Evaluating Community Building Effectiveness of Transportation Investments: Knowledge Transfer Webinar Series

Webinar II: Social Media Data Download and Analysis for Transportation Projects Part 2: Demo

Presented by
Samiul Hasan, Assistant Professor
Naveen Eluru, Professor
Jiechao Zhang, PhD Student

Civil, Environmental, and Construction Engineering
University of Central Florida

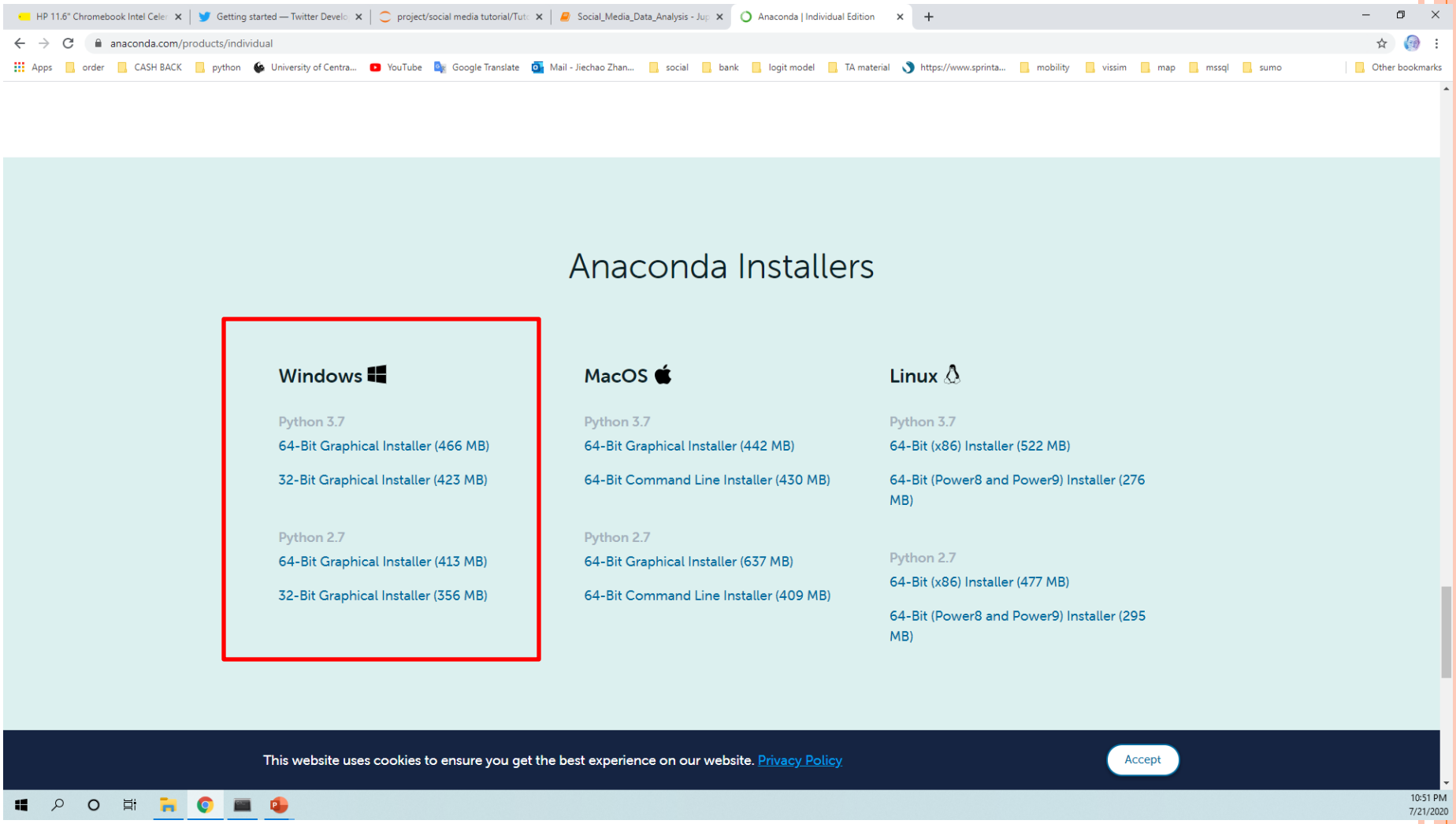
Outline

- Demo 1: Software installation and data downloading
- Demo 2: Sentiment Analysis and Visualization
- Demo 3: Topic Analysis and Visualization

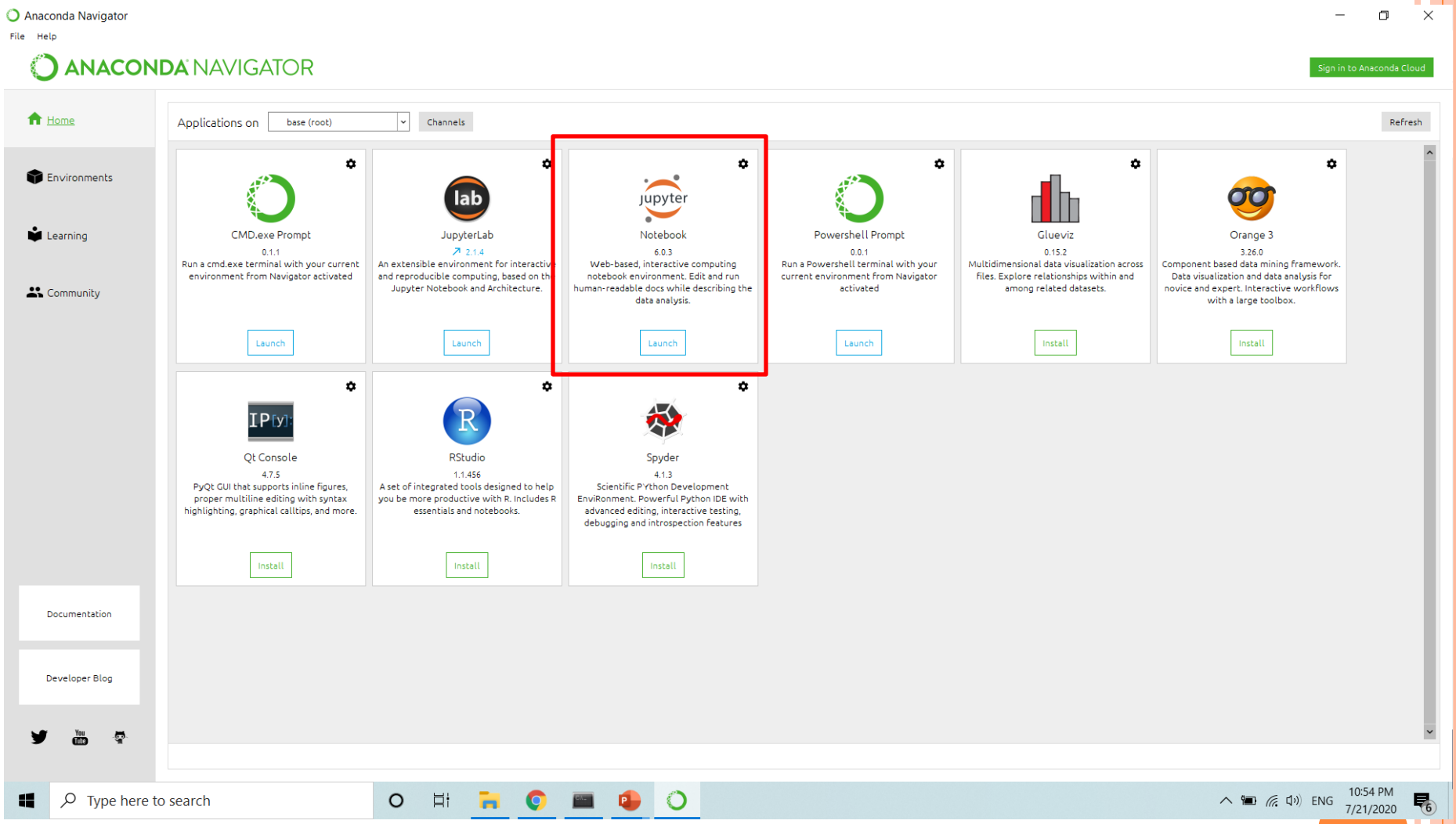
Demo 1:
Software Installation and Data Downloading

