NULL No Sentiments

2011-09-10 837874520421193000 Brad NULL No Sentiments

2016-07-11 228983761537704000 Frederick NULL No Sentiments

2011-09-10 275628796345959000 Douglas NULL No Sentiments

2011-09-10 274259281179532000 Christopher NULL No Sentiments

2011-09-10 134237176213536000 Jonathan NULL No Sentiments

2016-04-04 128176348280256000 Johnny NULL No Sentiments

2016-07-11 445255906427603000 Cameron NULL No Sentiments

2016-03-20 689669757146043000 Leo NULL No Sentiments

2016-07-11 577008061158285000 Lorraine NULL No Sentiments

2016-03-20 672626544417876000 Leah NULL No Sentiments

2016-07-11 626133151613473000 Donald NULL No Sentiments

2011-09-10 256149922478944000 Arnold NULL No Sentiments

2016-07-11 632587007832120000 June NULL No Sentiments

2011-09-10 254612845851866000 Yvonne NULL No Sentiments

2011-09-10 825606273268040000 Virginia NULL No Sentiments

2016-07-11 714956649901555000 Bradley NULL No Sentiments

2011-09-10 231902135998564000 Gina NULL No Sentiments

2011-09-10 843783996731268000 Donna NULL No Sentiments

2016-07-11 717799208125090000 Melanie NULL No Sentiments

2015-12-15 513137013318944000 Jennifer NULL No Sentiments

2011-09-10 108397478405756000 Rhonda NULL No Sentiments

2016-07-11 836741320762824000 Marvin NULL No Sentiments

2016-07-11 837156668028995000 Ruby NULL No Sentiments

2011-09-10 189602173921151000 Ben NULL No Sentiments

2009-03-18 550058458689286000 Kristen NULL No Sentiments

2016-08-23 268653953230593000 George NULL No Sentiments

2011-09-10 849509668772053000 Clifford NULL No Sentiments

2011-09-10 623294065756896000 Beth NULL No Sentiments

2016-07-24 371137043010186000 Justin NULL No Sentiments

2016-03-20 605793091233204000 Paige NULL No Sentiments

2009-03-18 375611165970496000 Diane NULL No Sentiments

2011-09-10 869789319080630000 Andrea NULL No Sentiments

2011-09-10 719247173706153000 Louis NULL No Sentiments

2011-09-10 719244398343400000 Cameron NULL No Sentiments

2015-12-18 583475249539407000 Susan NULL No Sentiments

2011-09-10 111565792997795000 Amy NULL No Sentiments

2011-09-10 857025223114037000 Valerie NULL No Sentiments

2011-09-10 710314746912019000 Jimmy NULL No Sentiments

2016-01-29 195582399958342000 Peter NULL No Sentiments

2011-09-10 632464301428428000 Maureen NULL No Sentiments

2011-09-10 706072806488439000 Rachel NULL No Sentiments

2016-01-29 835551609756185000 Penny NULL No Sentiments

2016-07-24 565047664539432000 Emma NULL No Sentiments

2011-09-10 719287043854373000 Julian NULL No Sentiments

2016-02-20 171071624081952000 Seth NULL No Sentiments

2011-09-10 724534260864285000 Ken NULL No Sentiments

2011-09-10 861739639947710000 Lori NULL No Sentiments

2011-09-10 498785734603842000 Tony NULL No Sentiments

2011-09-10 497904421215766000 Oscar NULL No Sentiments

2016-01-25 178185921882025000 Carlos NULL No Sentiments

2015-12-18 766254378098184000 Dana NULL No Sentiments

2016-04-26 202192313966323000 Carol NULL No Sentiments

2016-04-26 216297210501385000 Herman NULL No Sentiments

2011-09-10 780403501790242000 Sandra NULL No Sentiments

2016-04-26 775172105599004000 Rodney NULL No Sentiments

2011-09-10 785450721233129000 Karen NULL No Sentiments

2011-09-10 789826124283228000 Robyn NULL No Sentiments

2016-04-04 728963875740904000 Carlos NULL No Sentiments

1. **Did you find any problem in the previous question? If so briefly explain what exactly the problem is and what solution would you propose.**

**Note: You need not solve the proposed solution with the twitter file.**

Use of Serde: We haven’t use serde in our analysis above.

however, it is very powerful tool which can be used by hive to deal with complex data structure and we can also eliminate some of the insert statement steps in our examplWhat is a SerDe?

The SerDe interface allows you to instruct Hive as to how a record should be processed. A SerDe is a combination of a Serializer and a Deserializer (hence, Ser-De). The [Deserializer](http://javasourcecode.org/html/open-source/hive/hive-0.9.0/org/apache/hadoop/hive/serde2/Deserializer.html) interface takes a string or binary representation of a record, and translates it into a Java object that Hive can manipulate. The [Serializer](http://javasourcecode.org/html/open-source/hive/hive-0.9.0/org/apache/hadoop/hive/serde2/Serializer.html), however, will take a Java object that Hive has been working with, and turn it into something that Hive can write to HDFS or another supported system. Commonly, Deserializers are used at query time to execute SELECTstatements, and Serializers are used when writing data, such as through an INSERT-SELECT statement.

For that we have to use following Jar file in the Hive.



This will generate a file called hive-serdes-1.0-SNAPSHOT.jar in the target directory

Then we can create table like following query along with the data directly from our raw table json\_table\_2. (without need for the inserting the data in the main table)

CREATE EXTERNAL TABLE json\_assign7(

retweet\_count int,

created\_at timestamp,

text string,

id bigint,

source string,

in\_reply\_to\_screen\_name string,

user struct<location: string, id: bigint, id\_str: bigint, name:string, screen\_name: string, geo\_enabled: boolean, lang: string,protected: boolean, verified: boolean, followers\_count:int,friends\_count: int, listed\_count: int, favourites\_count:int,statuses\_count: int, profile\_background\_color: int>,

contributors string,

is\_quote\_status boolean,

entities struct <user\_mentions:ARRAY<struct <screen\_name : string,name : string, id : bigint >>>);

ROW FORMAT SERDE ‘com.cloudera.hive.serde.JSONSerDe’

LOCATION ‘/user/hive/warehouse/buan6346.db/json\_table\_2’;

And that is how we can eliminating the steps for the manually inserting the records in json\_assign7 main table.

2) Problems in sentiment analysis

* Sarcasm: We can not identify the sarcastic comments used into tweets.Someone may use good words to criticize the bad situation in good way, then our analysis will be worng.
* Words like “not good” or “not working” could be handled using this type of query. That problem will become much more complicated. Our code will separate not and good as both different words and then it will result into erroneous sentiment.

E.g Trump method of working is not good. 🡪 Sentiment score : 3 ( by our method)

Trump method of working is not good. 🡪 Sentiment score : -2( Originally)

* If words are become noun, adjective or adverb , it would not possible to find exact word for form.

Suppose , we have “need” as word in tweet, But we have “needy” as word in our dictinory file.In this case we can not compare both the words and derive the score for it.