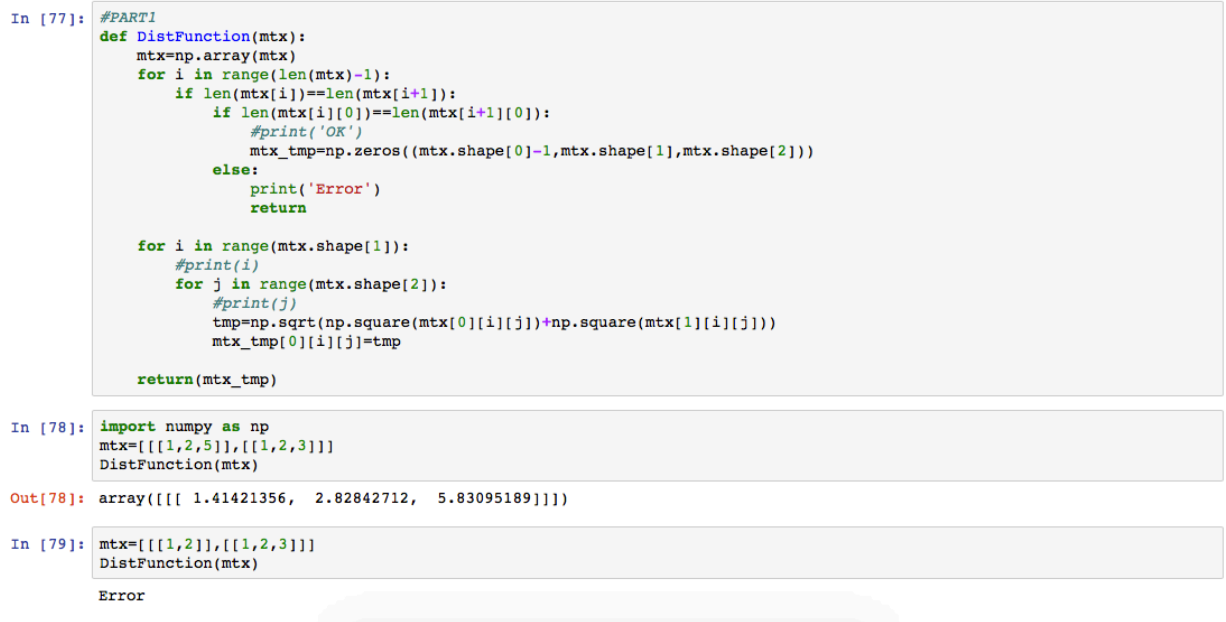
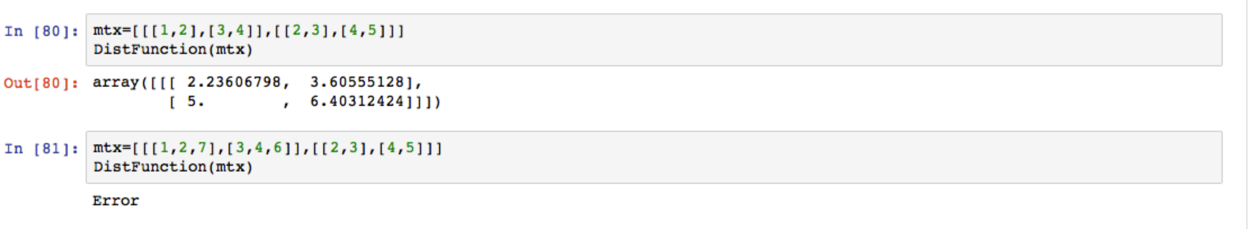
CSC495 Assignment#4

Email: kying@mail.depaul.edu
Date: 05-19-2017

Part-1: Write a function to compute Euclidian Distance between each individual value in each matrix.

