Wrangling and Analysis Data

(Act Report)

WeRateDogs™

@dog_rates

Your Only Source For Professional Dog Ratings IG, FB, Snapchat ⇨ WeRateDogs partnerships@weratedogs.com

DM YOUR DOGS

weratedogs.com

Joined November 2015

Introduction

This project is about a twitter account "WeRateDogs" it is a twitter account people send them pictures of their dogs to rate them, it just humorous ratings and comments about the dogs. The tweets might contain the name of the dog, his hobbies and his picture, the picture is most important to rate the dog. These ratings almost always have a denominator of 10 and the numerators Almost always greater than 10. 11/10, 12/10, 13/10, etc.

The aim of the project to applying the data wrangling processes which are gathering, assessing, and cleaning data. After that analyzing and visualizing it, using python packages and libraries in Jupyter Notebook environment.

A brief description of data assessing and cleaning in data wrangling:

Data Assessing: In this phase, we are assessing data quality and lack of tidiness, by discover the issues in datasets by using pandas’ functions.

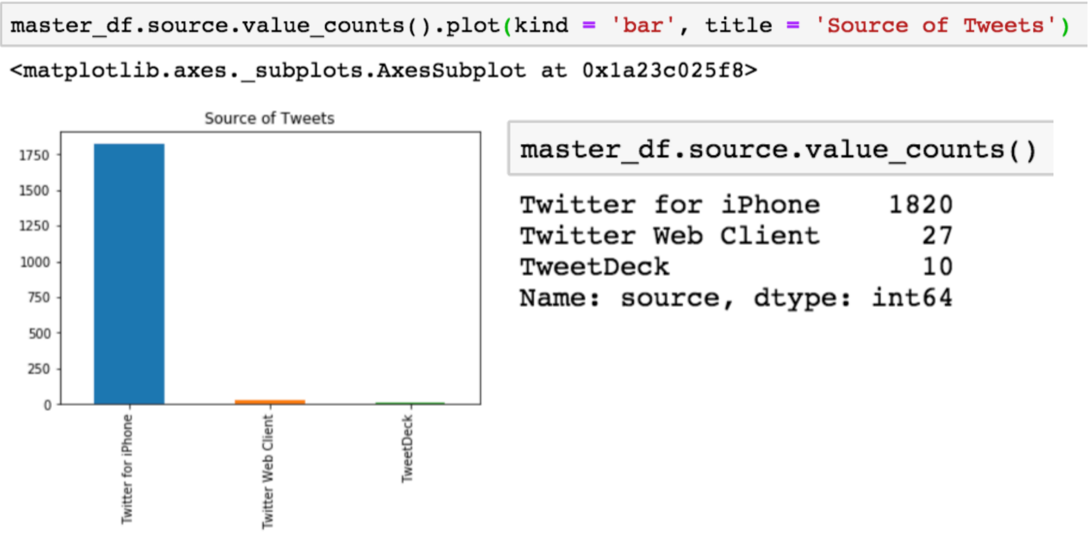
Data Cleaning: The actual changes on the dataset.

Analyzing and Visualizing

The shape of data frame after cleaning is 1857 entries and 22 columns.

Insight #1:

- 98% of tweets are from Twitter application on iPhone smartphone that means that means it rarely tweets from Web Client or TweetDeck.



```
master_df.source.value_counts().plot(kind = 'bar', title = 'Source of Tweets')
```

