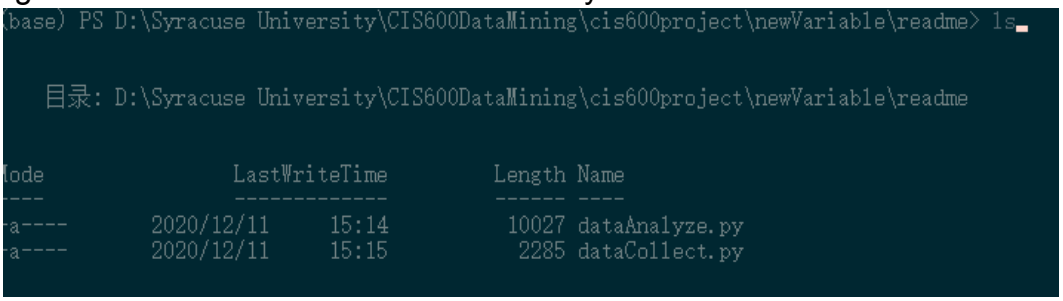
Sentiment Analysis of Presidential Candidates' Twitter Comments

1. **Installation**

   Our project system has two python files which correspond to the part of data collection and data analysis. To run our system, please download the code at first. Then start your python virtual environment like conda or anaconda and enter into our project directory. You can see in the current directory there are two python scripts:

   - dataCollect.py is the file processing comments based on candidates and converting them into csv files. These csv files will be used in the next data analysis step.
   - dataAnalyze.py with the input of csv files created by dataCollect.py will generate a dataset at first. Then both candidates datasets will be randomized and balanced in a fixed size. Finally it will generate several figures for each candidate for further analysis.



2. **Compile and Run**

   We recommend you to manage the dependencies by anaconda. Before you compile and run the files you have to set up the modules required in this system:

   - csv: CSV file reading and writing
   - tweepy: a module provide RESTful twitter Apis in python and can be installed by this tutorial: http://docs.tweepy.org/en/latest/install.html
   - numpy: a module to better perform python array
   - TextBlob: we use the already built in sentiment analysis tools in TextBlob so please install TextBlob module: https://textblob.readthedocs.io/en/dev/
   - pandas: we use pandas for manage our dataset

   To support static figure generation you also need to install kaleido

   - kaleido: pip install -U kaleido

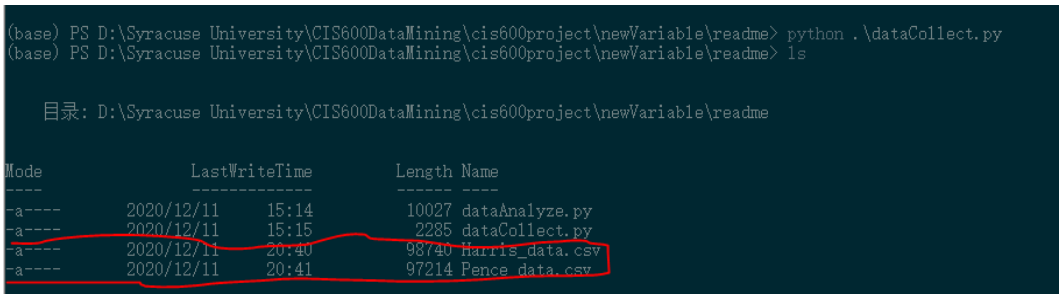
   Some plot module we will use in this project:

   - matplotlib: we use the pyplot in it to store image
   - seaborn: we generate displot by this module
   - plotly: we generate histogram and comment image by this model