Data Collection – Download Anaconda

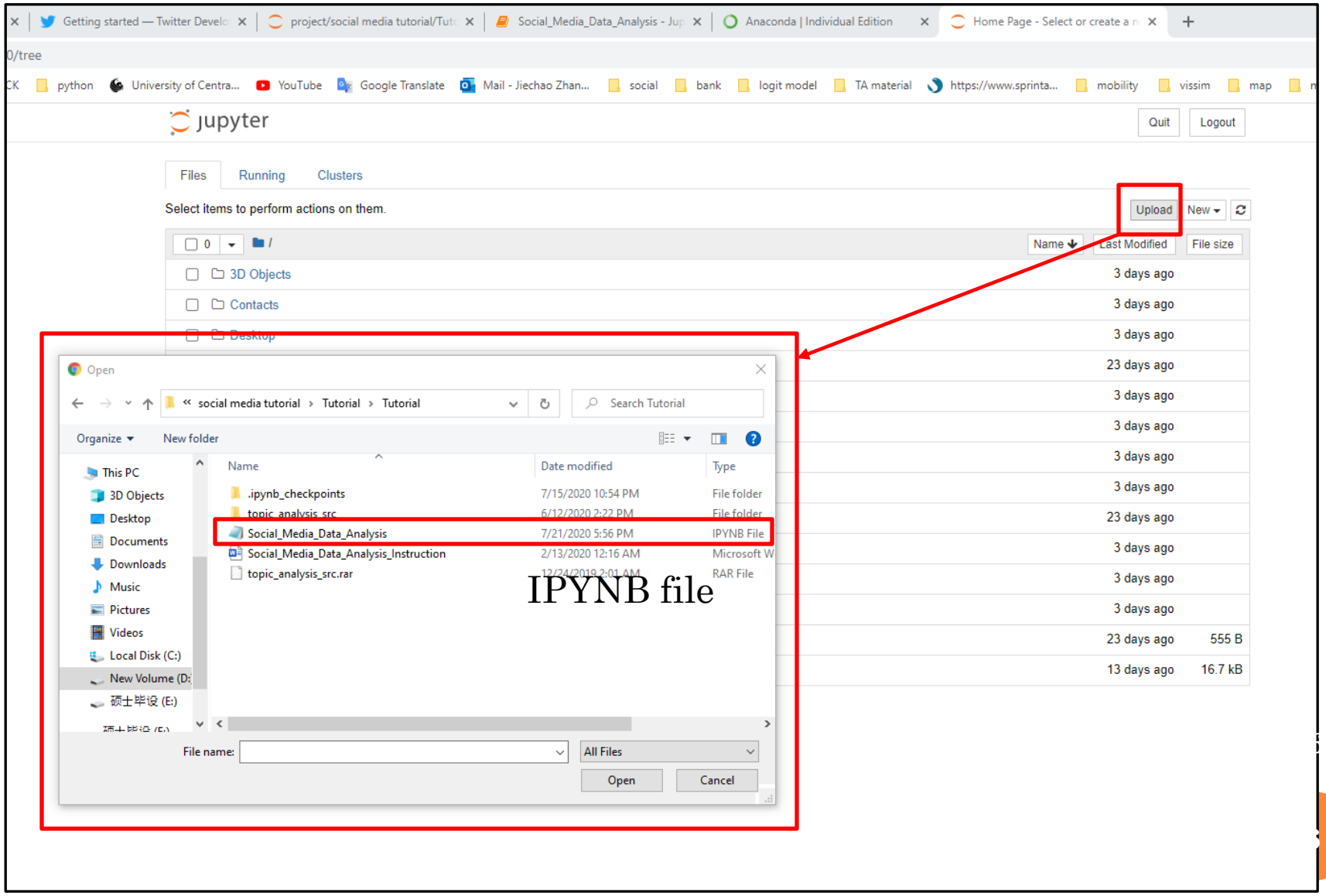
Download Link: https://www.anaconda.com/products/individual



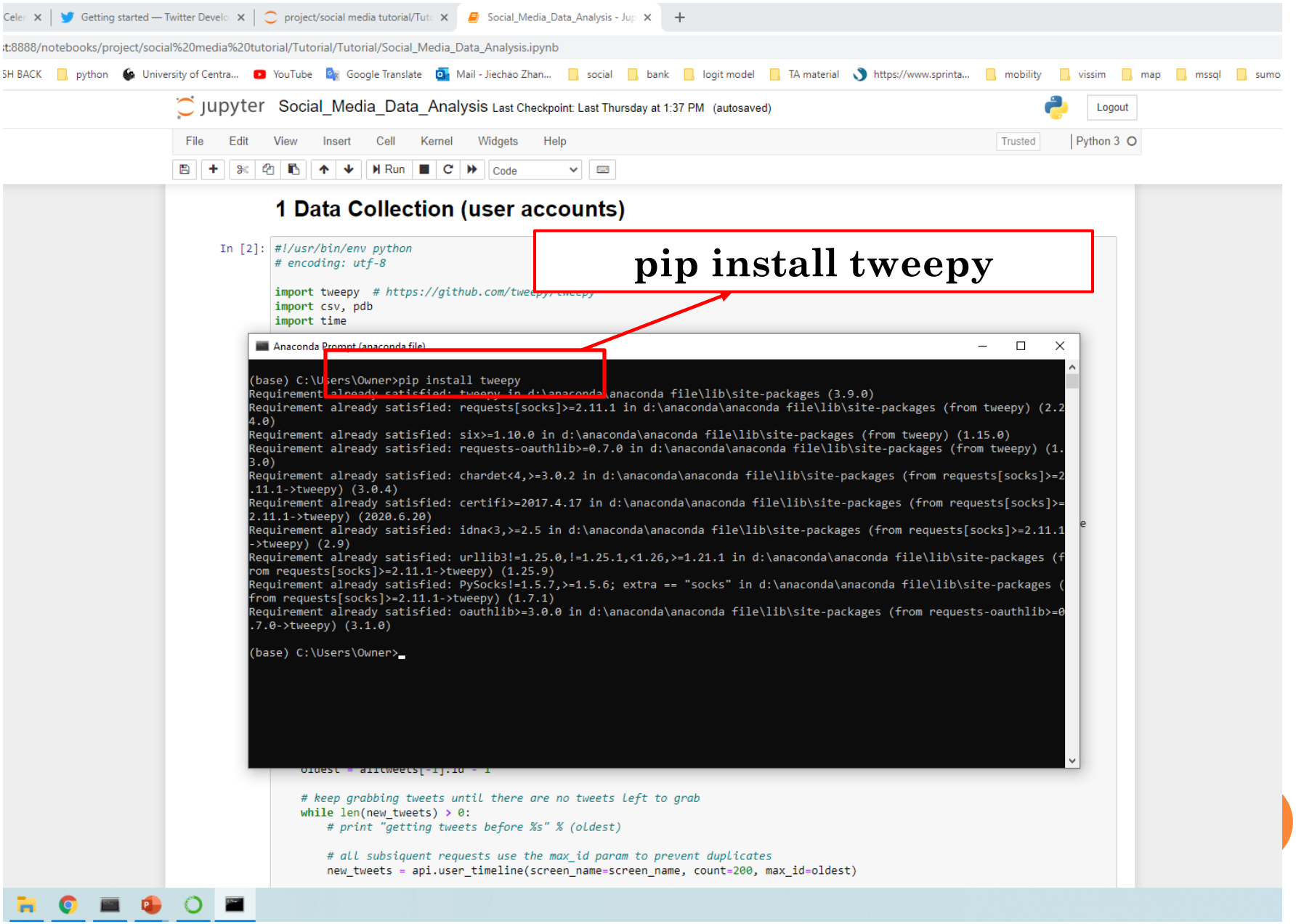
Data Collection – Open Anaconda Navigator