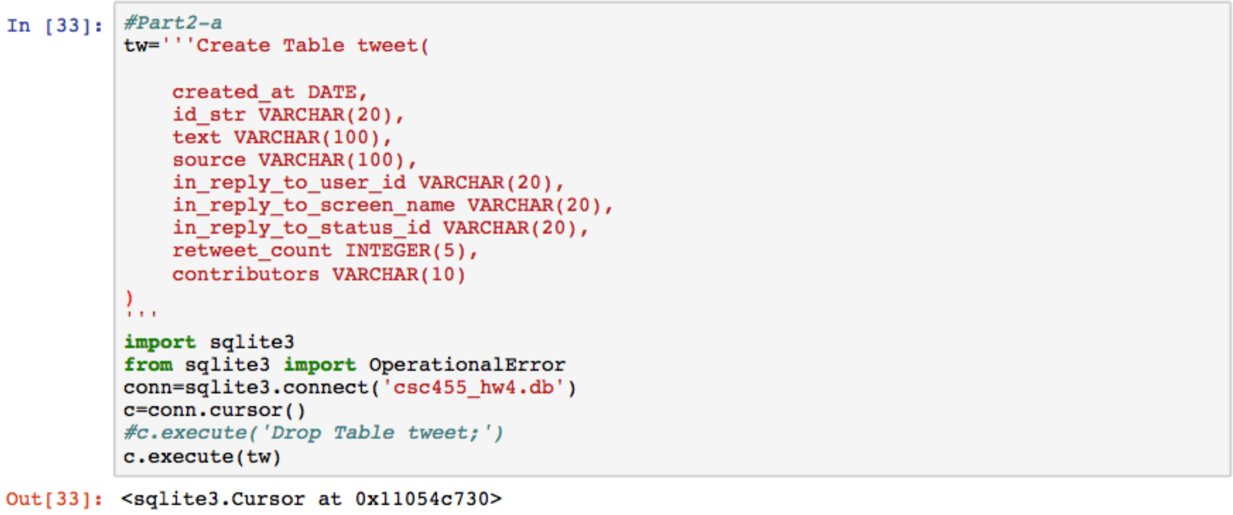


```
In [77]: #PART1
         def DistFunction(mtx):
             mtx=np.array(mtx)
             for i in range(len(mtx)-1):
                 if len(mtx[i])==len(mtx[i+1]):
                     if len(mtx[i][0])==len(mtx[i+1][0]):
                         #print('OK')
                         mtx_tmp=np.zeros((mtx.shape[0]-1,mtx.shape[1],mtx.shape[2]))
                     else:
                         print('Error')
                         return

             for i in range(mtx.shape[1]):
                 #print(i)
                 for j in range(mtx.shape[2]):
                     #print(j)
                     tmp=np.sqrt(np.square(mtx[0][i][j])+np.square(mtx[1][i][j]))
                     mtx_tmp[0][i][j]=tmp

             return(mtx_tmp)

In [78]: import numpy as np
         mtx=[[[1,2,5]],[[1,2,3]]]
         DistFunction(mtx)

Out[78]: array([[[ 1.41421356,  2.82842712,  5.83095189]]])

In [79]: mtx=[[[1,2]],[[1,2,3]]]
         DistFunction(mtx)

         Error
```



```
In [80]: mtx=[[[1,2],[3,4]],[[2,3],[4,5]]]
         DistFunction(mtx)

Out[80]: array([[[ 2.23606798,  3.60555128],
                 [ 5.        ,  6.40312424]]])

In [81]: mtx=[[[1,2,7],[3,4,6]],[[2,3],[4,5]]]
         DistFunction(mtx)

         Error
```

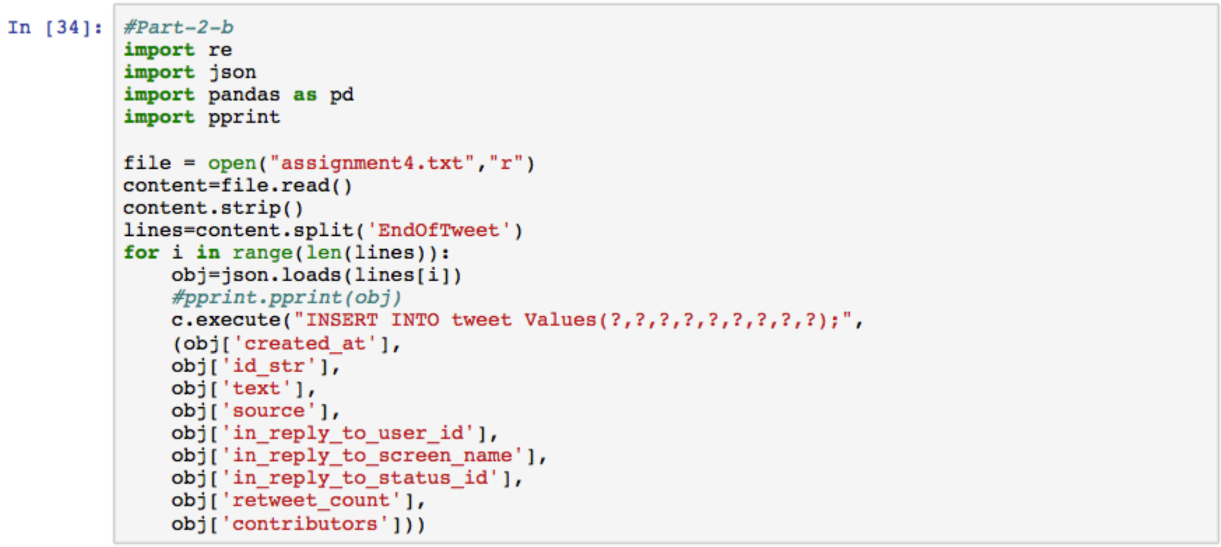
Part2-a: Create Table (9 Attributes) and Load into sqlite



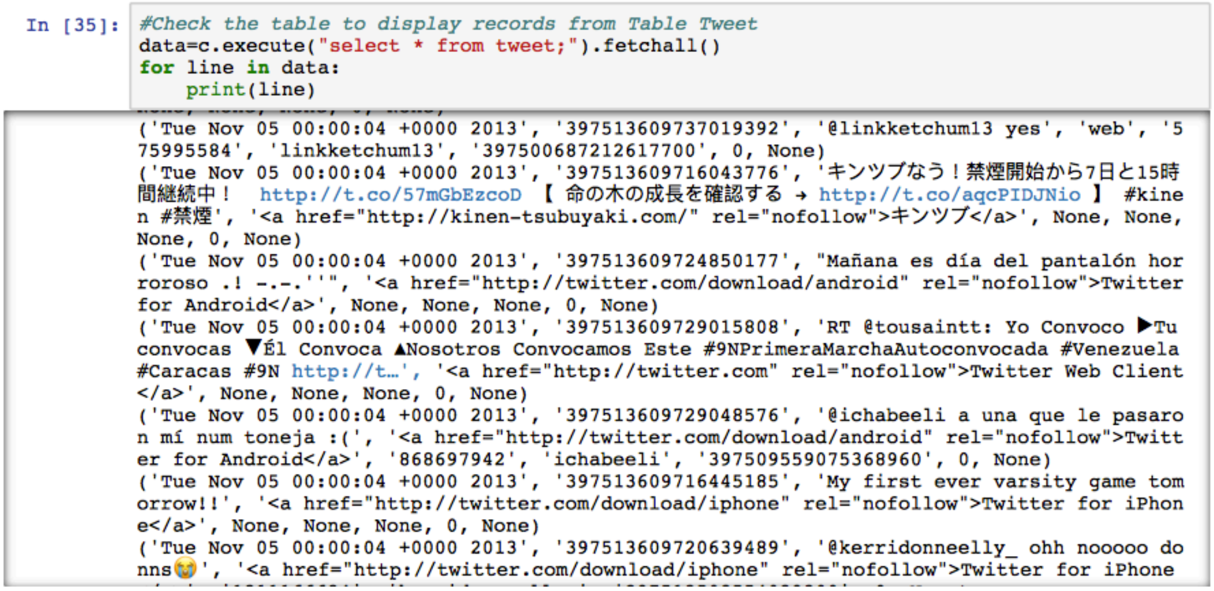
```
In [33]: #Part2-a
         tw='''Create Table tweet(

             created_at DATE,
             id_str VARCHAR(20),
             text VARCHAR(100),
             source VARCHAR(100),
             in_reply_to_user_id VARCHAR(20),
             in_reply_to_screen_name VARCHAR(20),
             in_reply_to_status_id VARCHAR(20),
             retweet_count INTEGER(5),
             contributors VARCHAR(10)
         )
         '''
         import sqlite3
         from sqlite3 import OperationalError
         conn=sqlite3.connect('csc455_hw4.db')
         c=conn.cursor()
         #c.execute('Drop Table tweet;')
         c.execute(tw)

Out[33]: <sqlite3.Cursor at 0x11054c730>
```

Date: 05-19-2017

Part2-b: Write python code to read through the Assignment4.txt file and populate table from part2-a including NULLs (i.e. None)



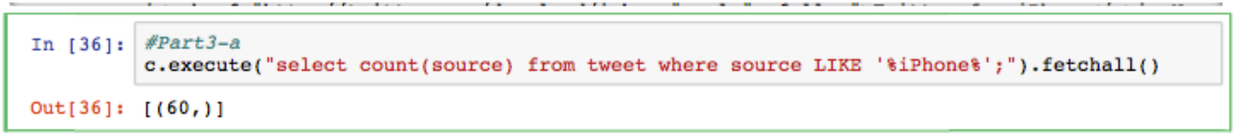
```
In [34]: #Part-2-b
         import re
         import json
         import pandas as pd
         import pprint

         file = open("assignment4.txt","r")
         content=file.read()
         content.strip()
         lines=content.split('EndOfTweet')
         for i in range(len(lines)):
             obj=json.loads(lines[i])
             #pprint.pprint(obj)
             c.execute("INSERT INTO tweet Values(?,?,?,?,?,?,?,?,?);",
             (obj['created_at'],
             obj['id_str'],
             obj['text'],
             obj['source'],
             obj['in_reply_to_user_id'],
             obj['in_reply_to_screen_name'],
             obj['in_reply_to_status_id'],
             obj['retweet_count'],
             obj['contributors']))
```



```
In [35]: #Check the table to display records from Table Tweet
         data=c.execute("select * from tweet;").fetchall()
         for line in data:
             print(line)
```