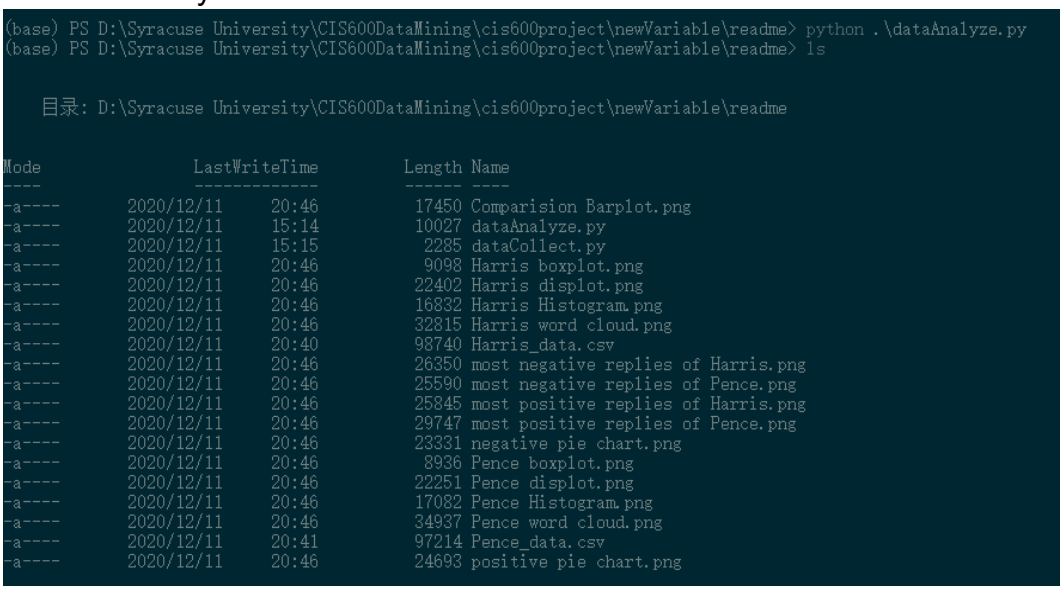
   After setting up the environment now we run this system by two steps in the command line

   - > python dataCollect.py

After running this command check your directory you will find two new csv files

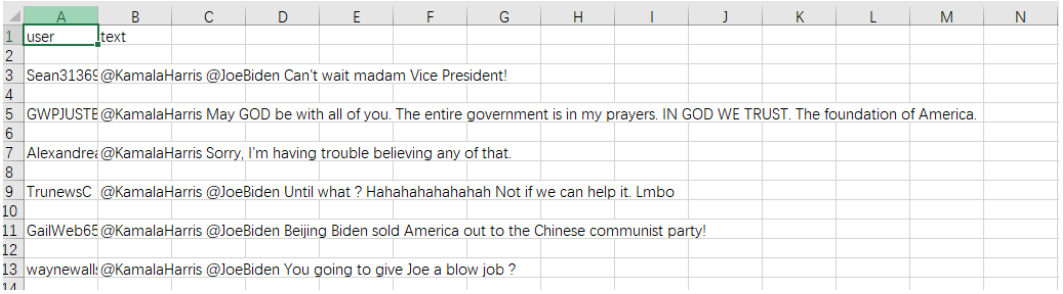


- > python dataAnalyze.py
  This command will start analyze the data and return several figures to show the analysis result



