



REEM ALRASHOUD

July , 2020

ACT REPORT

Project#5: Wrangle and Analyze Data

Data analyst nanodegree program

▪ Introduction:

This project points to urge quick data approximately puppy evaluations whereas illustrating progressed information wrangling and visualization strategies utilizing different Python libraries. I will assemble information from a assortment of sources and in a assortment of designs, evaluate its quality and tidiness, at that point clean it. Typically called information wrangling. The dataset that I will be wrangling (and analyzing and visualizing) is the tweet archive of Twitter user @dog_rates, moreover known as WeRateDogs.

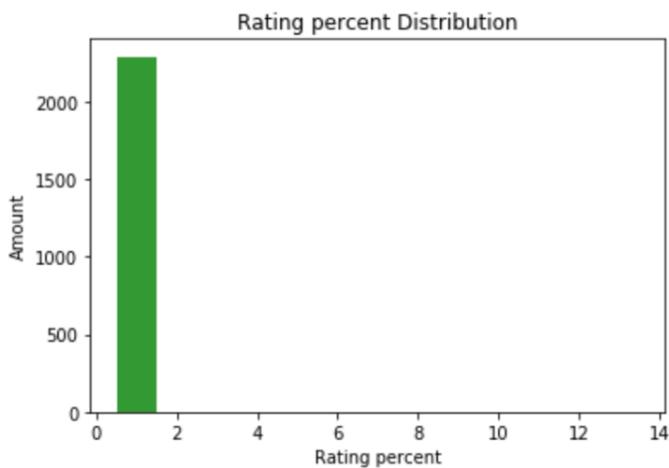
WeRateDogs may be a Twitter account that rates people's mutts with a amusing comment approximately the canine. The page has since developed greatly in popularity , with numerous clients sharing its substance and asking their pooches being appraised aswell. These appraisals nearly continuously have a denominator of 10. in spite of the fact that? Nearly continuously more prominent than 10. Since "they're great dogs."

Note: I did not use TWITTER API because they Couldnt not support me with an access to developer account.

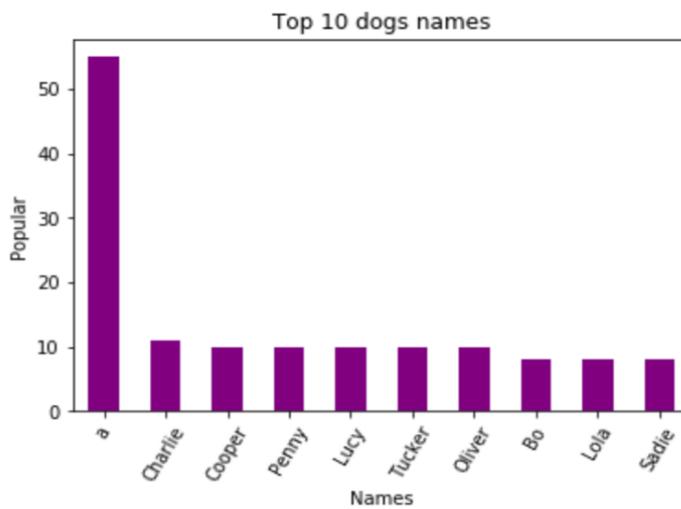
▪ Insights:

- 1- Display the number of entries and columns in the dataset.
- 2- Show the top 10 represented dog names in the dataset. To find the most 10 popular dog's names.
- 3- Shows descriptive information about timestamp. It helps to retrieves information about the time frame of the dataset.
- 4- shows descriptive information about rating_percent. Which is presented as rating_denominator and rating_numerator.
- 5- Show all retweet counts represented in the dataset.
- 6- Show all favorite counts represented in the dataset.

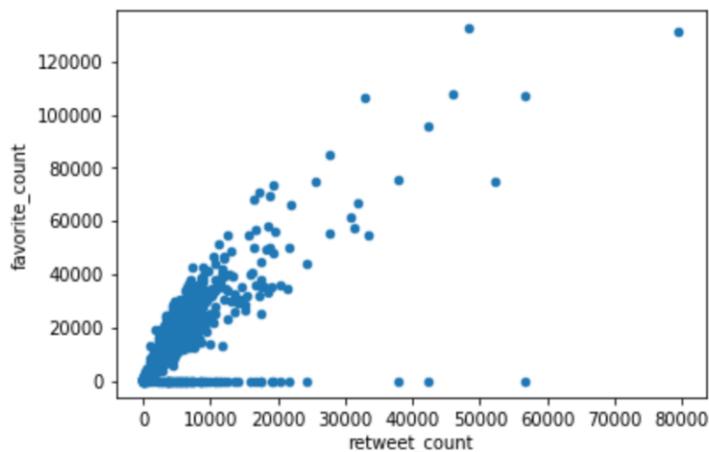
▪ Visualization:



Description: This plot represent distribution of ratings in histogram. As we can see the Rating increase as the Amount increase. we will find the highest rating which is equal 1 with amount above 2000. After that there is no impact between the amount and percent rating.



Description: If we exclude cases, where a letter 'A' is provided, the most popular names are: Charlie, Lucy, Oliver, Cooper, Lola, Penny, Tucker, Sadie and Bo. So we can summarize the most popular 10 dog's names. That can be helpful to make a special studies and test on a specific sample.



Description: This plot represent a relation between retweet and favorite in scatter graph. From this findings we notice the retweet count increase while the favorite count increase too! That means there is a direct correlation between them.