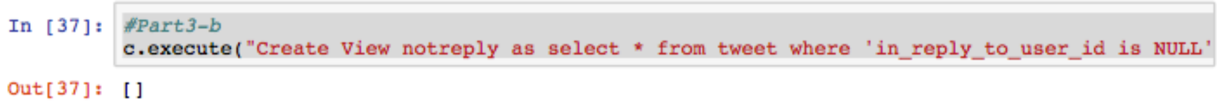
```
('Tue Nov 05 00:00:04 +0000 2013', '397513609737019392', '@linkketchum13 yes', 'web', '5
75995584', 'linkketchum13', '397500687212617700', 0, None)
('Tue Nov 05 00:00:04 +0000 2013', '397513609716043776', 'キンツブなう！禁煙開始から7日と15時
間継続中！  http://t.co/57mGbEzcoD 【 命の木の成長を確認する → http://t.co/aqcPIDJNio 】 #kine
n #禁煙', '<a href="http://kinen-tsubuyaki.com/" rel="nofollow">キンツブ</a>', None, None,
None, 0, None)
('Tue Nov 05 00:00:04 +0000 2013', '397513609724850177', "Mañana es día del pantalón hor
roroso .! -.-.''", '<a href="http://twitter.com/download/android" rel="nofollow">Twitter
for Android</a>', None, None, None, 0, None)
('Tue Nov 05 00:00:04 +0000 2013', '397513609729015808', 'RT @tousaintt: Yo Convoco ▶Tu
convocas ▼Él Convoca ▲Nosotros Convocamos Este #9NPrimeraMarchaAutoconvocada #Venezuela
#Caracas #9N http://t…', '<a href="http://twitter.com" rel="nofollow">Twitter Web Client
</a>', None, None, None, 0, None)
('Tue Nov 05 00:00:04 +0000 2013', '397513609729048576', '@ichabeeli a una que le pasaro
n mí num toneja :(', '<a href="http://twitter.com/download/android" rel="nofollow">Twitt
er for Android</a>', '868697942', 'ichabeeli', '397509559075368960', 0, None)
('Tue Nov 05 00:00:04 +0000 2013', '397513609716445185', 'My first ever varsity game tom
orrow!!', '<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhon
e</a>', None, None, None, 0, None)
('Tue Nov 05 00:00:04 +0000 2013', '397513609720639489', '@kerridonneelly_ ohh nooooo do
nns😭', '<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone
```

#Part3-a: The table name ‘tweet’ is assigned on part2-a
c.execute("select count(source) from tweet where source LIKE '%iPhone%';").fetchall()



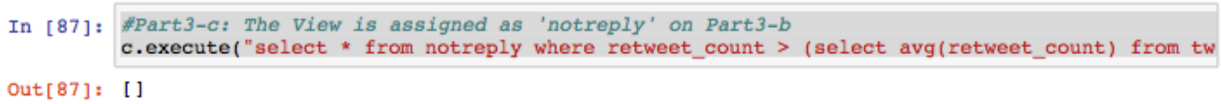
```
In [36]: #Part3-a
         c.execute("select count(source) from tweet where source LIKE '%iPhone%';").fetchall()

Out[36]: [(60,)]
```

#Part3-b
c.execute("Create View notreply as select * from tweet where 'in_reply_to_user_id is NULL'").fetchall()



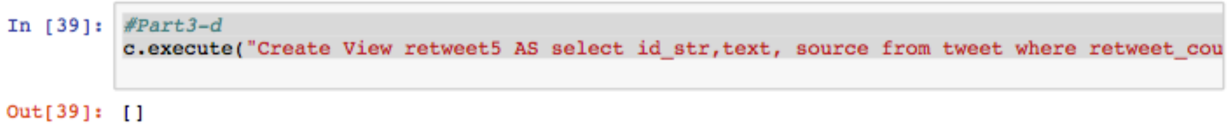
```
In [37]: #Part3-b
         c.execute("Create View notreply as select * from tweet where 'in_reply_to_user_id is NULL'

Out[37]: []
```

#Part3-c: The View is assigned as 'notreply' on Part3-b
c.execute("select * from notreply where retweet_count > (select avg(retweet_count) from tweet);").fetchall()