Data Collection – Upload Codes

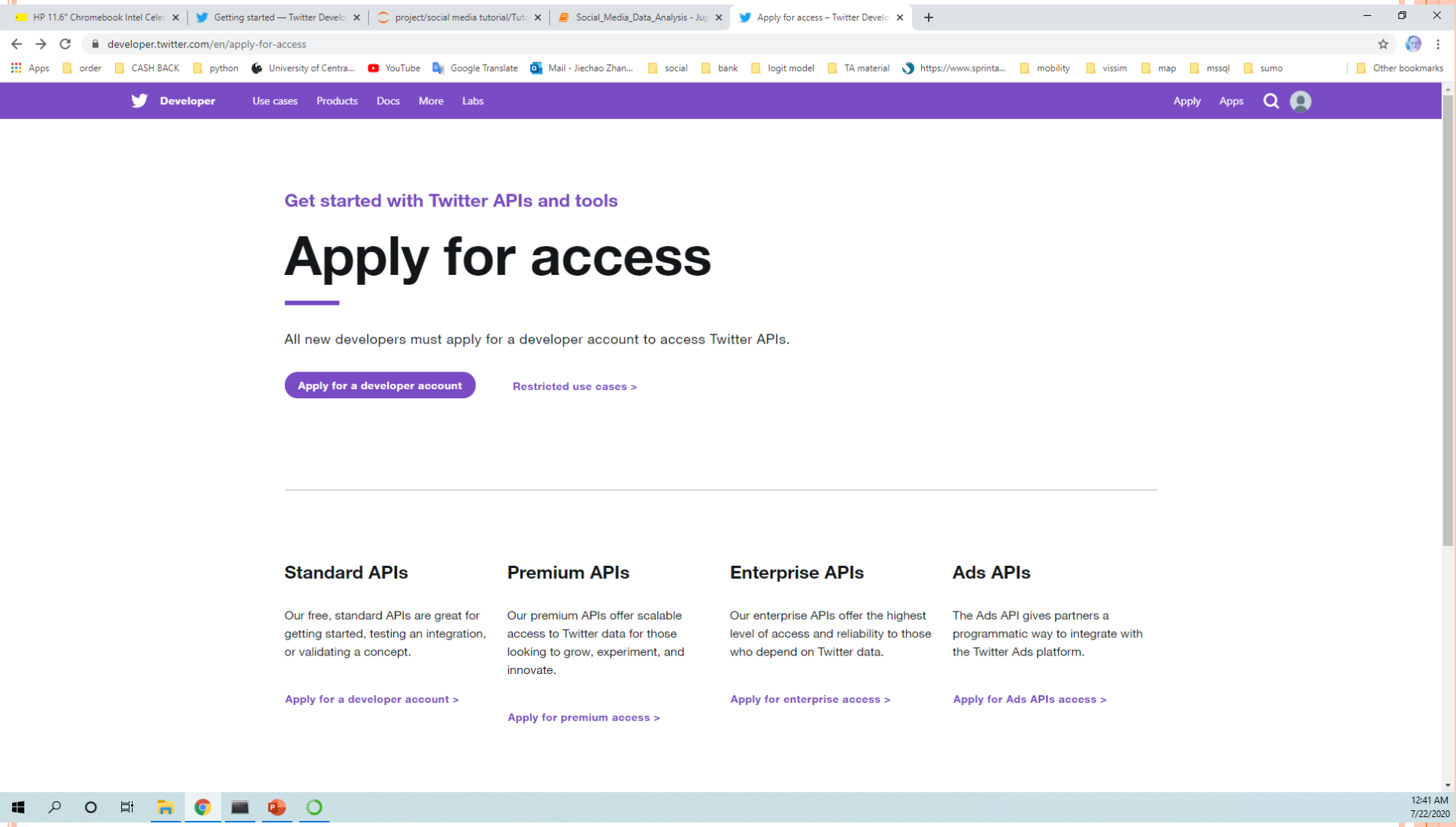


Data Collection – Install Packages



Data Collection – Apply Twitter Developer Account

Download Link: https://developer.twitter.com/en/apply-for-access



Data Collection – User accounts

jupyter Social Media Data Analysis Last Checkpoint: 12/24/2019 (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

1 Data Collection (user accounts) ¶

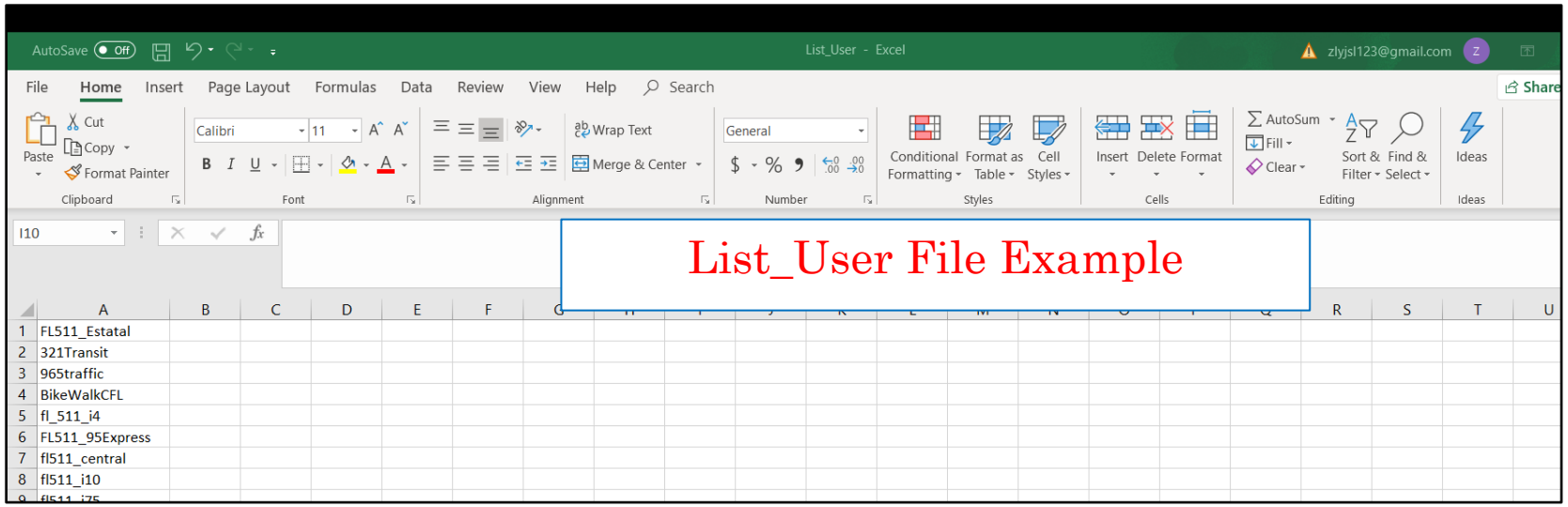
Necessary Packages

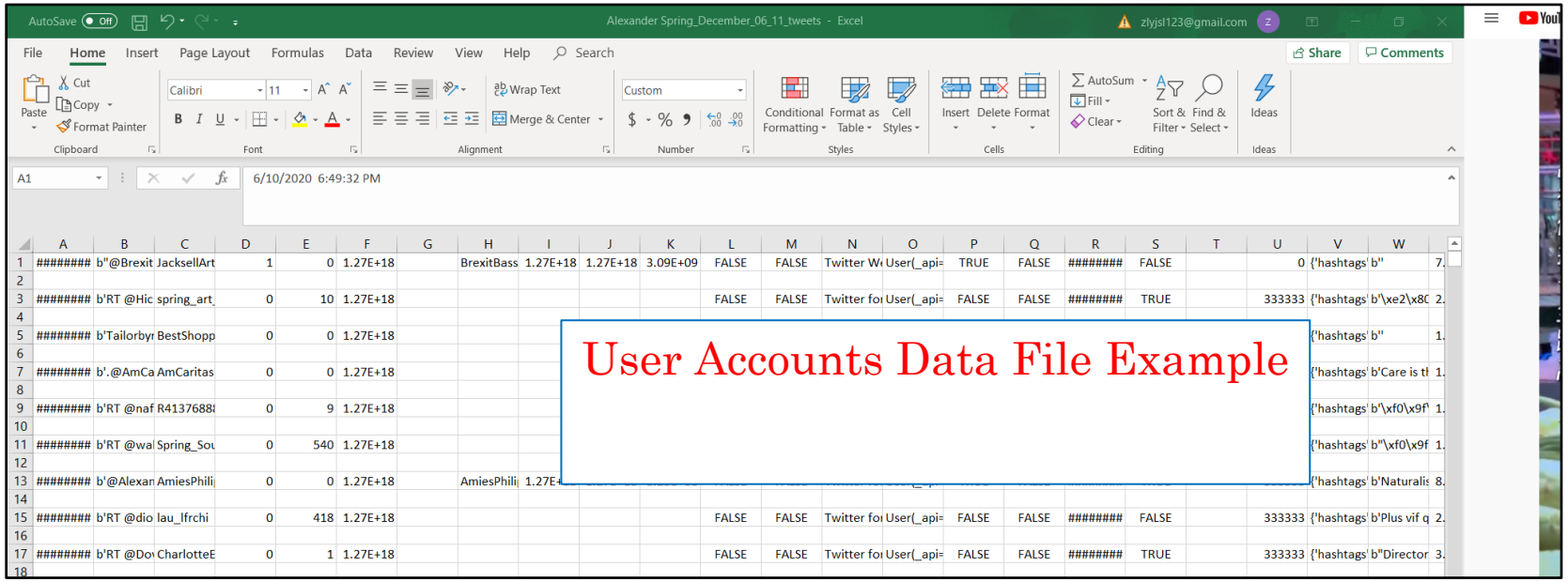
Path - List_User

Path – Save File

In [10]:
```
#!/usr/bin/env python
# encoding: utf-8

import tweepy  # https://github.com/tweepy/tweepy
import csv, pdb
import time

# Twitter API credentials
twitter_app_auth = {
    'consumer_key':
    'consumer_secret
    'access_token':
    'access_token_secret':
}

List_User = r'D:\project\social media tutorial\example data\List_User.csv'


def get_all_tweets(screen_name):

    # Twitter only allows access to a users most recent 3240 tweets with this method
    Collection_Path = r'D:\project\social media tutorial\example data\collection data\%s_June_10_tweets.csv' % screen_name
```

