Presidential Election Prediction

TASK 1: Data Collection

Steps followed to collect Twitter Data:

- Use the API key, API secret, token key, token secret to connect to twitter and access the API.
- Made use of tweepy library to collect twitter data.
- Defined search term, eg: "#DonaldTrump" and made use of tweepy.Cursor to download tweets that contain the search term in them. Specified the time frame and language for which the tweets should get collected. Specified the number of tweets that are to be collected.
- Save the tweets in json format so we don't have to collect them using the API every time we have to re-run the program. We can just read the saved data and proceed from there.
- Repeat the same process with search term: "#JoeBiden".
- Save the new data in a different json file for ease of usage.

Sample Tweets (text of the tweets):

```
In [6]: dataDonaldTrump['text'][15]

Out[6]: 'Coronavirus: US to halt funding to WHO, says Trump\n#WHO #DonaldTrump #COVID19 #science #funding #research\nhttps://t.co/inNrf
        pHe9i'
```

```
In [7]: dataDonaldTrump['text'][35]

Out[7]: '@Strandjunker #DonaldTrump wouldn\'t just "let us die" he would "kill is" if it meant keeping him powerful… https://t.co/nzqZy
        fFi7B'
```

```
In [8]: dataJoeBiden['text'][15]

Out[8]: "@joyannreid; @K_JeanPierre, #Warren's supporters would have no issue moving over to #JoeBiden, so it matters little… https://
        t.co/kIfMnz5RYW"
```

```
In [9]: dataJoeBiden['text'][35]

Out[9]: "When #Trump losses in the #2020Election - I wager he won't even be on stage when #JoeBiden is sworn in as President… https://
        t.co/ofrlGd5FoS"
```

TASK 2: Data Analysis

Check when were the earliest and latest timestamps of tweets of both candidates in the dataset:

```
Earliest tweet extracted for #DonaldTrump is at :  2020-04-14 23:31:24
Latest tweet extracted for #DonaldTrump is at :  2020-04-15 23:59:58
Earliest tweet extracted for #JoeBiden is at :  2020-04-14 22:40:12
Latest tweet extracted for #JoeBiden is at :  2020-04-15 23:59:35
```
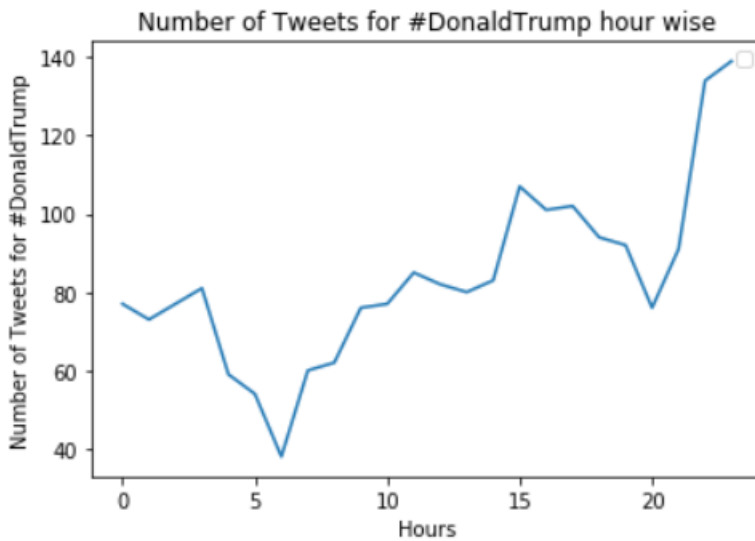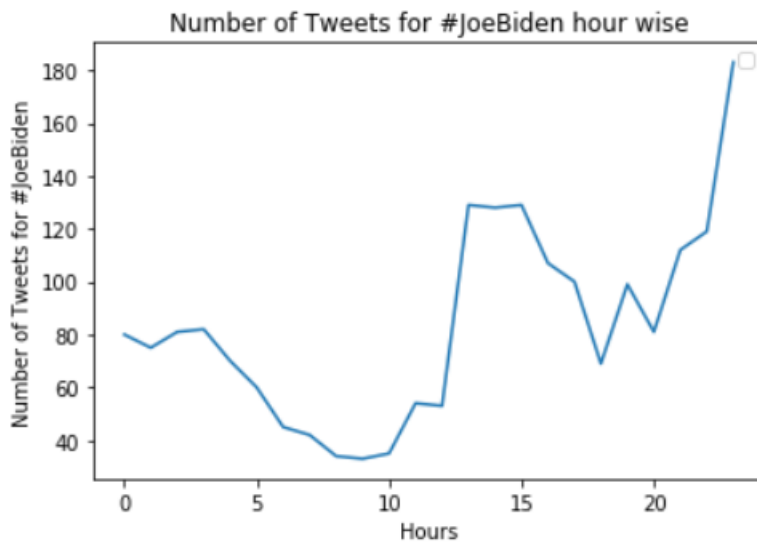
1.

