Tweeter Data Analysis

MOHAMED OLOU

- For this analysis we had to use 3 dataset: image-predictions.tsv, tweet_json.txt and twitter-archive-enhanced.csv

- First, we start by data gathering, and go to data accessing and finally cleaning data. And when we cleaning the data, we try to merge all dataset in one to make a global report about our tweet.

During the analysis we find some problems about the datasets. There was many problems like: non-explicit columns names, useless column, First letter of name non-capitalize, null values, non-good format for column data, column with html code, non-descriptive column data.
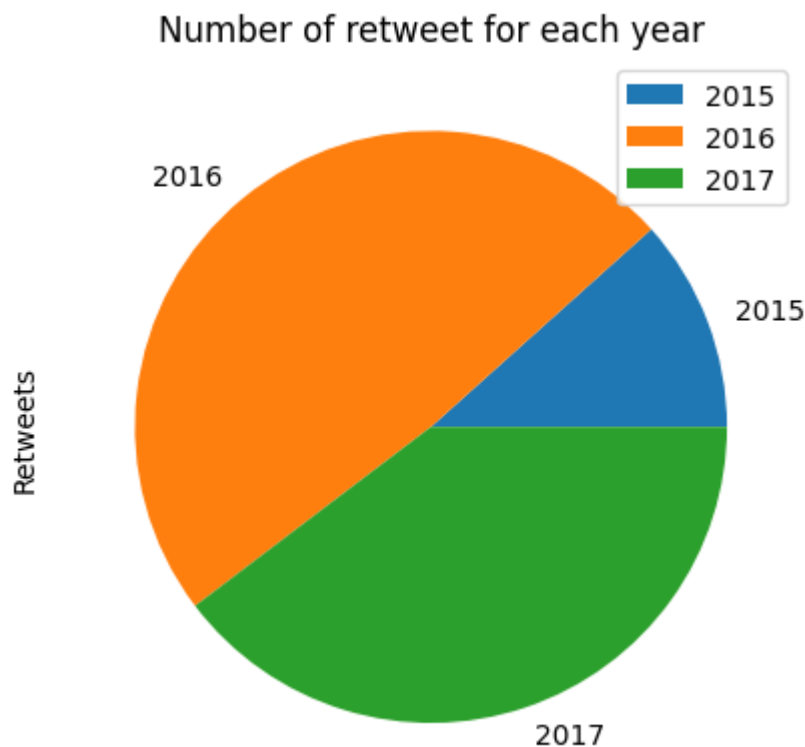
To solve all the problems, we start to clean datasets we drop some columns, we drop some columns, merge some columns, etc.

After all these analysis, we observed many thing.

Finally we have 2073 tweets.

RETWEETS

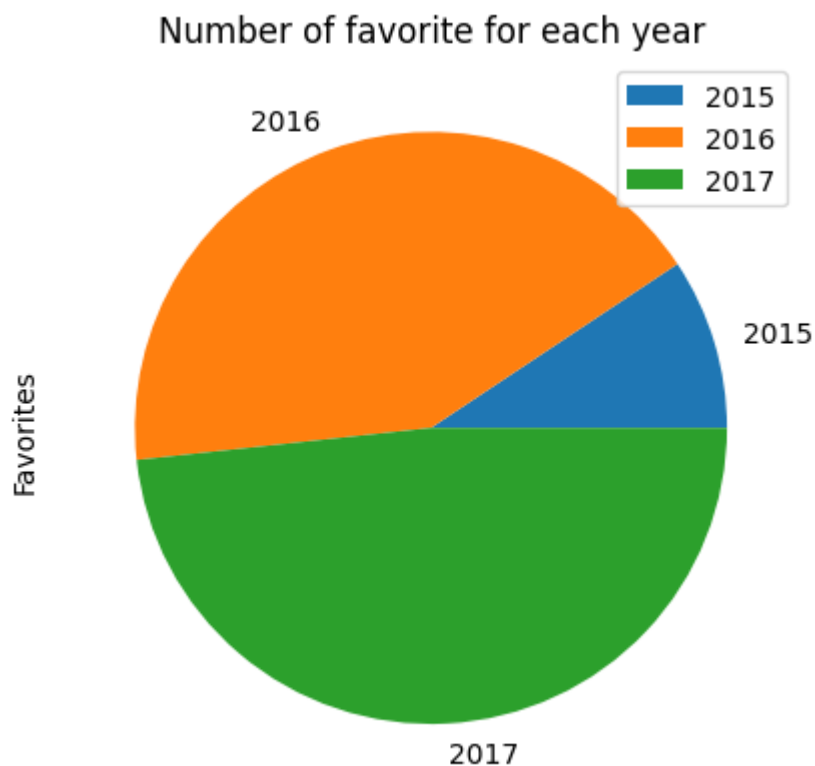On these 2073 we finally observed that there was more retweet in 2016.

Number of retweet for each year

This is the ranking of the years with the most retweets:

- 2016
- 2017
- 2015

FAVORITES

On these 2073 we finally observed that there was more retweet in 2016.

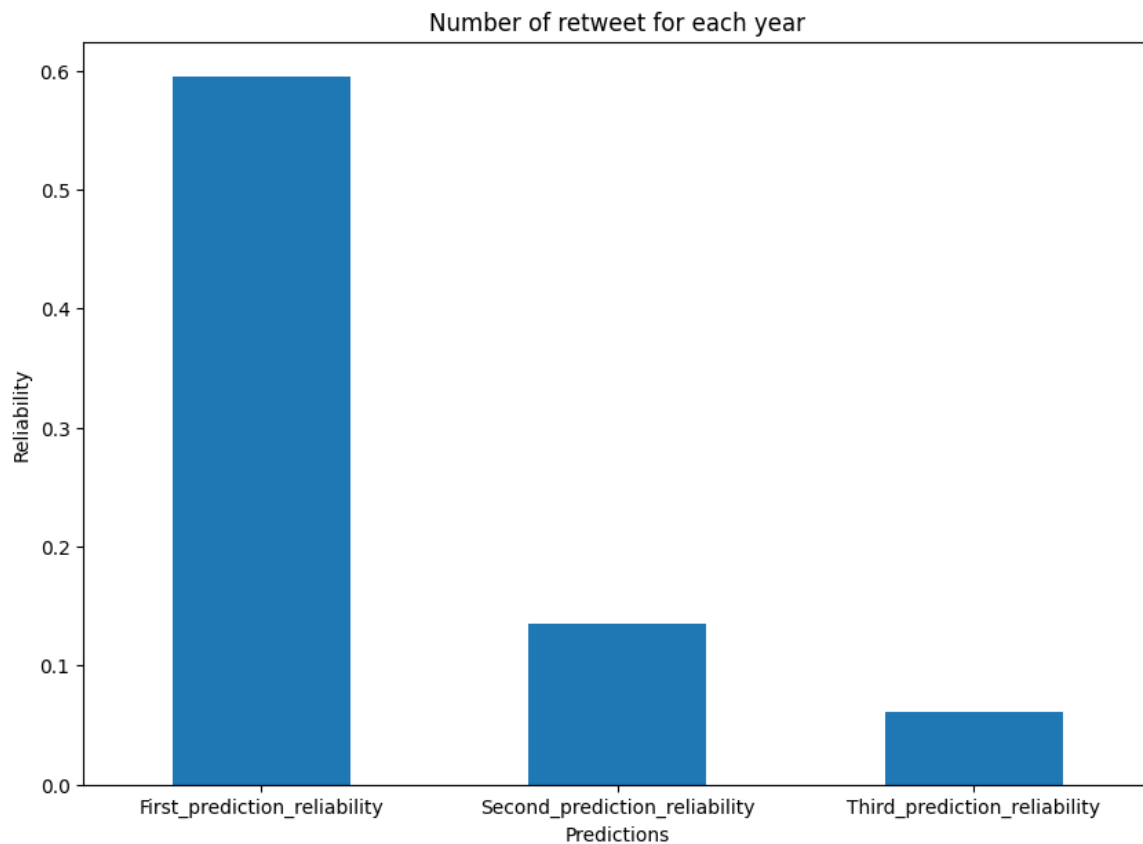Number of favorite for each year

This is the ranking of the years with the most retweets:

- 2017
- 2016
- 2015

PREDICTION RELIABILITY:

For each image we have 3 predictions. After getting all cleaned datasets in one we observed that the most reliable prediction is the first.
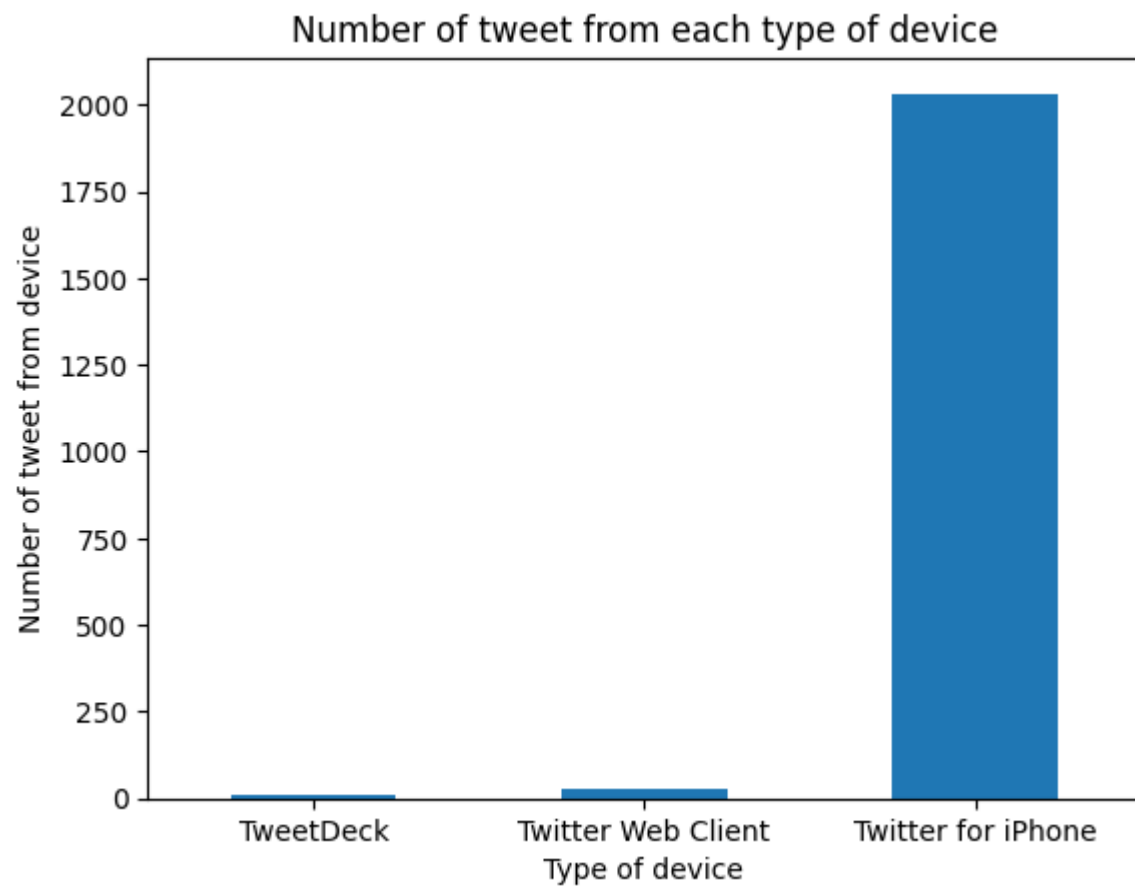
Number of retweet for each year

This is the ranking of the most reliable prediction:

- First prediction
- Second prediction
- Third prediction

TWEET SOURCE:

We also analyze to know which devices is most used to tweet and we observe that the most used device is iPhone.

Number of tweet from each type of device

TWEET FROM WeRateDogs

WeRateDogs® ✔
@dog_rates

This is Phineas. He's a mystical boy. Only ever appears in the hole of a donut. 13/10

5:23 PM · Aug 1, 2017 · Twitter for iPhone

6,955 Retweets **219** Quote Tweets **33.6K** Likes