Data Collection – User accounts



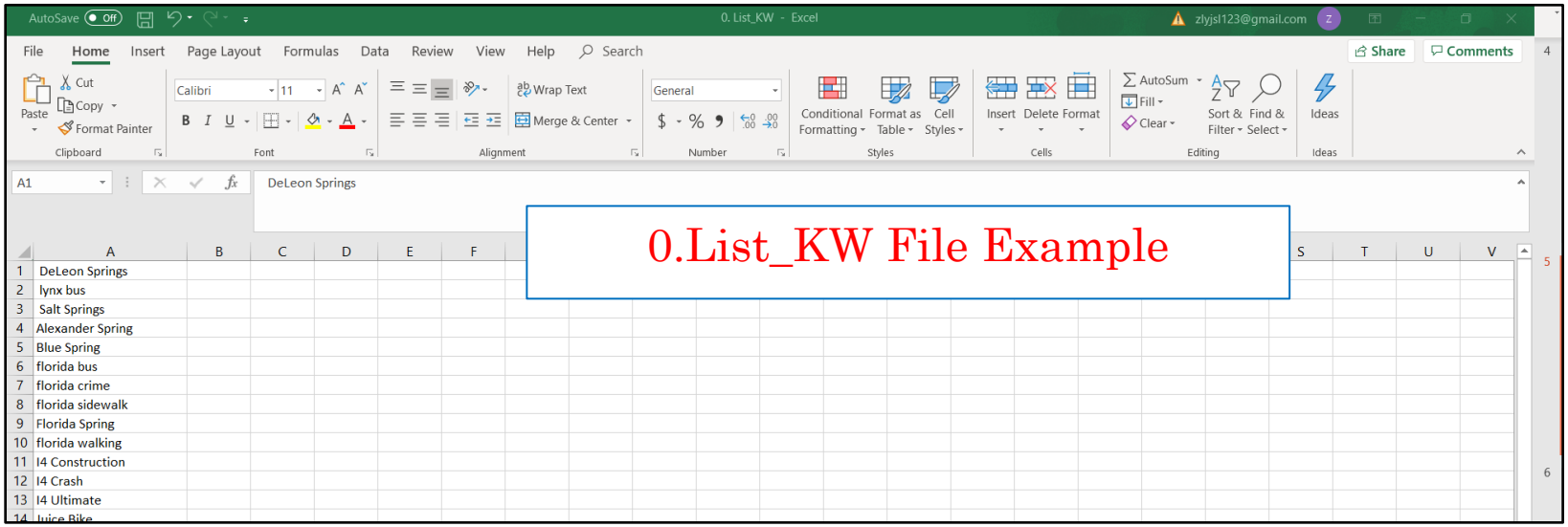


Data Collection – Keywords

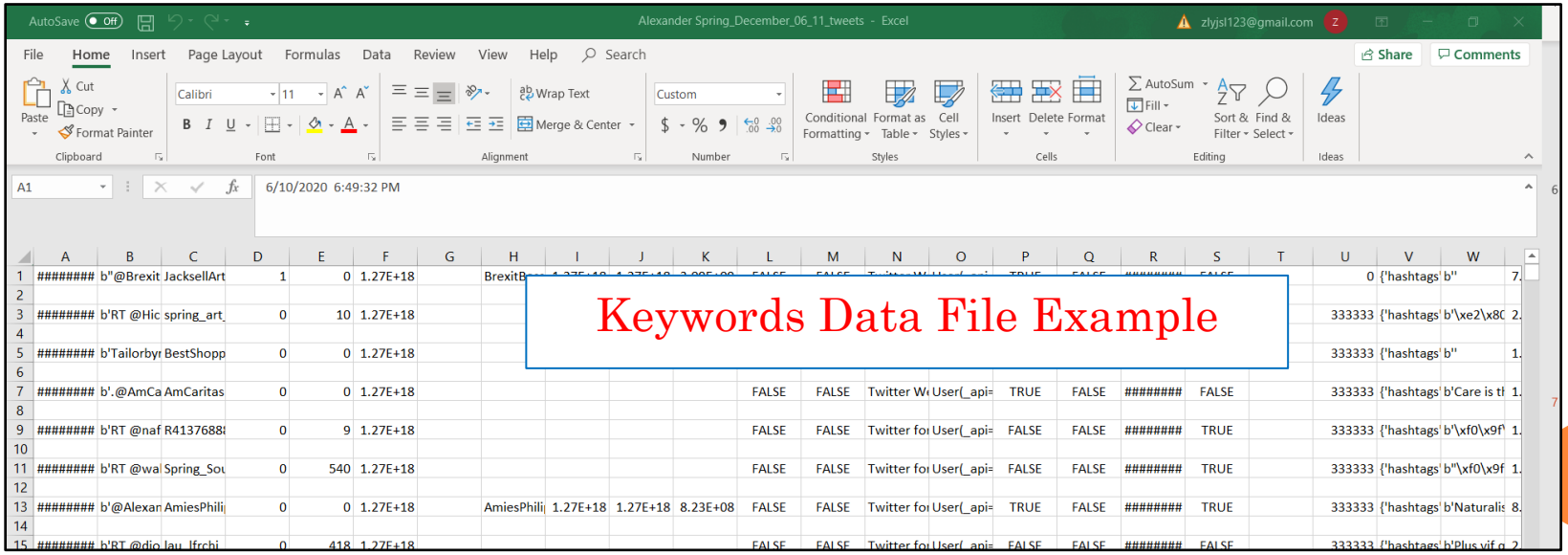
2 Data Collection (keywords)

14]:
```
import tweepy
import csv
from tweepy import Stream
from tweepy import OAuthHandler
from tweepy.streaming import StreamListener

consumer_key =
consumer_secret
access_token = '
access_token_secret =

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)

# Open/Create a file to append data
#csvFile = open('%s_tweets.csv' % == q, 'w', encoding='utf-8')
#Use csv Writer
#csvWriter = csv.writer(csvFile)

List_User = r'D:\project\social media tutorial\example data\0. List_KW.csv' #define the path


for line in open(List_User,'r', encoding='utf-8'):
    keyword = line.strip()

    save_path = r'D:\project\social media tutorial\example data\key word\%s_December_06_11_tweets.csv' % keyword #define the pat
```

Data Collection – Keywords



	A
1	DeLeon Springs
2	lynx bus
3	Salt Springs
4	Alexander Spring
5	Blue Spring
6	florida bus
7	florida crime
8	florida sidewalk
9	Florida Spring
10	florida walking
11	I4 Construction
12	I4 Crash
13	I4 Ultimate
14	Juice Bike



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	########	b"@Brexit	JacksellArt	1	0	1.27E+18		BrexitB													0	{'hashtags	b''
3	########	b'RT @Hic	spring_art_	0	10	1.27E+18															333333	{'hashtags	b'\xe2\x8(
5	########	b'Tailorbyr	BestShopp	0	0	1.27E+18															333333	{'hashtags	b''
7	########	b'.@AmCa	AmCaritas	0	0	1.27E+18						FALSE	FALSE	Twitter W	User(_api=	TRUE	FALSE	########	FALSE		333333	{'hashtags	b'Care is t
9	########	b'RT @naf	R4137688	0	9	1.27E+18						FALSE	FALSE	Twitter fo	User(_api=	FALSE	FALSE	########	TRUE		333333	{'hashtags	b'\xf0\x9f\
11	########	b'RT @wa	Spring_Sou	0	540	1.27E+18						FALSE	FALSE	Twitter fo	User(_api=	FALSE	FALSE	########	TRUE		333333	{'hashtags	b"\xf0\x9f
13	########	b'@Alexan	AmiesPhili	0	0	1.27E+18		AmiesPhili	1.27E+18	1.27E+18	8.23E+08	FALSE	FALSE	Twitter fo	User(_api=	TRUE	FALSE	########	TRUE		333333	{'hashtags	b'Naturalis
15	########	b'RT @dio	Jau_Ifrchi	0	418	1.27E+18						FALSE	FALSE	Twitter fo	User(_api=	FALSE	FALSE	########	FALSE		333333	{'hashtags	b'Plus vif q

Demo 2:
Sentiment Analysis and Visualization

Sentiment Analysis

3 Sentiment Analysis - keywords

24]:
```
import os
import pandas as pd
from textblob import TextBlob


path = r'D:\project\social media tutorial\example data\key word' #define the path


files = os.listdir(path) #define the files in the path

def modifystr(s):
    #s = s.str.replace('[^\w\s]','')
    s = s.replace('/','')
    s = s.replace(',','')
    s = s.replace('@','')
    s = s.replace('//','')
    s = s.replace('#','')
    s = s.replace('%','')
    s = s.replace('','')
    s = s.replace('\\','')
```

