

ACTION REPORT

This report includes the actions that were carried out as a result of doing this project.

The first action is to import all the necessary libraries, gather the csv files using `pd.read_csv`, and name them with easy names like `image`, `twitter` and `tweet_json`. Some issues were detected and the following actions were taken to clean the messy data:

- Cleaning of columns such as those that contain missing values
- Removal of all second and third predictions
- Changing the timestamp to datetime
- Cleaning source column from complex names to simple names for the purpose of analysis
- Dropping retweet columns because they are not necessary
- Correcting incorrect names
- Changing 'None' string values to NaN in Dog stage column
- Converting `img_num` and `p1_conf` to string
- Put `doggo`, `puppo`, `floofer` and `pupper` as categorical variable with variable name `dog_stage`
- Merge all necessary dataframes to have a more concise dataframe .

After all the cleaning, a dataframe was formed which was later visualized

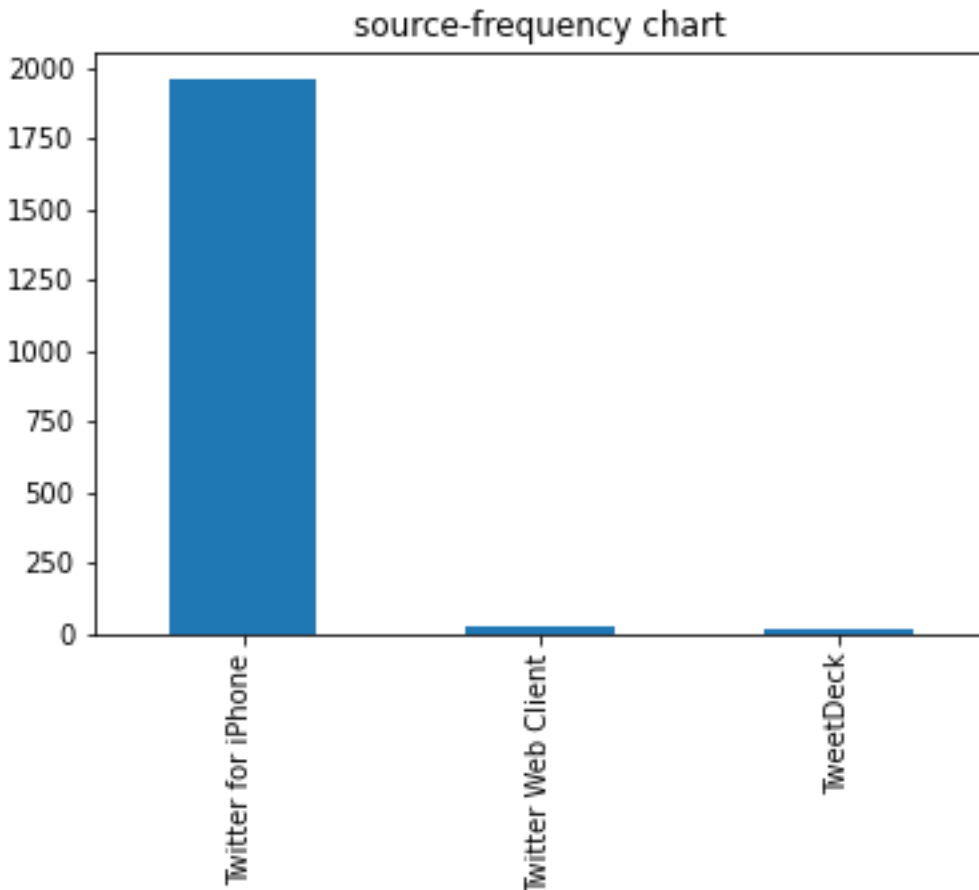


Figure 1

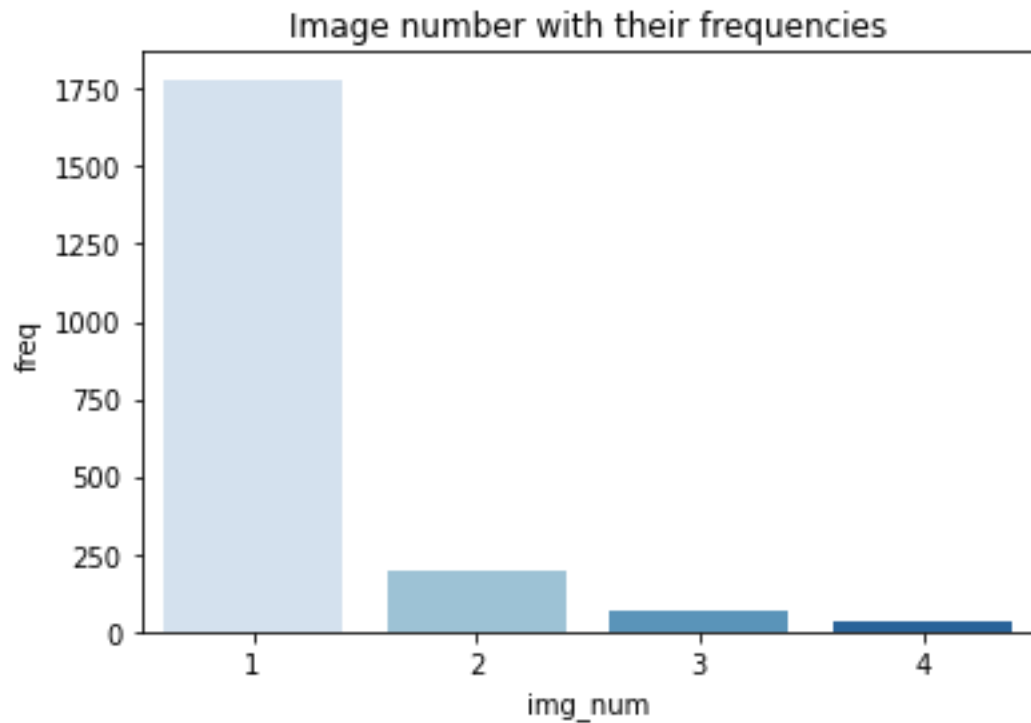


Figure 2

Number of Prediction based on the year 2015, 2016 and 2017

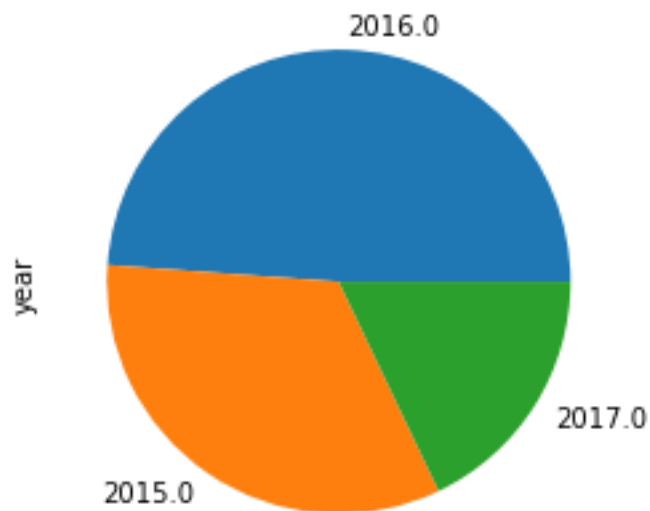


Figure 3

From the graphs above, the following insights were obtained:

Insight 1: From figure 1 above, it can be seen that data was gotten most from 'twitter for iPhone' which may suggest also that most participants are iPhone users

Insight 2: From figure 2 above, it can be established that the img_num 1 has highest frequency. This means that the image 1 has the largest population from the sources. We can also see here that it is possible that dog with image number 1 are more popular which would be an assumption.

Insight 3: From figure 3 above, it can be seen that most data are from 2016