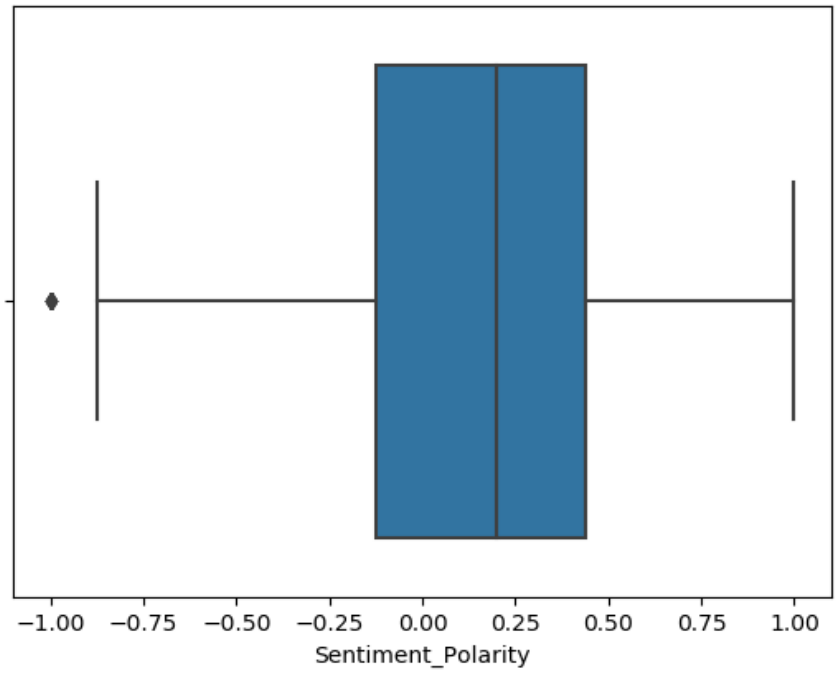
3. Data Files and Result Files

- Data files
  In the example we showed before you can see after the first command it generates two csv files the detail of the csv file name can be found in the comment in dataCollect.py. The generate csv file contains two column: user and comment.

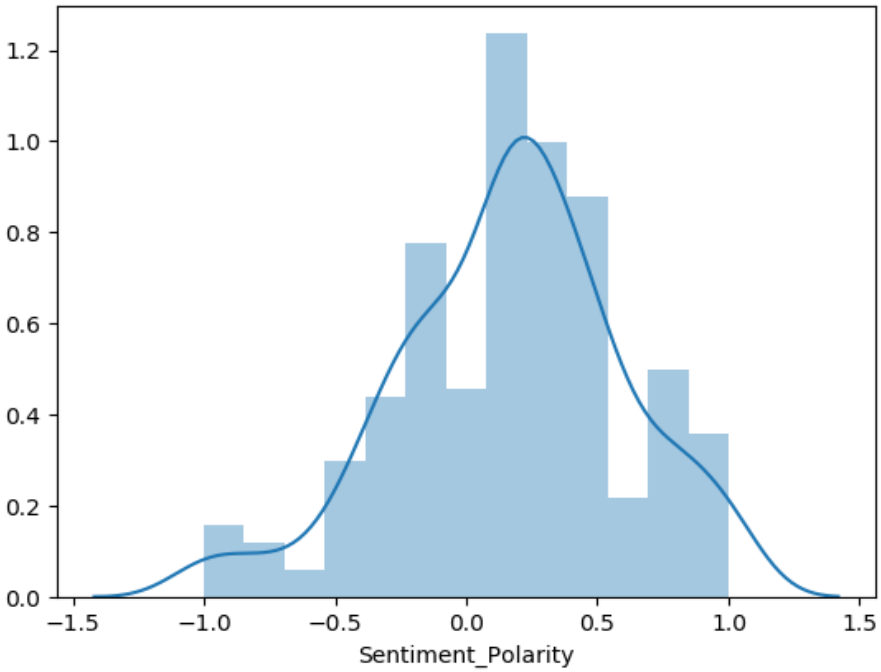


	A	B
1	user	text
2		
3	Sean31369	@KamalaHarris @JoeBiden Can't wait madam Vice President!
4		
5	GWPJUSTE	@KamalaHarris May GOD be with all of you. The entire government is in my prayers. IN GOD WE TRUST. The foundation of America.
6		
7	Alexandre	@KamalaHarris Sorry, I'm having trouble believing any of that.
8		
9	TrunewsC	@KamalaHarris @JoeBiden Until what ? Hahahahahahahah Not if we can help it. Lmbo
10		
11	GailWeb65	@KamalaHarris @JoeBiden Beijing Biden sold America out to the Chinese communist party!
12		
13	waynewall	@KamalaHarris @JoeBiden You going to give Joe a blow job ?