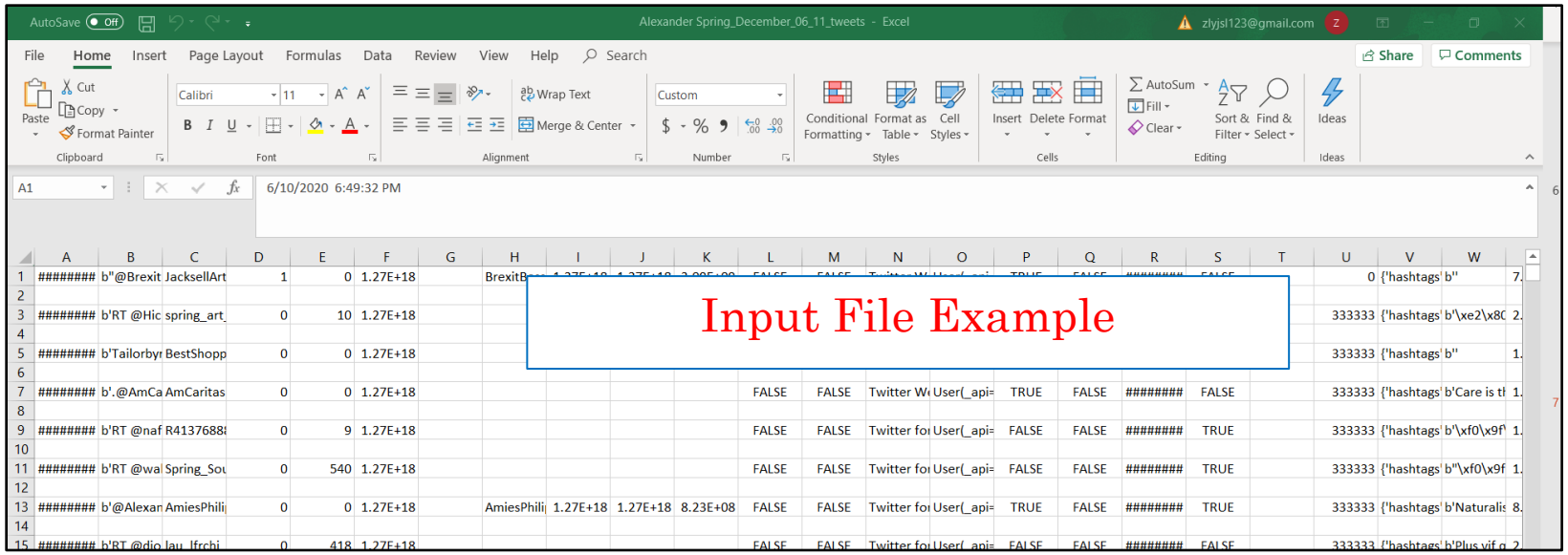
Path – Input Folder

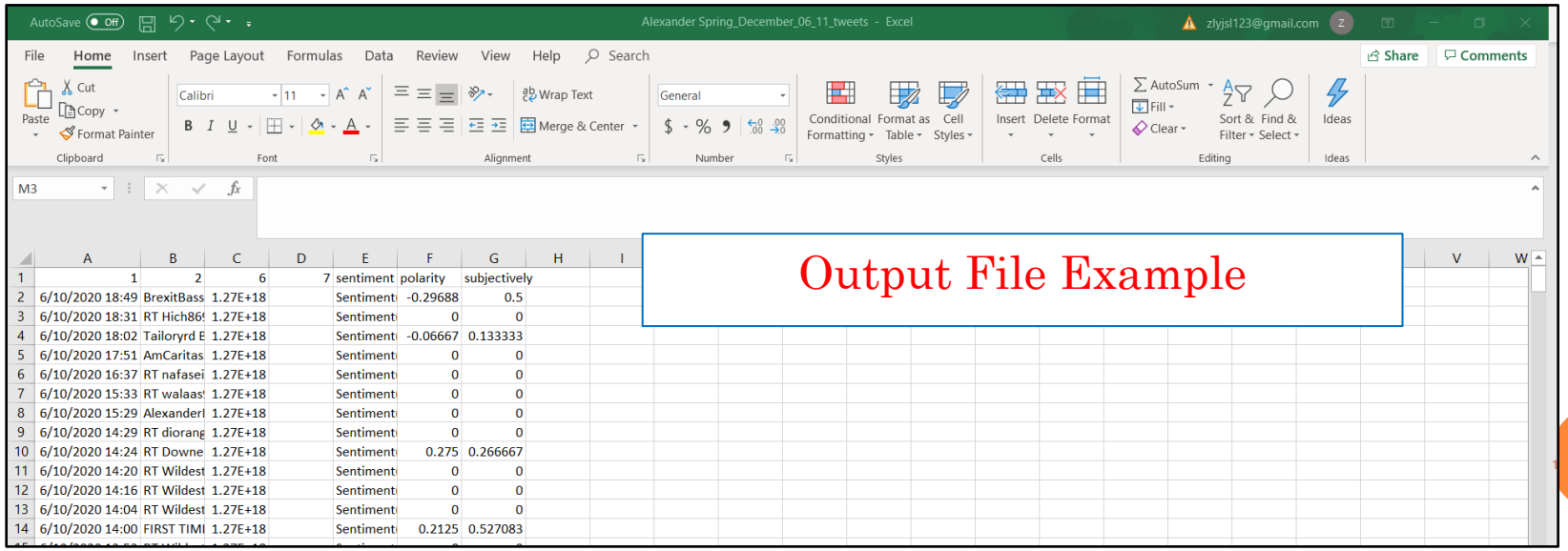
Sentiment Analysis

```
for i in range(0,len(files)):

    df = files[i]
    df_name = path + "\\" + df
    df_name_save_path = r'D:\\project\\social media tutorial\\example data\\key word senti\\' +df
    print(df_name_save_path)
    df_final = pd.read_csv(df_name,
                       names = ['1','2','3','4','5','6','7','8','9','10','11','12','13',
                                '14','15','16','17','18','19','20','21','22','23','24'] )
    df_final.dropna(axis = 0, how = 'all', inplace = True)


    df_final['2']=list(map(modifystr,df_final['2']))
    df_final['sentiment']=list(map(sentiment,df_final['2']))
    df_final.index = range(0,len(df_final))
    df_final_sentiment = pd.DataFrame(df_final[['1','2','6','7','sentiment']])
```

Path – Output Files

Sentiment Analysis





Sentiment Analysis Visualization

5 Sentiment Analysis Visualization

```
]: import pandas as pd
   import matplotlib.pyplot as plt
   import numpy as np

   path = r'F:\sentiment result\final_data\whole result\Sunshine  Skyway.csv'
   df = pd.read_csv(path, header = 0, names = ['id','time','text','account','geotagged','sentiment','polarity','subjectivity'])
   output_path = r'F:\sentiment result\final_data\whole result\Sunshine  Skyway.png'


   df.time = pd.to_datetime(df.time)

   #select data based on the time (half year)
   df_1 = df[(df.time.dt.year == 2017)&(df.time.dt.month>1)&(df.time.dt.month<8)]
   df_2 = df[(df.time.dt.year == 2017)&(df.time.dt.month>7)&(df.time.dt.month<13)]
   df_3 = df[(df.time.dt.year == 2018)&(df.time.dt.month>0)&(df.time.dt.month<9)]


   fig, axes = plt.subplots(3, 1, sharex=True, sharey=True)

   fig.set_size_inches(5,10)

   axes[0].hist(df_1.polarity, density = 1, bins=20, color='r',)
   axes[0].set_title('February 2017 - July 2017')
   axes[0].set_ylabel('Density')

   axes[1].hist(df_2.polarity,density = 1, bins=20, color='r')
   axes[1].set_title('August 2017 - December 2017')
   axes[1].set_ylabel('Density')
```

