



Report

On
Analysis and Insights into Final Data

By:Potros Shenoda

[Pick the date]

Introduction:

The data came mess and untidy and our job to make it tide and visualize it to make decisions here we have a tweet archive of Twitter user @dog_rates which a Twitter account in which people like and comment on dogs

Gather:-

- The Twitter archive 'twitter_archive_enhanced.csv' This archive contains basic tweet data (tweet ID, timestamp, text, etc.) for all 5000+ of their tweets as they stood on August 1, 2017
- The Tweet image prediction, i.e., what breed dogs (or another object, animal, etc.)
- Twitter API and Python's Tweepy library to gather each tweet's retweet count and favourite ("like") count

Assessing Data:

1. Completeness
2. Validity
3. Accuracy
4. