Katherine Pollan

ACC- Python for Data Analytics

Final Project

For my final project, I pulled data from President Trump's (@realDonaldTrump) twitter feed from his first term (1/20/2017-9/6/2020) and sorted it by hour to show the most active hour it was used across ~3 years and 17.7k of tweets. The most active hour for Trump's twitter was 8am EST, with 1,327 instances of it being used.

The question that prompted this line of thought and twitter scraping was honestly, 'how often does President Trump tweet at 3am in the morning' – which ended up being a non-zero number, Trump had 27 instances of twitter activity at 3am EST. Trumps most inactive twitter hour was 4am EST with only 20 instances.

@realDonaldTrump Twitter Activity: Most Active Hours
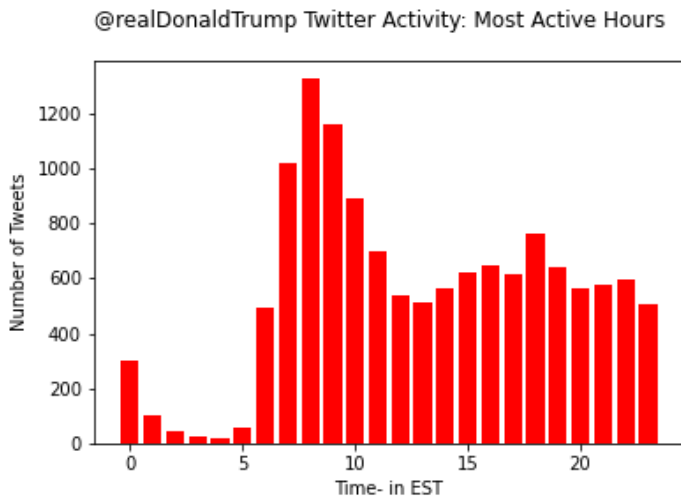
```
#Trump's twitter data
#I had Trump's tweets in a CSV called trump_tweets

#splitting up the time from the datetime stamp in the CSV
trump_time = [item.split(" ")[1] for item in
trump_tweets['Date'].values]

#isolating the hour and making a list of hours tweeted
trump_hour = [item.split(":")[0] for item in trump_time]

#converting UTC to EST for readability
trump_hourEST = [(int(item)-4)%24 for item in trump_hour]

#turning the previous list into a dictionary with 'counts' of
how many instances
trhours_tweeted = collections.Counter(trump_hourEST)

#ordering the dictionary so it goes 00 to 23
trhours_tweeted_od =
collections.OrderedDict(sorted(trhours_tweeted.items()))

trump = pd.DataFrame.from_dict([trhours_tweeted_od])
#Plotting Trump's twitter data
#The dictionary values are acting as the X,Y coordinates- the
# of instances (or values) are the Y axis
#The hour of the day (the key of the dictionary) is the x axis
#techinically I didn't need to reorder/resort the dictionary,
but made more visual sense
fig, ax = plt.subplots()
plt.bar(trhours_tweeted_od.keys(),
trhours_tweeted_od.values(), color='r')
plt.suptitle('@realDonaldTrump Twitter Activity: Most
Active Hours')
plt.xlabel("Time- in EST")
plt.ylabel("Number of Tweets")
plt.show()
print('Tweets collected from 1/20/2017 to 9/6/20')
```

The method's I used could not account for staff posting on his account, bot activity (auto liking a tweet with a specific hashtag for example) or time zone differences if he traveled which undoubtedly accounts for some percentage of the activity- but I don't know how much specifically.

I also pulled the twitter data from former president Barack Obama's (@barackobama) twitter from his second term. I initially pulled from his first term as a more 1:1 comparison, but his twitter was not as frequently used in his first term as his second, so I re-pulled using his second term (1/20/13-1/20/17).

Obama had 6.4k instances of twitter activity in his second term, and his most frequent hour was noon EST with 800 instances of twitter activity, and his most infrequent hour was 7am with 1 instance and a gap between 1am and 7am. (Trump has no such gap with instances of twitter activity at every hour).

@barackobama Twitter Activity: Most Active Hours
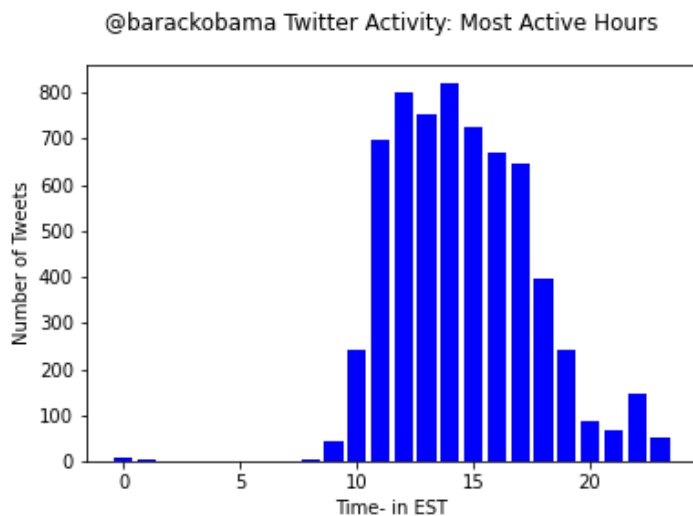
```
#Obama's twitter data
#Obama's tweets were in a CSV called obama_tweets

#splitting up the time from the datetime stamp in the CSV
obama_time = [item.split(" ")[1] for item in
obama_tweets['Date'].values]

#isolating the hour and making a list of hours tweeted
obama_hour = [item.split(":")[0] for item in obama_time]

#converting UTC to EST for readability
obama_hourEST = [(int(item)-4)%24 for item in
obama_hour]

#turning the previous list into a dictionary with 'counts' of
how many instances
obhours_tweeted = collections.Counter(obama_hourEST)

#ordering the dictionary so it goes 00 to 23
obhours_tweeted_od =
collections.OrderedDict(sorted(obhours_tweeted.items()))

obama = pd.DataFrame.from_dict([obhours_tweeted_od])

#Plot of President Obama's Twitter activity by hour
fig2, ax = plt.subplots()
plt.bar(obhours_tweeted_od.keys(),
obhours_tweeted_od.values(), color='b')
plt.suptitle('@barackobama Twitter Activity: Most Active
Hours')
plt.xlabel("Time- in EST")
plt.ylabel("Number of Tweets")
plt.show()
print('Tweets collected from 1/20/2013 to 1/20/2017')
```