Path – Input Files

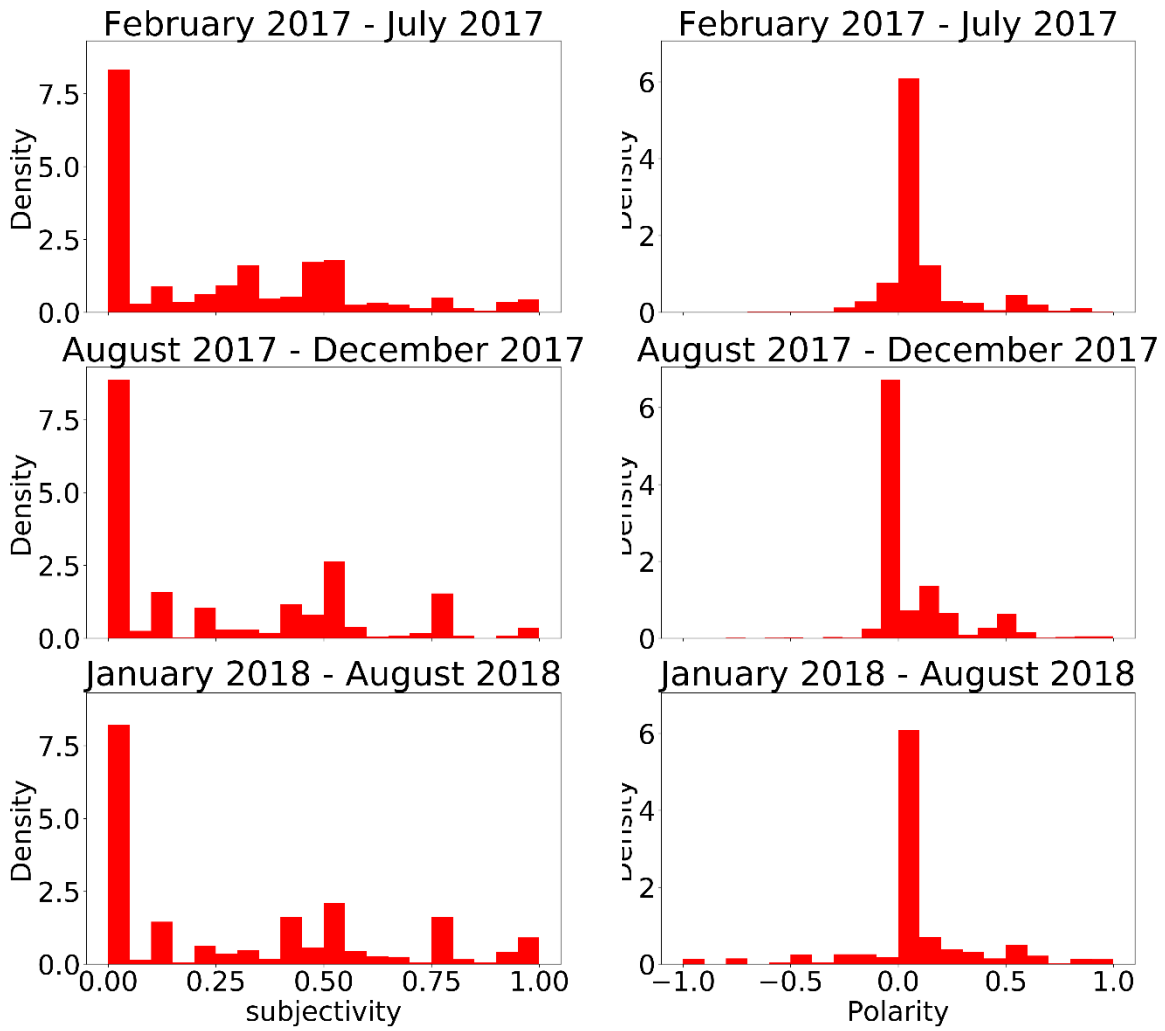
Path – Output Files

Set up the time periods

Time periods name in the figure

Sentiment Analysis Visualization

Example – Sentiment Analysis Visualization



Demo 3:
Topic Analysis

Topic Analysis – Data Processing

7 Topic Model Data Processing

In [68]:

```
import pandas as pd
import os
from textblob import TextBlob

path = 'F:\\sentiment result\\final_data\\whole result'

path_save = 'D:\\project\\social media tutorial\\example data\\topic model'

files = os.listdir(path)


for j in range(0,len(files)):

    df = files[j]
    path1 = path + "\\" + df

    df_txt = df.replace('.csv','')
    path_final = path_save + "\\" + df_txt + '.txt'

    names = ['index', 'time', 'text', 'id', 'geolocation','sentiment', 'polarity', 'subjectivity']

    df = pd.read_csv(path1)
```

Path – Input Folder

Path – Output Files

Topic Analysis – Data Processing

florida bus - Excel

	A	B	C	D	E	F	G	H
1		1	2	6	7	sentiment	polarity	subjectively
2	0	########	bRT ChrisK	1.03E+18		Sentimenti	0	0.0625
3	1	########	bRT ChrisK	1.03E+18		Sentimenti	0	0.0625
4	2	########	bRT ChrisK	1.03E+18		Sentimenti	0	0.0625
5	3	########	bRT joanne	1.03E+18		Sentimenti	0.519048	0.590476
6	4	########	bRT hamm	1.03E+18		Sentimenti	0	0
7	5	########	brealDona	1.03E+18		Sentimenti	0.25	0.333333
8	6	########	bRT Gwen	1.03E+18		Sentimenti	0.2	0.45
9	7	########	bSubtle hir	1.03E+18		Sentimenti	0	0
10	8	########	bRT ChrisK	1.03E+18		Sentimenti	0	0.0625
11	9	########	bRT Gwen	1.03E+18		Sentimenti	0.5	0.6
12	10	########	bRT ChrisK	1.03E+18		Sentimenti	0	0.0625
13	11	########	bEverywhe	1.03E+18		Sentimenti	0	0.0625
14	12	########	bRT joanne	1.03E+18		Sentimenti	0.519048	0.590476
15	13	#						
16	14	#						
17	15	#						
18	16	#						
19	17	########	bRT joanne	1.03E+18		Sentimenti	0.519048	0.590476
20	18	########	bRT joanne	1.03E+18		Sentimenti	0.519048	0.590476
21	19	########	bAwesome	1.03E+18		Sentimenti	0.278571	0.385714
22	20	########	bRT Gwen	1.03E+18		Sentimenti	0.5	0.6
23	21	########	bRT hamm	1.03E+18		Sentimenti	0	0
24	22	########	bRT Gwen	1.03E+18		Sentimenti	0.2	0.45
25	23	########	bRT Gwen	1.03E+18		Sentimenti	0.2	0.45
26	24	########	bRT hamm	1.03E+18		Sentimenti	0	0
27	25	########	bRT Gwen	1.03E+18		Sentimenti	0.2	0.45

Input File Example

florida bus - Notepad

```
1.0309664470679512e+18 "RT ChrisKingFL Everywhere we go in
1.0309663678377124e+18 "RT ChrisKingFL Everywhere we go in
1.030964885092528e+18 "RT ChrisKingFL Everywhere we go in
1.0309567547865496e+18 "RT joannefea Awesome bus tour educ
1.0309562966049956e+18 "RT hammel11 I havenxe2x80x99t said
1.0309554328677088e+18 "realDonaldTrumpnNRA FLGovScott Thi
1.030954321599447e+18 "RT GwenGraham South Florida friends
1.0309539232496681e+18 "Subtle hint from the Panama CityFl
1.0309516505743809e+18 "RT ChrisKingFL Everywhere we go in
1.0309511678561894e+18 "RT GwenGraham Our South Florida Ea
1.0309509267633888e+18 "RT ChrisKingFL Everywhere we go in
1.0309498999935384e+18 "Everywhere we go in Southwest Flor
1.0309494540426076e+18 "RT joannefea Awesome bus tour educ
1.0309452314060841e+18 "RT Cissy4Judge Cissy is very proud
1.0309363598239908e+18 "RT joannefea Awesome bus tour educ
1.0309362734255226e+18 "Awesome bus tour educating the vot
1.0309319836381184e+18 "RT GwenGraham Our South Florida Ea
1.0309318630980198e+18 "RT hammel11 I havenxe2x80x99t said
1.0309313480962662e+18 "RT GwenGraham South Florida friend
1.0309252249870458e+18 "RT GwenGraham South Florida friend
1.030918707630039e+18 "RT hammel11 I havenxe2x80x99t said
1.0309062709482291e+18 "RT GwenGraham South Florida friend
1.0309047633936957e+18 "RT GwenGraham South Florida friend
1.030901339650502 5e+18 "RT hammel11 I havenxe2x80x99t said
1.0309007449484737e+18 "I havenxe2x80x99t said a damn thin
1.0309000108194816e+18 "WDWToday Ray from Davenport Florid
```