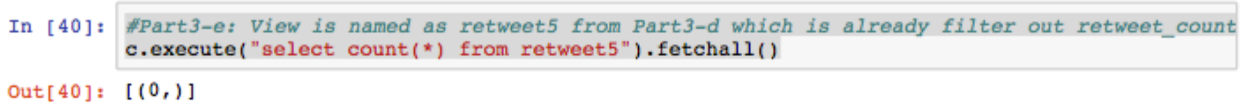
```
In [87]: #Part3-c: The View is assigned as 'notreply' on Part3-b
         c.execute("select * from notreply where retweet_count > (select avg(retweet_count) from tw

Out[87]: []
```

#Part3-d: The name of View is assigned as ‘retweet5’
c.execute("Create View retweet5 AS select id_str,text, source from tweet where retweet_count>=5").fetchall()



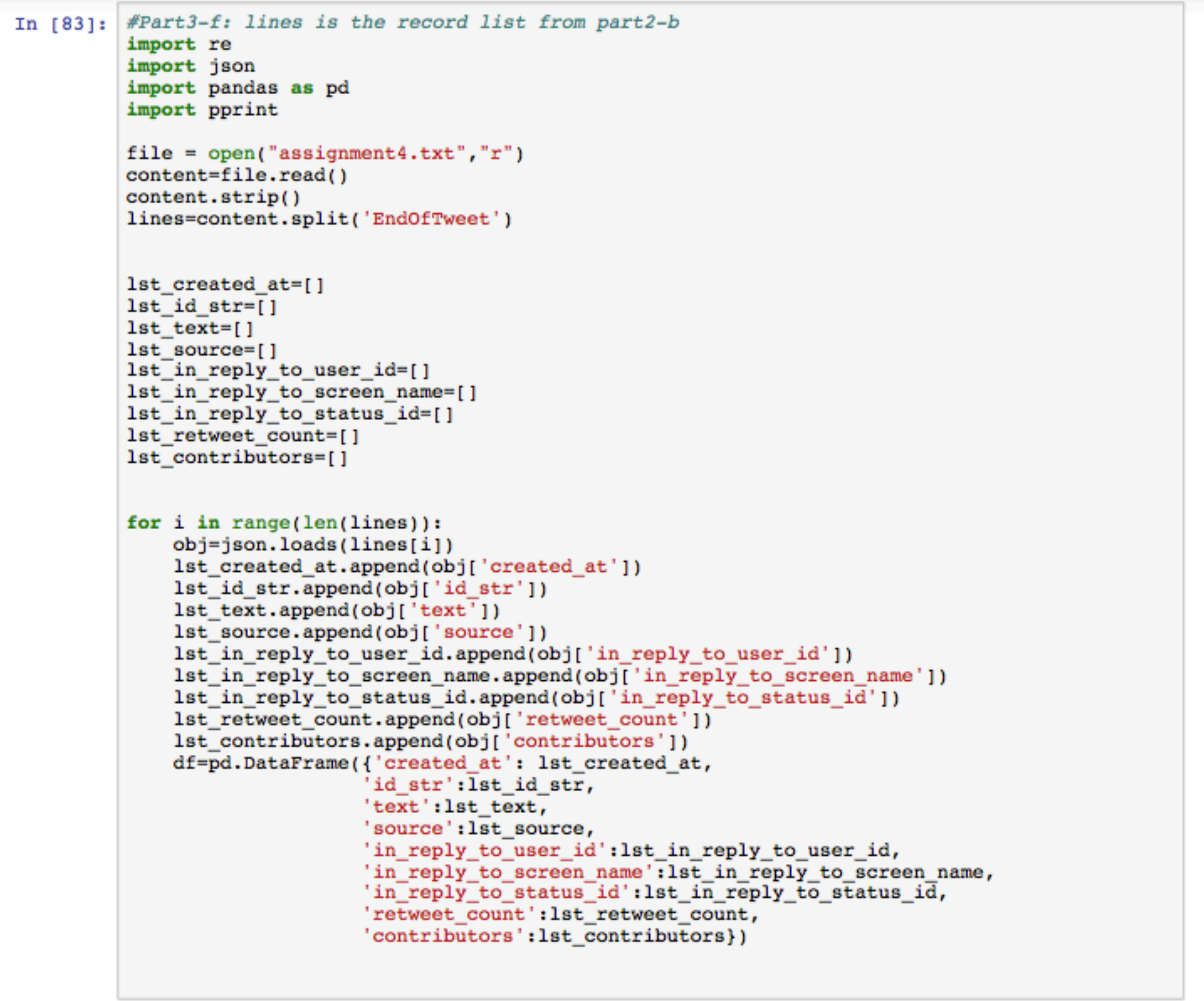
```
In [39]: #Part3-d
         c.execute("Create View retweet5 AS select id_str,text, source from tweet where retweet_cou

Out[39]: []
```