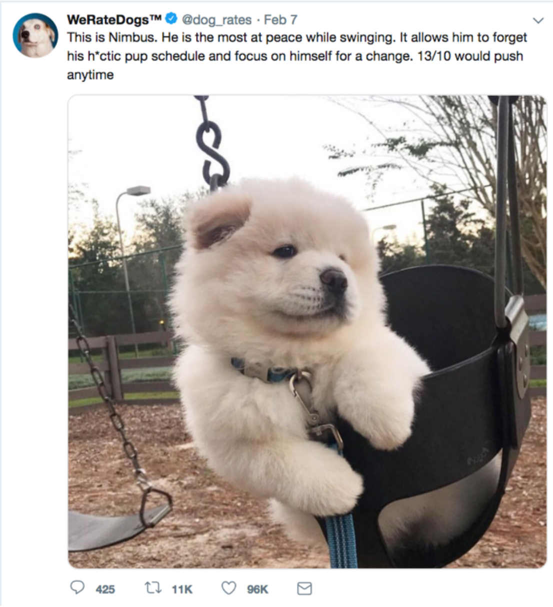
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a23c025f8>
```

```
master_df.source.value_counts()
```

```
Twitter for iPhone     1820
Twitter Web Client       27
TweetDeck                10
Name: source, dtype: int64
```

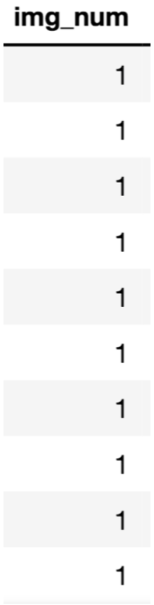
Insight #2:

- Most tweets contain one image approximately 85% of tweets.

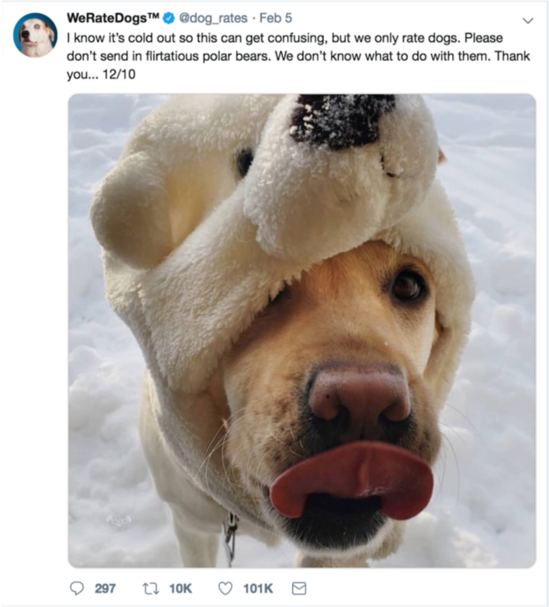


WeRateDogs™ @dog_rates · Feb 7

This is Nimbus. He is the most at peace while swinging. It allows him to forget his h*ctic pup schedule and focus on himself for a change. 13/10 would push anytime



img_num
1
1
1
1
1
1
1
1
1
1

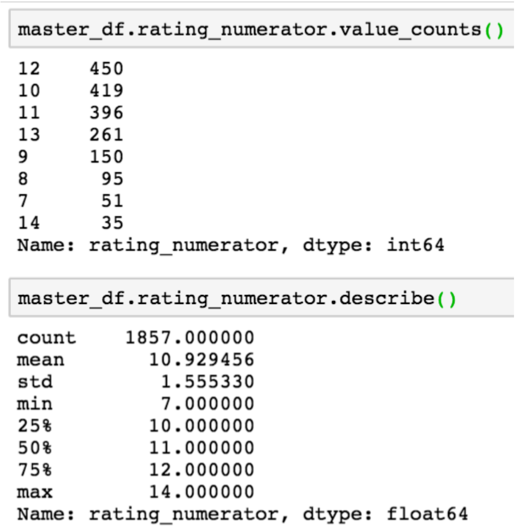


WeRateDogs™ @dog_rates · Feb 5

I know it's cold out so this can get confusing, but we only rate dogs. Please don't send in flirtatious polar bears. We don't know what to do with them. Thank you... 12/10

Insight #3:

- Average ratings between 10 and 11 and the most tweets rated 12.

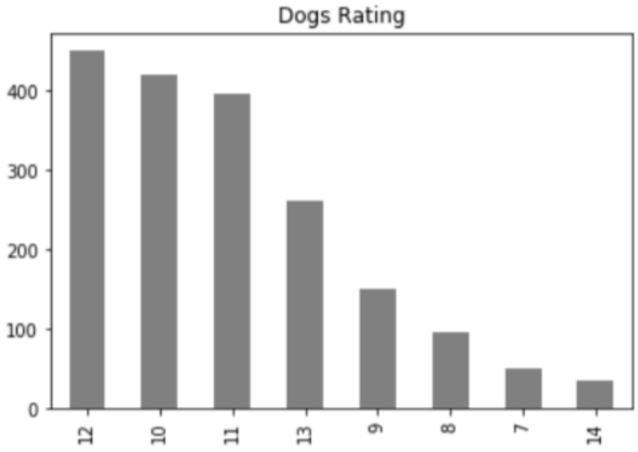