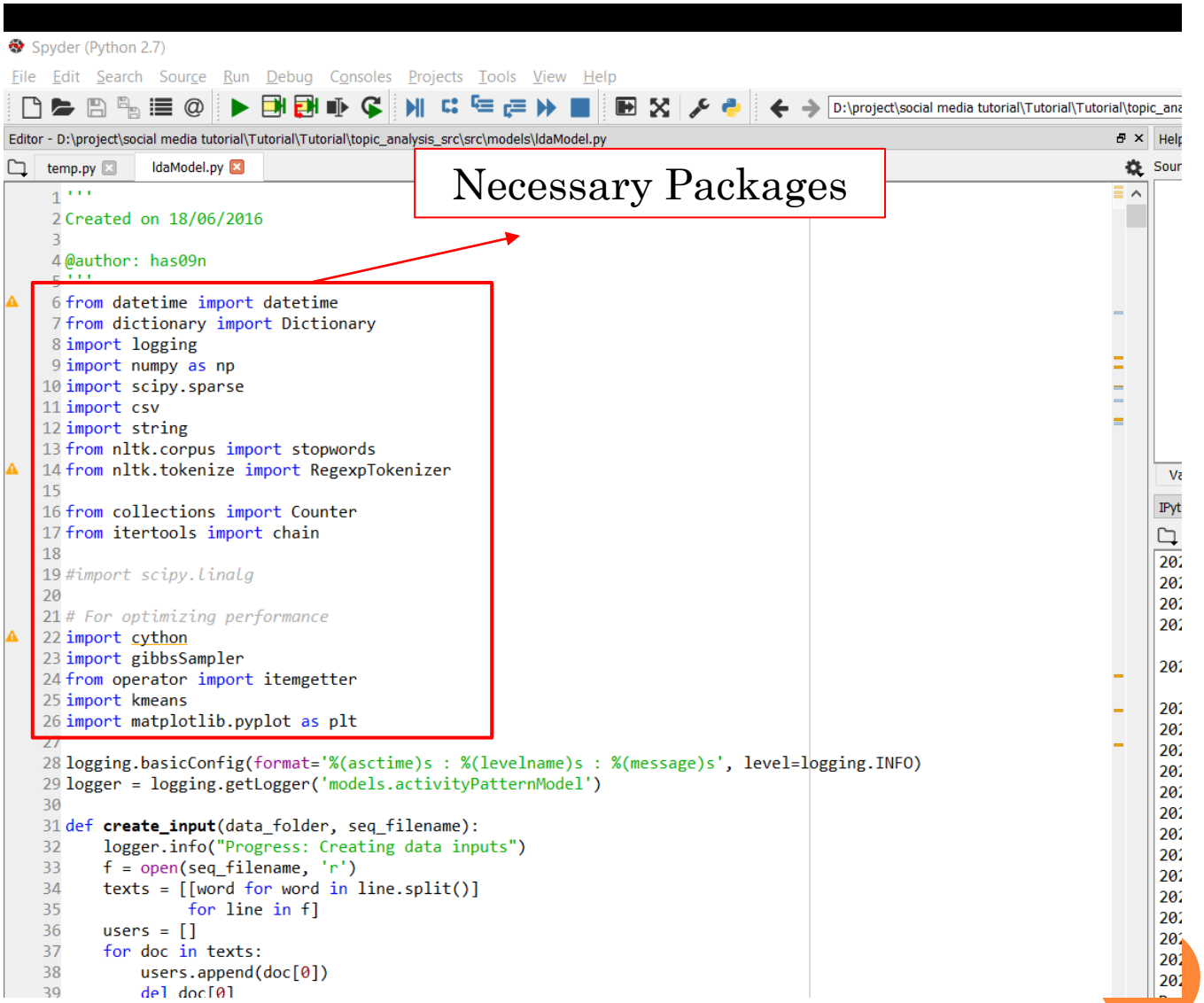
Output File Example

Topic Analysis – Run Topic Model

- Download Anaconda (python 2.7) 32-bit Graphical Installer
- Open Spyder from the anaconda navigator (python 2.7) version
- Open the ldaModel.py
- Install all the necessary python packages
- Change the input path and file name
- Run the model