My question at this point started to become 'well, when DO world leaders get on twitter? Maybe late at night are the only free moments that they have' and in thinking of country leaders who are active on Twitter, I thought of Canadian Prime Minister, Justin Trudeau.

I grabbed Trudeau's tweets from his first term as PM- 11/4/2015 to 11/20/2019. In that timeframe, he had 10.8k instances of twitter activity, with his most active hour being 2pm EST with 909 instances. His least active hour was 5am

with only one instance of activity. Interestingly (to me at least), Trudeau shows more late-night activity then Obama did- there were over 400 instances of twitter activity at 11pm EST for instance, where Obama only had over 100 but Trudeau does have a gap from 2am to 5am with no activity. As noted above, the volume could also be bot activity, staff, or time zone changes from EST- but (to me) it is still interesting numbers, especially regarding the sheer volume of tweets from each leader.

@JustinTrudeau Twitter Activity: Most Active Hours
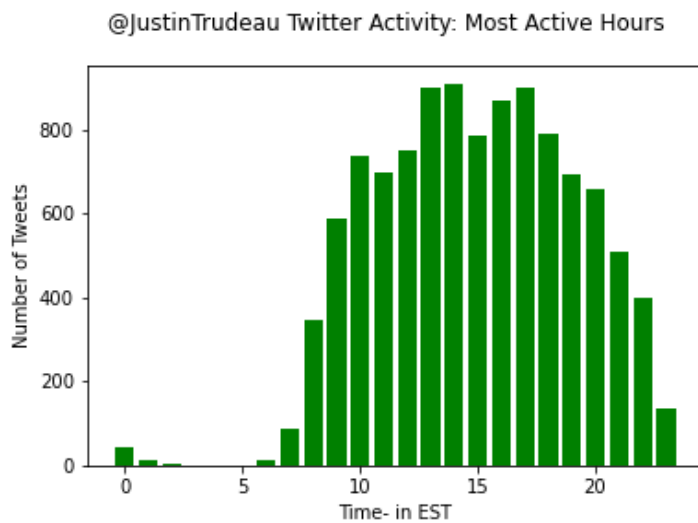
```
#Trudeau's twitter data
#Trudeau's tweets were in a CSV called 'Trudeau
tweets'

#splitting up the time from the datetime stamp in the
CSV
trudeau_time = [item.split(" ")[1] for item in
trudeau_tweets['Date'].values]

#isolating the hour and making a list of hours tweeted
trudeau_hour = [item.split(":")[0] for item in
trudeau_time]

#converting UTC to EST for readability
trudeau_hourEST = [(int(item)-4)%24 for item in
trudeau_hour]

#turning the previous list into a dictionary with 'counts'
of how many instances
jthours_tweeted =
collections.Counter(trudeau_hourEST)

#ordering the dictionary so it goes 00 to 23
jthours_tweeted_od =
collections.OrderedDict(sorted(jthours_tweeted.items(
)))

trudeau =
pd.DataFrame.from_dict([jthours_tweeted_od])

#plot of PM Trudeau's Twitter activity by hour
fig3, ax = plt.subplots()
plt.bar(jthours_tweeted_od.keys(),
jthours_tweeted_od.values(), color='g')
plt.suptitle('@JustinTrudeau Twitter Activity: Most
Active Hours')
plt.xlabel("Time- in EST")
plt.ylabel("Number of Tweets")
plt.show()
print('Tweets collected from 11/4/2015 to
11/20/2019')
```

That is where I am at with the data that I have currently- I want to keep working on this project and these numbers, because again- it is interesting to me to see the active times.

Some things I want to work on/expand on in the future with this:

- Grab more twitter activity from different politicians from all ends of the political spectrum- for this, I stuck with 'world leaders' to make it 'fair' and have an easy comparison of dates (i.e., 1 term), but I think it would be interesting to see if conservative leaders tweet more/different hours then liberals or vice versa, etc., does the age of the politician play into it-are younger, 'hipper' politicians more active on social media then the older ones, etc. I realize here I have one conservative vs two liberals which is not a whole ton of comparison, so I want to get more even if the office they hold is not part of the comparison any longer.
- I also want to put all the political twitters on one stacked graph to better see the twitter activity quantities vs hours with each other vs having three (or more) separate graphs.
- Have a comparison of the @WhiteHouse twitter, @barackobama, @realDonaldTrump vs gallup/PEW polls- is there more activity at more frequent intervals when polls are low or no correlation?
- Check twitter history over time- have the active hours and times changed since being in office, or stayed the same? Do they trend later or earlier? Has the quantity changed since being in office, do they tweet at the same rate?