UDACITY

2021

Data Analyst Nanodegree Program

Project 3: Wrangel and Analyzed Data

REFAH M. ALDOSSARY

Act Report

Report Objective:

This report communicates the insights and displays the visualizations produced from the wrangling process.

Analysis and Findings:

I started checking the dataset info and statistics using .info() and .describe() which result in the following:

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2018 entries, 0 to 2017
Data columns (total 6 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   tweet_id        2018 non-null   object
 1   created_at      2018 non-null   datetime64[ns, UTC]
 2   retweet_count   2018 non-null   int64
 3   favorite_count  2018 non-null   int64
 4   stages_of_dogs  316 non-null    object
 5   source          2018 non-null   object
dtypes: datetime64[ns, UTC](1), int64(2), object(3)
memory usage: 110.4+ KB
```

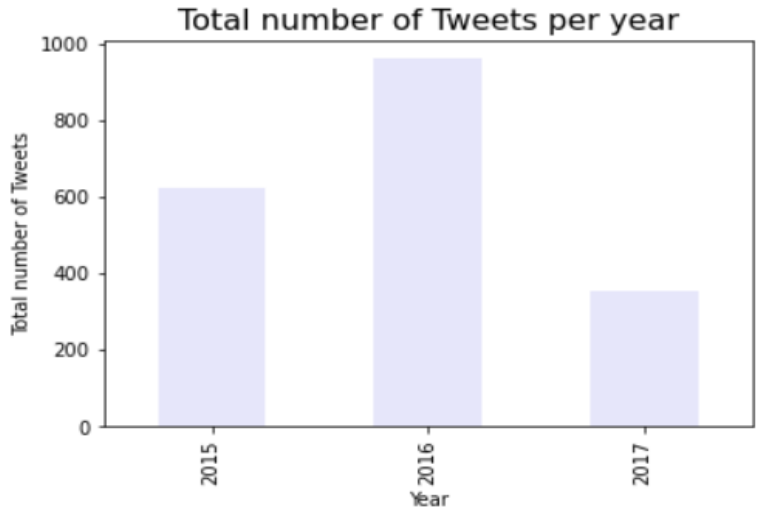
	retweet_count	favorite_count
count	2018.000000	2018.000000
mean	3028.937066	8708.240337
std	5098.711145	12171.987523
min	16.000000	0.000000
25%	661.000000	1784.250000
50%	1448.000000	3972.500000
75%	3517.500000	11181.500000
max	79515.000000	132810.000000

Looking at the above information, I selected only the data fields that I want to investigate more and draw conclusions from. We can notice that the average retweet counts is around 3028.94 while the average favorite counts is approximately 8708.24. Also, notice that we have tweets with 0 favorite which is normal. However, for better visualization, I filtered all tweets with favorite_count that equals 0 using the following code:

```
#removing favorite_count = 0 for better visulization:
final_tweets_data=final_tweets_data[final_tweets_data['favorite_count']!= 0]
```

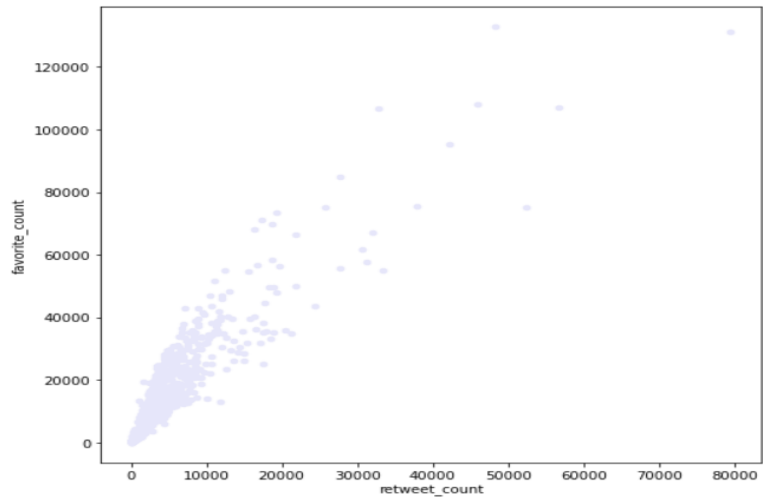
Visualization and Insights:

1. Number of tweets per year:



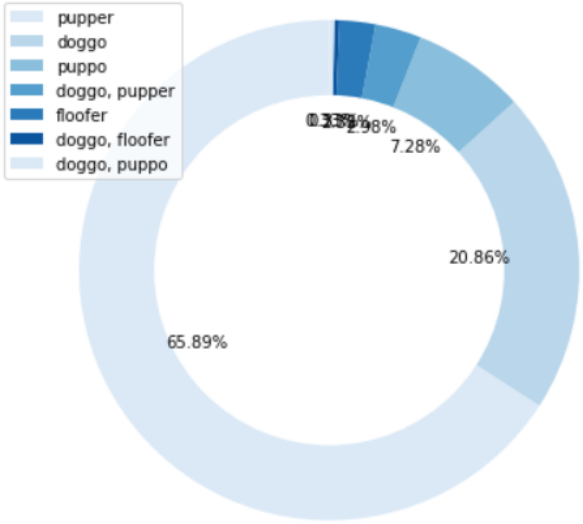
The above bar chart shows the total number of tweets per year. As the graph clearly communicates, the highest number of Tweets was in 2016.

2. The relationship between favorites and retweets:



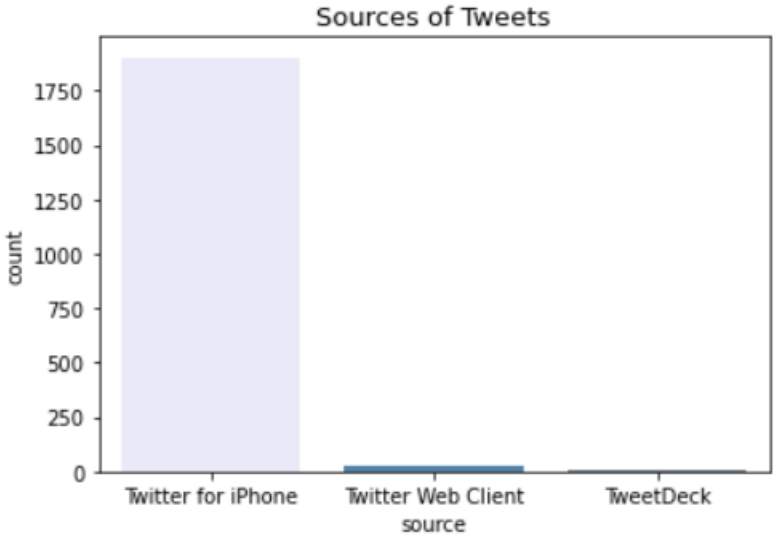
The above scatter plot shows the relationship between favorite counts and retweet counts. As the graph clearly communicates, the relationship between favorites and retweets is positive, meaning that as the number of retweets increase, the number of favorites increase as well.

3. The most common stage of doges:



The above doughnut chart shows the different dogs stages and highlights the most common stage which is pupper with 65.89% of total stages.

4. The most used source to tweet is iphone



The above bar chart shows the different sources of tweets used by users. As it clearly presents that the mostly used source to tweet by users is the iphone.