--MetaData

SELECT NAME, CREATOR, CTIME, COLCOUNT

FROM SYSIBM.SYSTABLES

WHERE CREATOR ='CZH52602'

AND NAME='DEFLATEGATE';

SELECT NAME, TBNAME, TBCREATOR, COLTYPE, NULLS, LENGTH

FROM SYSIBM.SYSCOLUMNS

WHERE TBCREATOR='CZH52602'

AND TBNAME='DEFLATEGATE';

SELECT COUNT(\*) num\_rows

FROM Deflategate;

--check for missing data

Select Deflate\_Sentiment, count(Text) Count

from deflategate

where deflate\_sentiment != 'positive'

and deflate\_sentiment != 'negative'

and deflate\_sentiment != 'neutral'

and deflate\_sentiment != 'slightly positive'

and deflate\_sentiment != 'slightly negative'

group by deflate\_sentiment;

--number of post per sentiment

Select Distinct(Deflate\_Sentiment), count(TEXT) Count

From Deflategate

Group by Deflate\_Sentiment;

Select

sum(case when deflate\_sentiment = 'positive' then 1 else 0 end) positive,

sum(case when deflate\_sentiment = 'negative' then 1 else 0 end) negative,

sum( case when deflate\_sentiment = 'neutral' then 1 else 0 end) neutral,

sum( case when deflate\_sentiment = 'slightly positive' then 1 else 0 end) slightly\_positive,

sum( case when deflate\_sentiment = 'slightly negative'

or deflate\_sentiment = 'slighly negative' then 1 else 0 end) slightly\_negative,

Count(Deflate\_sentiment) Total\_Counts

From Deflategate;

--number of post per location

select \* from (

select user\_timezone, count(text) Count

from deflategate

group by user\_timezone

having length(user\_timezone) >1

order by Count Desc)

where count > 40;

select \* from (

select tweet\_location, count(text) Count

from deflategate

group by tweet\_location

having length(tweet\_location) > 1

order by Count Desc)

where count > 40;

select \* from (

select tweet\_location, count(text) Count

from deflategate

group by tweet\_location

having length(tweet\_location)>1

order by Count Desc)

where count > 40;

-- 1. Sentiment by location

select tweet\_location,

sum(case when deflate\_sentiment = 'positive'

or deflate\_sentiment = 'slightly positive' then 1 else 0 end) positive,

sum(case when deflate\_sentiment = 'negative'

or deflate\_sentiment = 'slighly negative' then 1 else 0 end) negative,

sum(case when deflate\_sentiment ='neutral' then 1 else 0 end) neutral,

count(text) Count

from deflategate

group by tweet\_location

having LENGTH(tweet\_location) > 1

order by Count Desc;

select user\_timezone,

sum(case when deflate\_sentiment = 'positive'

or deflate\_sentiment = 'slightly positive' then 1 else 0 end) positive,

sum(case when deflate\_sentiment = 'negative'

or deflate\_sentiment = 'slighly negative' then 1 else 0 end) negative,

sum(case when deflate\_sentiment ='neutral' then 1 else 0 end) neutral,

count(text) Count

from deflategate

group by user\_timezone

having LENGTH(user\_timezone) > 1

order by Count Desc;

-- 2. Did he do it.

select deflategate\_yn, count(deflategate\_yn) Count

from deflategate

group by deflategate\_yn

order by Count Desc;

select tweet\_location,

sum(case when deflategate\_yn= 'yes' then 1 else 0 end) Yes,

sum(case when deflategate\_yn ='no' then 1 else 0 end) No,

count(deflategate\_yn) Count

from deflategate

group by tweet\_location

having LENGTH(tweet\_location) > 1

order by Count Desc;

select user\_timezone,

sum(case when deflategate\_yn= 'yes' then 1 else 0 end) Yes,

sum(case when deflategate\_yn ='no' then 1 else 0 end) No,

count(deflategate\_yn) Count

from deflategate

group by user\_timezone

having LENGTH(user\_timezone) > 1

order by Count Desc;

--3.

--What are the people saying:

-- Top tweeters

Select Name, count(Text) Count

From deflategate

group by Name

order by Count DESC;

Select \* from (

Select Name, regexp\_replace(text,'[^a-zA-Z\d\s]','')

From deflategate

where name = 'BreathingMemes'

limit 5)

union

select \* from (

Select Name, regexp\_replace(text,'[^a-zA-Z\d\s]','')

From deflategate

where name = 'politicalsculpt'

limit 5)

order by name;

--top retweets... jokes

Select Name, regexp\_replace(text,'[^a-zA-Z\d\s]',''), RETWEET\_COUNT

from deflategate

where length(trim(text)) > 1

order by retweet\_count desc;

--4. Tweets with tom brady

Select Name, regexp\_replace(text,'[^a-zA-Z\d\s]',''), retweet\_count

from deflategate

where length(trim(text)) > 1

and text like '%brady%'

or text like '%tom%'

order by retweet\_count desc;