#Part3-e: View is named as retweet5 from Part3-d which is already filter out retweet_count>=5
c.execute("select count(*) from retweet5").fetchall()

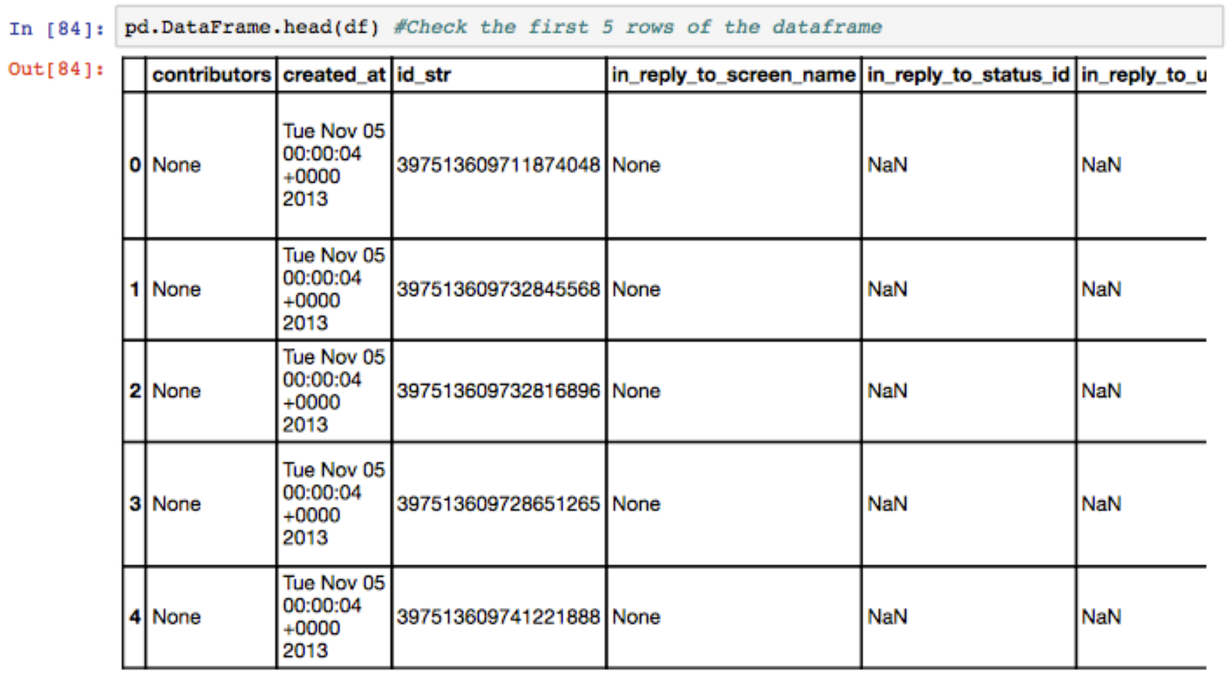


```
In [40]: #Part3-e: View is named as retweet5 from Part3-d which is already filter out retweet_count
         c.execute("select count(*) from retweet5").fetchall()

Out[40]: [(0,)]
```

Part3-f: Write Python script to find out the number of tweet with retweet_count>=5



