

## The Truth Behind Donald Trump's Twitter Feed

### Project 2 Summer 2018 [W18]

Cait

#### Context

Our team analyzed the twitter feed of @realDonaldTrump to draw conclusions about the current US President's tweet behavior. This project demonstrated the process of joining and cleaning up multiple points of data from the websites twitter.com and ds100.org. We used user-friendly numerical processors from Python packages, like Pandas and NumPy. After analyzing the data set, we made conclusions about the characteristics of Trump tweets and the devices used for the tweets and so on. Furthermore, we expressed these conclusions in charts, images, and figures.

#### Data Scope

Initially, we downloaded Donald Trump's recent tweets (3203 tweets) using public APIs from the Twitter Developer platform, then we retrieved the older tweets (5210) from the website [www.ds100.org/sp18/assets/datasets/old\\_trump\\_tweets.json.zip](http://www.ds100.org/sp18/assets/datasets/old_trump_tweets.json.zip). Then, we performed a quick analysis after saving it. This came out to a total of 8257 tweets.

Name	trump_tweets	old_trump_tweets	all_tweets
Type	json → list	json → list	list + list → list
Shape	3203 Rows x 25 Columns	5210 Rows x 25 Columns	8257 Rows x 25 Columns

Figure 1A.

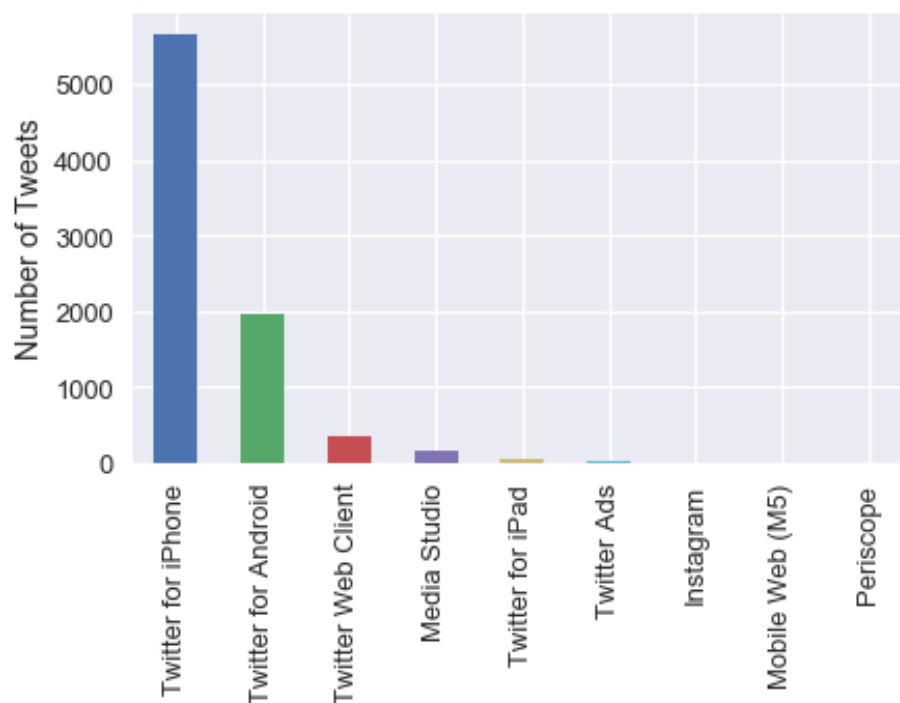


Figure 1B.

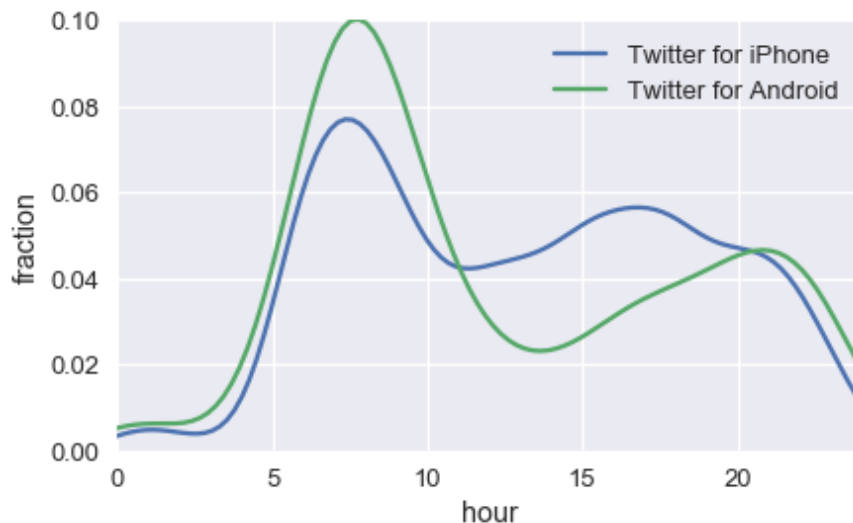
Source	Count
Twitter for iPhone	5668
Twitter for Android	1982
Twitter Web Client	364
Media Studio	150
Twitter for iPad	55
Twitter Ads	34
Instagram	2
Mobile Web (M5)	1
Periscope	1

## Data Analysis

We observed in the Figure 1A bar graph that the two device types that are most commonly used were iPhone and Android. This means 68.6% of the time he used an app downloaded from the App Store and 24% of the time he used an app downloaded from the Google Play Store. We analyzed his twitter behavior across these devices by exploring the time and sentiments of his tweets.

