
0.0.1 Question 1d

There are many ways we could choose to read tweets. Why might someone be interested in doing data analysis on tweets? Name a kind of person or institution which might be interested in this kind of analysis. Then, give two reasons why a data analysis of tweets might be interesting or useful for them. Answer in 2-3 sentences.

Someone might be interested in doing data analysis on tweets in order to compare, contrast, or read the patterns of someone's tweets. For example, a worker at a political campaign for a certain candidate might use data analysis on another political candidate's tweets in order to undermine any inconsistencies of their political beliefs or words/opinions. Doing data analysis on tweets would also be useful for them in a way that they can compare and contrast different political candidates' views on a certain political topic or issue.

0.0.2 Question 2e

What might we want to investigate further? Write a few sentences below.

We might want to investigate why all AOC, Christiano, and Elon Musk all tweet mainly from iPhone. We can investigate whether it is inconvenient for users to use Twitter on Android, or if it was simply due to the three people owning an iPhone. We can also investigate why Christiano tweeted from other sources other than iPhone.

0.0.3 Question 2f

We just looked at the top 5 most commonly used devices for each user. However, we used the number of tweets as a measure, when it might be better to compare these distributions by comparing *proportions* of tweets. Why might proportions of tweets be better measures than numbers of tweets?

Proportions of tweets may be better measures than numbers of tweets in order to determine the most commonly used devices for each user because it helps us better view the data in the perspective of how many tweets someone has tweeted from a certain device over how often/many times they tweet. For example, AOT has tweeted a lot of tweets in general, so it might look skewed if the person who tweets often primarily used one device and many other people who don't use that device primarily use a different device to tweet. It gets rid of potential mistakes we can make while reading and analyzing the data.

0.0.4 Question 3b

Compare Cristiano's distribution with those of AOC and Elon Musk. In particular, compare the distributions before and after Hour 6. What differences did you notice? What might be a possible cause of that? Do the data plotted above seem reasonable?

Compared to AOC and Elon Musk, Cristiano barely tweets anything before the Hour 6, and after the Hour 6, Cristiano starts tweeting, while AOC and Elon Musk seem to tweet less tweets. AOC and Elon Musk lives in USA and Cristiano lives in Europe, so that might be the cause of this difference. Also, the difference in their life styles could be another cause of the difference. The data plotted above seems reasonable.

0.0.5 Question 4a

Please score the sentiment of one of the following words, using your own personal interpretation. No code is required for this question!

- police
- order
- Democrat
- Republican
- gun
- dog
- technology
- TikTok
- security
- face-mask
- science
- climate change
- vaccine

What score did you give it and why? Can you think of a situation in which this word would carry the opposite sentiment to the one you've just assigned?

I give the word 'gun' a score of -3. The word 'gun' have a very negative sentiment because most of the times where gun is involved, it is about crime, death, and injury. Guns have caused many deaths and violences throughout history and were rarely used in anything ethical or just, so it receives a score of -3 from me. A situation where a gun would carry an opposite sentiment to the one I've assigned would be where the word 'gun' was used in 'gun control law passed' or a gun control nearrative.

0.0.6 Question 4g

When grouping by mentions and aggregating the polarity of the tweets, what aggregation function should we use? What might be one drawback of using the mean?

We should use the mean function in order to group by mentions and aggregating the polarity of tweets. We should use mean in order to get the average from the tweet data. One drawback of using the mean might be potential outliers that might skew our data.

0.0.7 Question 5a

Use this space to put your EDA code.

```
In [184]: #dig deeper into when devices were used
tweets['Cristiano'][['device']]
tweets['Cristiano'][['converted_hour']]
which_device_when = (tweets["Cristiano"][["converted_hour", "device"]].groupby("device")
                    .mean().sort_values('converted_hour').rename(columns={'converted_hour':
which_device_when
```

```
Out[184]:
```

device	tweeted_time_avg
Crowdfire Inc.	8.000833
Twitter for Websites	10.540278
Google	13.482639
swonkie	14.039120
Facebook	14.511806
Players' Post	15.005328
MobioINsider.com	15.373169
Twitter Web Client	15.540183
Viva Ronaldo	15.595133
Twitter for BlackBerry®	15.614444
Twitter for Android	15.716615
WhoSay	15.852240
Instagram	15.963715
10 - Sport Through Your Lens	16.079236
Twitter Media Studio	16.196217
Twitter for iPhone	16.278526
Twitter for iPad	16.306135
Twitter Web App	17.479980
iOS	21.116667

0.0.8 Question 5b

Use this space to put your EDA description.

My EDA displays the average times the tweets by Cristiano were posted for different devices. Based on the data table displayed from the code above, we can see that no matter what device Cristiano used to tweet his tweets, most of them average out to be between the Hours 14-17. This shows the prime time that Cristiano likes to tweet his tweets, which gazes our focuses on the anomolies such as iOS (with the average time 21), Crowdfire Inc. (with the average time 8), and Twitter for Websites (with the average time 10.5). These times differ from most other devices by a lot of hours, which shows that Cristiano's uses of these devices were very different than his 'normal patterns'. This data also suggests and raises new questions, such as, "What were the special occasions where Cristiano used these devices to tweet his tweets?"