```
master_df.rating_numerator.value_counts()
```

```
12    450
10    419
11    396
13    261
9     150
8      95
7      51
14     35
Name: rating_numerator, dtype: int64
```

```
master_df.rating_numerator.describe()
```

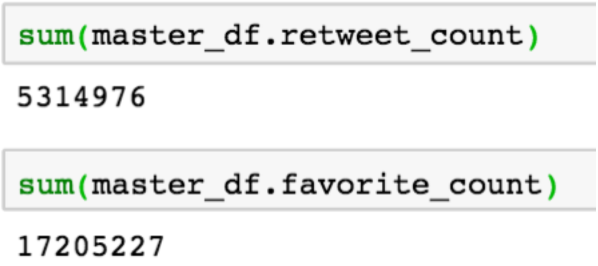
```
count    1857.000000
mean       10.929456
std         1.555330
min         7.000000
25%        10.000000
50%        11.000000
75%        12.000000
max        14.000000
Name: rating_numerator, dtype: float64
```



Insight #4:

- Followers interact in the account by favorite tweets more than retweet it.

Retweet < Favorite



```
sum(master_df.retweet_count)
```

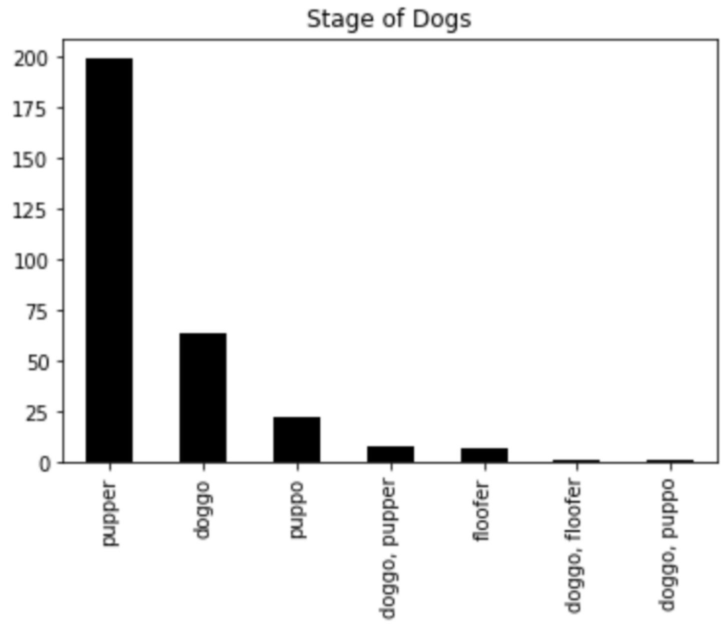
```
5314976
```

```
sum(master_df.favorite_count)
```

```
17205227
```

Insight #5:

- pupper breed are the most popular which are 68% of dogs are pupper.


Stage of Dogs
pupper
doggo
puppo
doggo, pupper
floofer
doggo, floofer
doggo, puppo