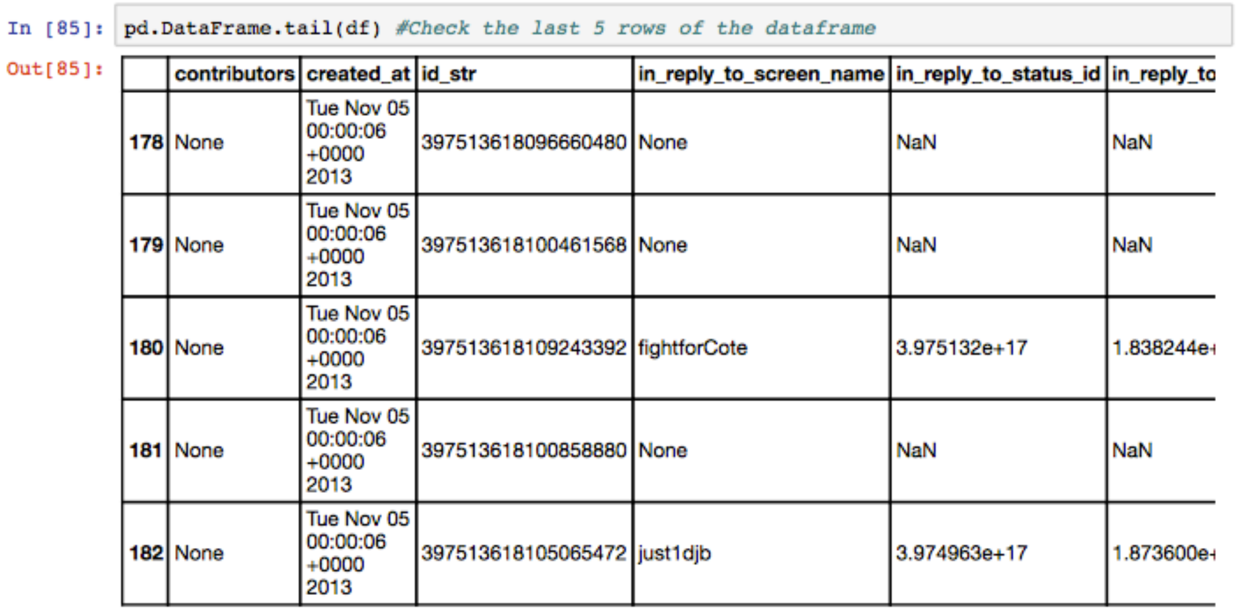
```
In [83]: #Part3-f: lines is the record list from part2-b
         import re
         import json
         import pandas as pd
         import pprint

         file = open("assignment4.txt","r")
         content=file.read()
         content.strip()
         lines=content.split('EndOfTweet')


         lst_created_at=[]
         lst_id_str=[]
         lst_text=[]
         lst_source=[]
         lst_in_reply_to_user_id=[]
         lst_in_reply_to_screen_name=[]
         lst_in_reply_to_status_id=[]
         lst_retweet_count=[]
         lst_contributors=[]


         for i in range(len(lines)):
             obj=json.loads(lines[i])
             lst_created_at.append(obj['created_at'])
             lst_id_str.append(obj['id_str'])
             lst_text.append(obj['text'])
             lst_source.append(obj['source'])
             lst_in_reply_to_user_id.append(obj['in_reply_to_user_id'])
             lst_in_reply_to_screen_name.append(obj['in_reply_to_screen_name'])
             lst_in_reply_to_status_id.append(obj['in_reply_to_status_id'])
             lst_retweet_count.append(obj['retweet_count'])
             lst_contributors.append(obj['contributors'])
             df=pd.DataFrame({'created_at': lst_created_at,
                             'id_str':lst_id_str,
                             'text':lst_text,
                             'source':lst_source,
                             'in_reply_to_user_id':lst_in_reply_to_user_id,
                             'in_reply_to_screen_name':lst_in_reply_to_screen_name,
                             'in_reply_to_status_id':lst_in_reply_to_status_id,
                             'retweet_count':lst_retweet_count,
                             'contributors':lst_contributors})
```



```
In [84]: pd.DataFrame.head(df) #Check the first 5 rows of the dataframe
```

Out[84]:

	contributors	created_at	id_str	in_reply_to_screen_name	in_reply_to_status_id	in_reply_to_u
0	None	Tue Nov 05 00:00:04 +0000 2013	397513609711874048	None	NaN	NaN
1	None	Tue Nov 05 00:00:04 +0000 2013	397513609732845568	None	NaN	NaN
2	None	Tue Nov 05 00:00:04 +0000 2013	397513609732816896	None	NaN	NaN
3	None	Tue Nov 05 00:00:04 +0000 2013	397513609728651265	None	NaN	NaN
4	None	Tue Nov 05 00:00:04 +0000 2013	397513609741221888	None	NaN	NaN

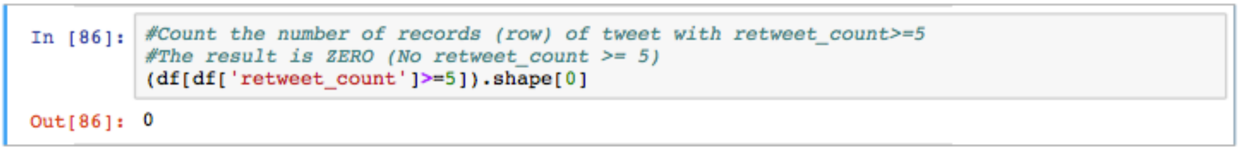


```
In [85]: pd.DataFrame.tail(df) #Check the last 5 rows of the dataframe
```

Out[85]:

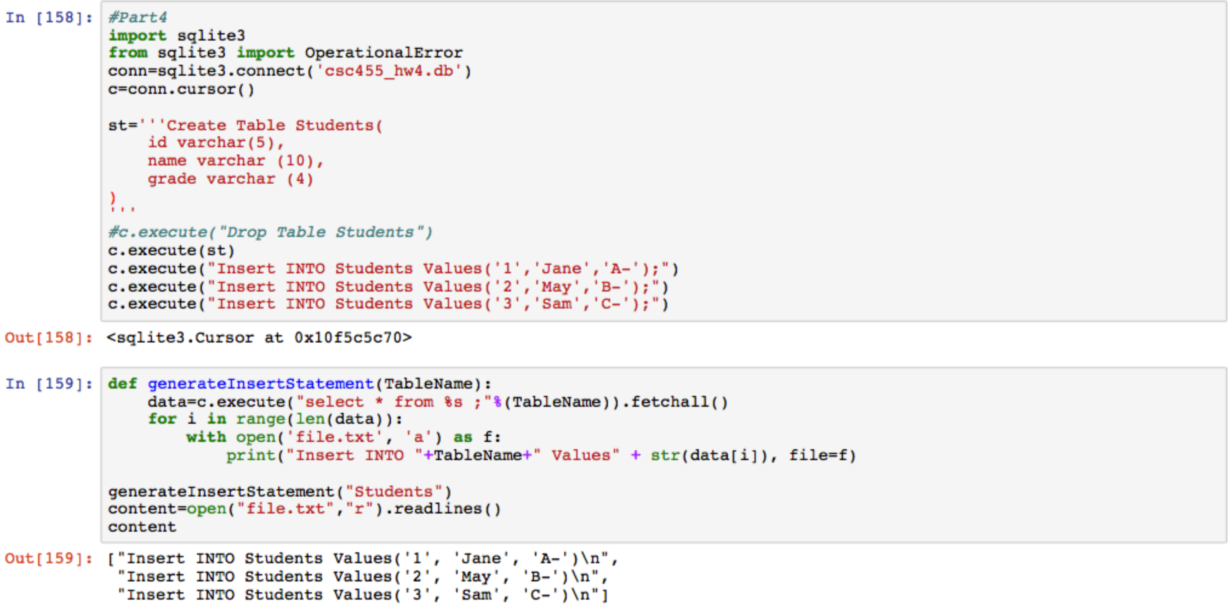
	contributors	created_at	id_str	in_reply_to_screen_name	in_reply_to_status_id	in_reply_to
178	None	Tue Nov 05 00:00:06 +0000 2013	397513618096660480	None	NaN	NaN
179	None	Tue Nov 05 00:00:06 +0000 2013	397513618100461568	None	NaN	NaN
180	None	Tue Nov 05 00:00:06 +0000 2013	397513618109243392	fightforCote	3.975132e+17	1.838244e+
181	None	Tue Nov 05 00:00:06 +0000 2013	397513618100858880	None	NaN	NaN
182	None	Tue Nov 05 00:00:06 +0000 2013	397513618105065472	just1djb	3.974963e+17	1.873600e+

#***The number of tweet_count >= 5 equals to ZERO



```
In [86]: #Count the number of records (row) of tweet with retweet_count>=5
         #The result is ZERO (No retweet_count >= 5)
         (df[df['retweet_count']>=5]).shape[0]

Out[86]: 0
```

Part-4: Write python function with Table Name as parameter to output INSERT statement to a file. In this case, I named the file as “file.txt”.



```
In [158]: #Part4
          import sqlite3
          from sqlite3 import OperationalError
          conn=sqlite3.connect('csc455_hw4.db')
          c=conn.cursor()

          st='''Create Table Students(
              id varchar(5),
              name varchar (10),
              grade varchar (4)
          )
          '''
          #c.execute("Drop Table Students")
          c.execute(st)
          c.execute("Insert INTO Students Values('1','Jane','A-');")
          c.execute("Insert INTO Students Values('2','May','B-');")
          c.execute("Insert INTO Students Values('3','Sam','C-');")

Out[158]: <sqlite3.Cursor at 0x10f5c5c70>

In [159]: def generateInsertStatement(TableName):
              data=c.execute("select * from %s ;"%(TableName)).fetchall()
              for i in range(len(data)):
                  with open('file.txt', 'a') as f:
                      print("Insert INTO "+TableName+" Values" + str(data[i]), file=f)

          generateInsertStatement("Students")
          content=open("file.txt","r").readlines()
          content

Out[159]: ["Insert INTO Students Values('1', 'Jane', 'A-')\n",
           "Insert INTO Students Values('2', 'May', 'B-')\n",
           "Insert INTO Students Values('3', 'Sam', 'C-')\n"]
```