## Data Exploration

First, we looked to see if Trump's tweets from an Android come at different times than his tweets from an iPhone. The initial tweet time was in UTC timezone, and then we converted it into US Eastern Time. We chose EST due to the fact that it was the timezone of New York and Washington D.C., and that is where Trump lives. Therefore, those are the places we expected the most tweet activity from Trump.



**Figure 2.**

We utilized the seaborn distplot function to examine the distribution over hours of the day with this data.

f

**Figure 3.**

We also examined the distribution over time with biannual year increments.

From **Figures 2 and 3** we can visualize the differences of device usage. For instance:

- From 12 am to 11 am (0 hour to 11 hour), the android usage is above apple.
- From 12pm - 8pm iphone is higher.
- From 9pm - 12am
- Started using twitter on android 2015
- Android until 2017
- As president switched to more secure option

\*explain fraction\*

- Trump tends to tweet early in the morning rather than later in the evening
- Personal device android?
- It turns out that in the early years of 2017 Trump stopped using Android to tweet due to safety concerns. His iphone usage has been consistent since 2016 indicating that the majority of his tweets since the election are from the President's staff

### **Tweet Sentiment**

Sentiment analysis is the use of natural language processing techniques to quantify attitudes about a certain topic. Marketers often use sentiment analysis to understand how the public feels toward their brand. Twitter is the perfect data source for this type of analysis because the length of a tweet is capped at 140 characters, just enough for someone to convey a single idea or emotion about a particular subject, but also not too much.

the lexicon contains emojis too! The first column of the lexicon is the *\*token\**, or the word itself. The second column is the *\*polarity\** of the word, or how positive / negative it is.

Now, let's use this lexicon to calculate the overall sentiment for each of Trump's tweets. Here's the basic idea:

For each tweet, find the sentiment of each word.

Calculate the sentiment of each tweet by taking the sum of the sentiments of its words.

First, let's lowercase the text in the tweets since the lexicon is also lowercase. Set the text column of the trump DF to be the lowercase text of each tweet.

## Preliminary Practice Questions

As we familiarize ourselves with the data set, we will perform tests to answer these questions:

- What is the Trump's oldest tweet?
- What month is the oldest tweet from Trump?

Using pandas, we will create a table including the following columns: id of tweets, time the tweet was created, the source device of the tweet, the retweet count of the tweet, we can answer and plot the images of these questions:

- What is the distribution over hours that Trump's tweets on iPhone and Android?
  - During the hours of the morning, Trump tweets more on Android than on iPhone. It turns out that Trump used his personal Android in the morning. He later switched to iPhone during late 2017 due to security issues.
- Are there any striking difference between these distributions?
  - The device distributions of Android and iPhone are inverses of each other. As the iPhone usage increases, the android usage decreases and as the android usage increases the iPhone usage decreases.
- Which device might Trump use to tweet in the morning and then later in the evening?
  - It was found that typically Trump will use his iPhone to tweet in the morning and uses his android to tweet in the afternoon
- Which is the distribution over days that Trump's tweets on iPhone and Android?
- Is there the difference between his Tweet behavior across these devices?
  - There is no difference between his tweets on Android vs iPhone
- Did Trump use only iPhone and Android? If not, it turns out that some of his tweets were not from Trump himself but from his staff? Does the data support the article Take a look at this Verge article.?
  - Trump only used iPhone and Android. His iPhone use is consistent with 2016 indicating that the majority of his tweets since the election are from the President's staff.

Next, we use this tool <https://github.com/cjhutto/vaderSentiment> to analyze the sentiment of Trump's tweets. We need to answer these questions:

- What is the sentiment of a tweet?
  - According to the data collected during our experiment, the sentiment of a tweet, or how the vast populis feels about what is being said varies upon a variety of factors such as the source and buzzwords.
- What is the distribution of tweet sentiments broken down in the text of the tweet containing 'nyt' or 'fox'?

- During the course of our project we were able identify the tweet sentiments depending on the mention of the New York Times and Fox respectively. Tweets that mention the New York Times have an overwhelming negative sentiment while mentions of Fox News have a large positive sentiment. This means that Trump has favoritism toward Fox News.

Finally, we measure the most popular Trump's tweets by answering these questions

- What are the top 20 tweets having the most retweets?
  - Words that slandered his opponent such as "crooked Hillary" and Trump's comments on the media's lack of attention to the information revealed in the wikileaks hack, particularly the words "dishonest media" and "so rigged"
- Were there certain words that often led to more retweets?
  - *During our project we discovered that certain words did indeed have an affect on the number of retweets. Word such as illegally, lowest, jong, obstruction, FBI lead to more retweets.*