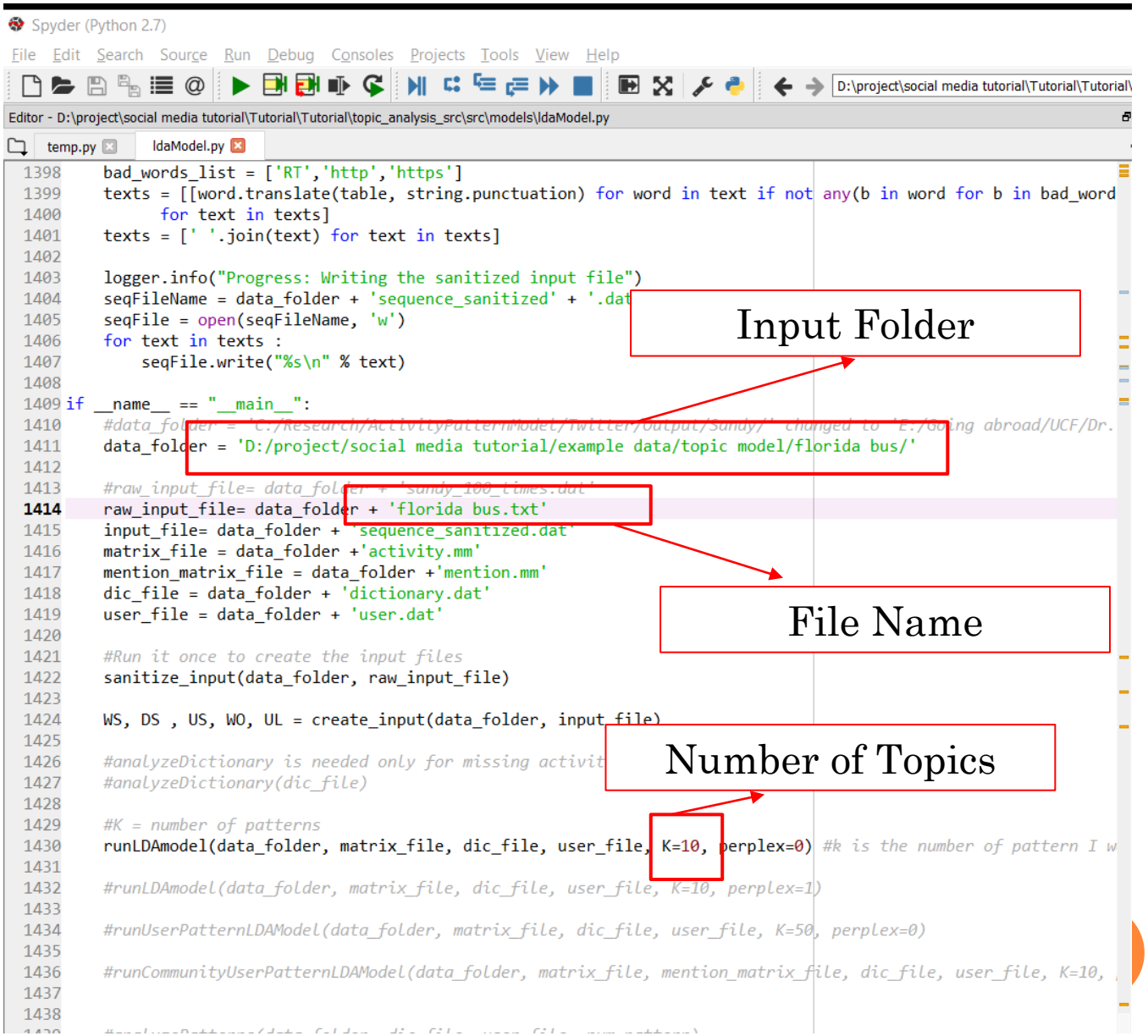
Topic Analysis – Run Topic Model

Example – Topic Analysis Model in Spyder



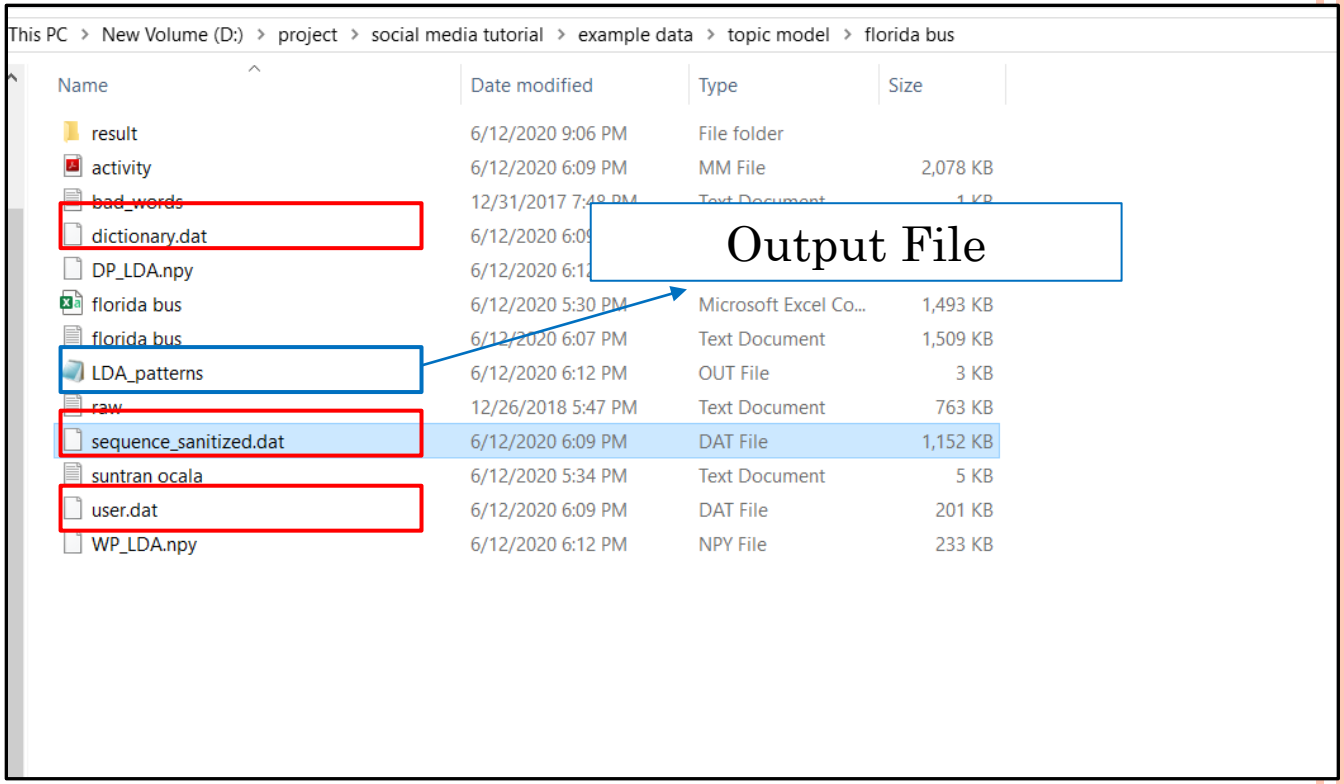
Topic Analysis – Run Topic Model

Example – Topic Analysis Model in Spyder



Topic Analysis – Run Topic Model

Tips: In the input file folder, there should be two necessary files – *user.dat*, *dictionary.dat* and *sequence_sanitized.dat* , which can be seen as right figure.



Topic Analysis – Data Processing for Visualization

Input File

8 Data Processing for Visualization

```
[92]: df = pd.read_csv(r'D:\project\social media tutorial\example data\topic model\florida bus\LDA_patterns.out', sep = ' ', names = [
      print(df)
      num_class = 10
      topic_num = 10

      topic_list = []
      words_list = []
      probability_list = []
      final_name = ['Topic', 'Words
      df_final = pd.DataFrame(columr

      for i in range(0, topic_num):

          topic_num = i + 1

          for j in range(2, num_class+2):

              id_num = j+i*12

              topic_name = 'Topic #' +str(topic_num)

              topic_list.append(topic_name)
              words_list.append(df.type[id_num])
              probability_list.append(df.prob[id_num])

      df_final['Topic'] = topic_list
      df_final['Words'] = words_list
      df_final['Probability'] = probability_list

      df final
```

Number of Topics

Topic Analysis – Data Processing for Visualization

LDA_patterns - Notepad

File Edit Format View Help

```
Pattern1 Prob. 0.109347

Activity Prob.
school 0.091001
There 0.073810
Florida 0.057103
bus 0.052753
week 0.050682
new 0.049716
year 0.048473
see 0.044330
ICE 0.040948
last 0.040809

Patte

Activity Prob.
tour 0.044066
bus 0.038879
MayorLevine 0.018654
The 0.017526
Tampa 0.013240
days 0.012939
LiveFromLivingRooms 0.012112
4 0.011436
Floridas 0.011135
1 0.010534

Pattern3 Prob. 0.096877
```

Input File

AutoSave Off

File Home Insert Page Layout Formulas Da

Paste Cut Copy Format Painter Clipboard

Calibri 11 Font

G193

	A	B	C	D	E
1	Topic	Words	Probability		
2	Topic #1	amp	0.053861		
3	Topic #1	11	0.040138		
4	Topic #1	near	0.039083		
5	Topic #1	like	0.035388		
6	Topic #1	He	0.033541		
7	Topic #1	Hes	0.031166		
8	Topic #1	moving	0.030638		
9					
10					
11	Topic #1	allowed	0.03011		
12	Topic #2	bus	0.167361		
13	Topic #2	Florida	0.06021		
14	Topic #2	driver	0.042303		
15	Topic #2	one	0.01735		
16	Topic #2	WATCH	0.017056		
17	Topic #2	street	0.016175		
18	Topic #2	florida	0.015295		
19	Topic #2	hitting	0.015295		
20	Topic #2	pedestrian	0.015001		
21	Topic #2	fell	0.014708		
22	Topic #3	bus	0.163686		
23	Topic #3	Florida	0.142392		
24	Topic #3	way	0.019499		
25	Topic #3	ON	0.013719		
26	Topic #3	Im	0.012502		
27	Topic #3	video	0.011894		

Output File

Topic Analysis –Visualization

8 Topic Model Visualization ¶

In [93]:
```
import csv,pdb
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from datetime import datetime
import pickle
from datetime import datetime
from matplotlib import style
import matplotlib.ticker as mticker
import matplotlib.dates as mdates
import matplotlib.cm as cm
import math

SMALL_SIZE = 12
MEDIUM_SIZE = 32
BIGGER_SIZE = 40

plt.rc('font', size=SMALL_SIZE)
plt.rc('axes', titlesize=BIGGER_SIZE)
plt.rc('axes', labelsize=MEDIUM_SIZE)
plt.rc('xtick', labelsize=MEDIUM_SIZE)
plt.rc('ytick', labelsize=MEDIUM_SIZE)
plt.rc('legend', fontsize=SMALL_SIZE)
plt.rc('figure', titlesize=BIGGER_SIZE)


path_input = r"F:\topic model\tm\florida bus\result\florida bus.csv"


path_output = r"D:\project\social media tutorial\example data\topic model\florida bus\Topic_non_RT_user_heatmap_1.png"
```

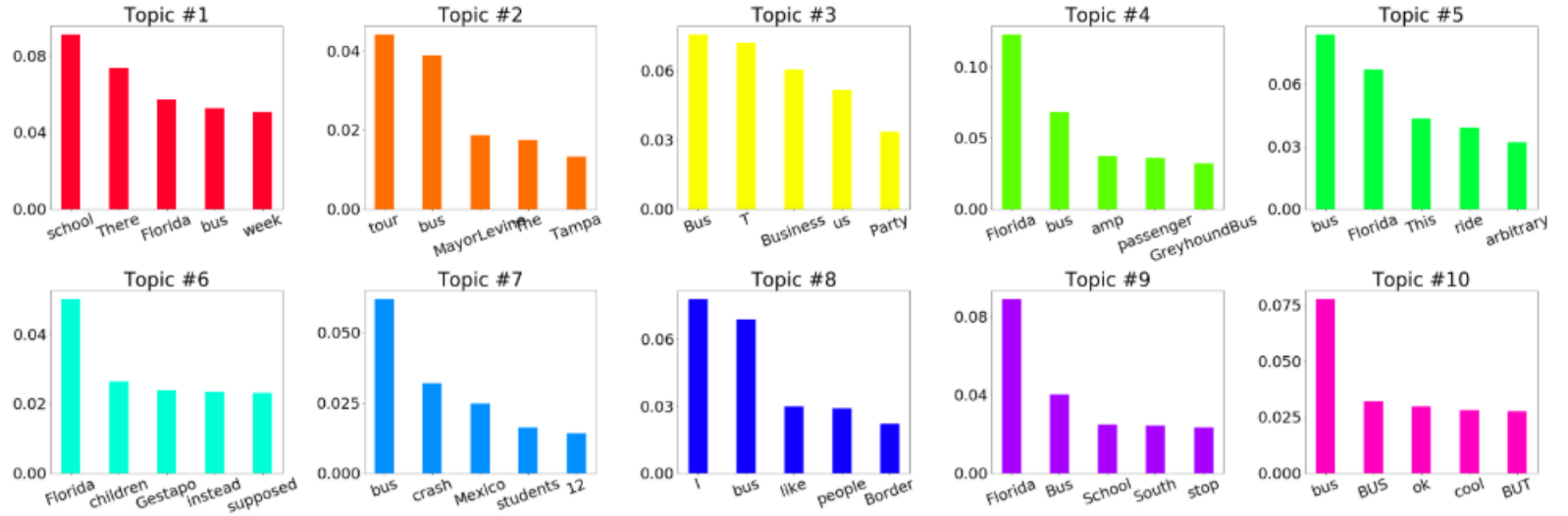
Input File

Output Figure

Topic Analysis –Visualization

Example of output figure

```
plt.tight_layout()
plt.show()
```



Questions