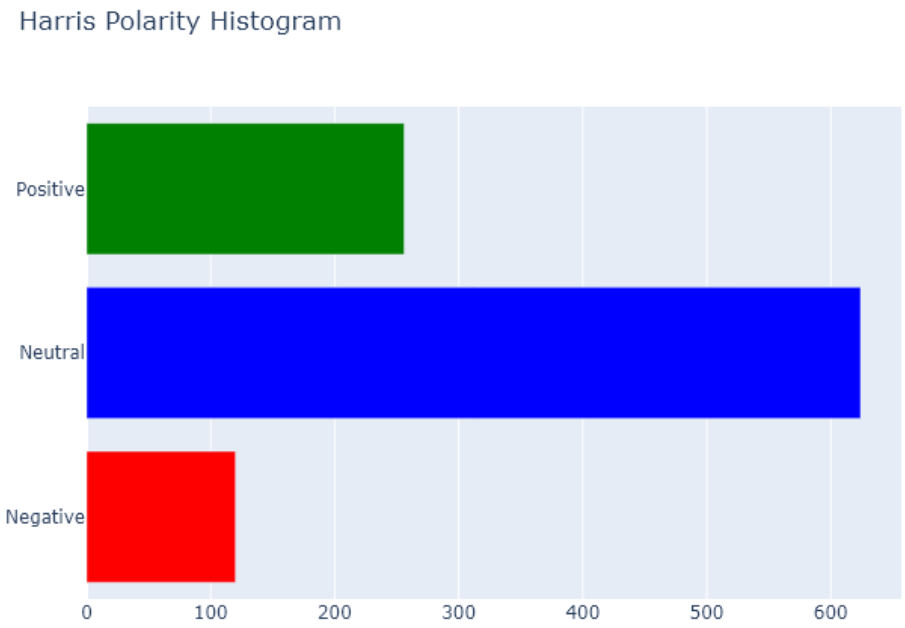
- Result files
  - Boxplot figure



  - Displot figure

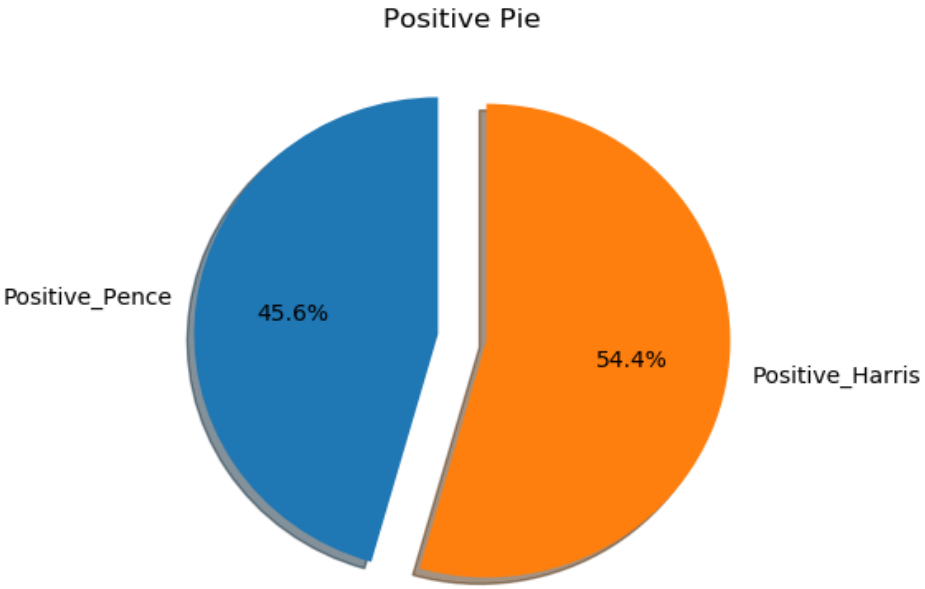


- Histogram figure



- Word Cloud

- Pie Charts



- Comparision barplot