

Data Wrangling & Analysis

Introduction

Real-world data rarely comes clean.

Our goal: wrangling WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations. The dataset that we will be wrangling (and analyzing and visualizing) is the tweet archive of Twitter user @dog_rates, also known as WeRateDogs.

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10.

The Twitter archive only contains very basic tweet information. Using Python and its libraries, we will gather data from a variety of sources and in a variety of formats, assess its quality and tidiness, then clean it

Quality

1- Delete columns that won't be used for analysis

2-unnecessary html tags in source column in place of utility name e.g. <a href=""<http://twitter.com/download/iphone>" rel=""nofollow"">Twitter for iPhone

3-extract the rating numerator and rating denominator from the text and Rating denominator should have 10 .

4-Delete unreasonable rate rows at the column that has the value of (rating_numerator / rating_denominator) 'rate'

5-Remove outlier rating.

6-Remove all un-original tweets (retweets).

7-Change datatype of tweet_id column to a string and Change datatype timestamp column to data time.

8-Change missing values in 'name' from 'None' to NaN (dog stages already covered).

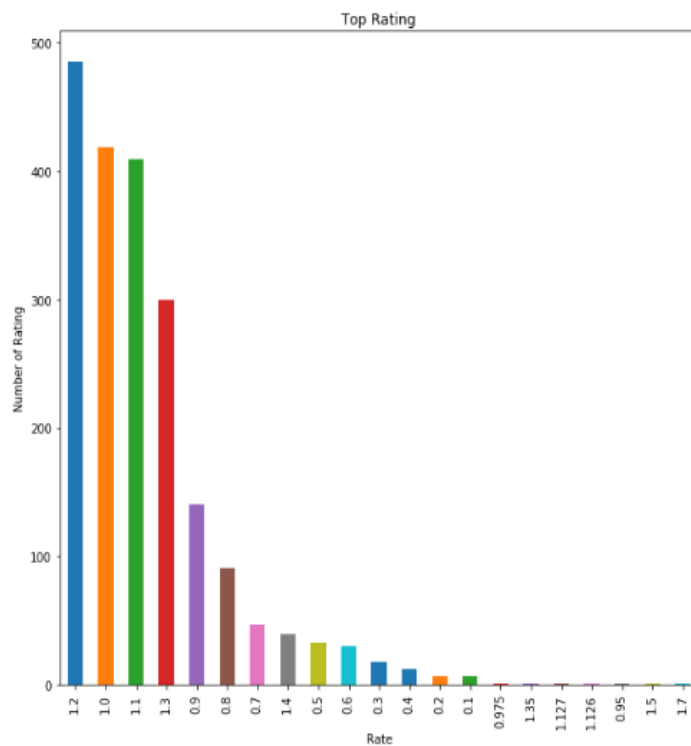
9-Remove uncorrct names

Tidiness

1-Dog "stage" variable in four columns: doggo, floofer, pupper, puppo

2-Join 'json_tweets' and 'arc_df'

Visualization



After a lot of work and base on plot the conclusion is the most rate choose from tweets is 1.2 and 1 .

Analyses

After a lot of work and base on analyses the conclusion is :

the most common name is **Charlie , Lucy , Cooper, Oliver , Penny**

the most retweet on tweet is 79515 it was in stage doggo in 18/06/2016.

the most favorite on tweet is 132810 it was in stage puppo in 21/01/2017.

Number of dog has stage pupper is 213 and this is most stage .

11 dog has two stage like doggo , pupper or doggo ,floofer