Plot of number of tweets per hour for #DonaldTrump:

Plot of number of tweets per hour for #JoeBiden:

2. Perform sentiment analysis on tweets of both candidates.

- The most positive score observed for #DonaldTrump is 0.950 and that for #JoeBiden is 0.938. The most negative score for #DonaldTrump is -0.957 and that for #JoeBiden is -0.955. The image below shows what is written in the tweet along with the scores.

```
Most positive score and tweet for #DonaldTrump:    0.950 : "Seeking for a lovely and adorable sugar babe,a baby I can take good
care of and make her happy always...… https://t.co/i9N4edqniT"

Most negative score and tweet for #DonaldTrump:  -0.957 : "Poor poor stupid #DonaldTrump. He to stupid to realize that Blaming
the WHO had absolutely no political upside. What a dumb-ass."

Most positive score and tweet for #JoeBiden:    0.938 : "@ewarren @JoeBiden What a fantastic endorsement, Senator!! Wow, so spo
t on and so glad you're in this with us! #LetsGetToWork  #JoeBiden"

Most negative score and tweet for #JoeBiden:  -0.955 : "#JoeBiden is a rapist #JoeBiden is a cheater #JoeBiden is establishmen
ts bitch right now. I would have said whore b… https://t.co/3OM86hBIff"
```

- We observed that there is more positive and less negative sentiment for Joe Biden as compared to Donald Trump. The count of positive tweets is more for Donald Trump than Joe Biden.

```
Total positive sentiment for #DonaldTrump :   264.37720000000013

Total negative sentiment for #DonaldTrump :   -349.23339999999956

 Total tweet count for negative sentiment for #DonaldTrump :   735

Total tweet count for positive sentiment for #DonaldTrump :   571

Total tweet count for neutral sentiment for #DonaldTrump :   694

Total positive sentiment for #JoeBiden :   379.3479999999994

Total negative sentiment for #JoeBiden :   -250.41899999999993

Total tweet count for negative sentiment for #JoeBiden :   551

Total tweet count for positive sentiment for #JoeBiden :   836

Total tweet count for neutral sentiment for #JoeBiden :   613
```

- Check if the tweets with most retweet counts have a greater number of positive sentiment tweets or negative sentiment tweets for each candidate.

  For #DoanldTrump, the top 5 maximum retweet count tweets have two tweets with negative sentiment, 0 tweets with positive sentiment and 3 tweets with neutral sentiment.

	retweet_count	text	scores
1545	13	US President #DonaldTrump says some US states ...	0.0000
382	13	Donald Trump says deciding when to reopen the ...	0.0000
804	12	Indian fake journalists need taste of #DonaldT...	-0.6093
41	12	President Trump has said he doesn't like shaki...	-0.1511
283	12	One like=One vote. Trump 2020!!!! #GOP #Trump ...	0.0000

  For #JoeBiden, the top 5 maximum retweet count tweets have four tweets with positive sentiment, 0 tweet with negative sentiment, and 1 tweets with neutral sentiment.

	retweet_count	text	scores
943	16	Vote Joe Biden Come with me if you want to liv...	0.0772
986	16	Thank you President Obama for your moving word...	0.5346
161	16	When asked if she would accept an offer to run...	0.7506
1448	16	#How2020CouldGetWorse \n\nBy electing this dud...	0.0000
1515	15	Yep. This. Obama's legacy is a Heritage Founda...	0.2960

  All the above shown analysis indicate that the possibility of Joe Biden winning the elections is more than Donald Trump.

3.

Word cloud for #DonaldTrump:

Word cloud for #JoeBiden:

The word cloud indicates that the tweets contain the above shown words for each candidate and the bigger size indicates that those words were used more number of times.

We can observe that people are talking about things related to coronavirus and world health when it comes to Donald Trump. But when it comes to Joe Biden, we see a word "vote" in the word cloud. This shows that people actually are rooting for Joe Biden to win the elections.

The result is same for both the techniques used. According to my analysis of the collected twitter data, Joe Biden will win the elections of 2020.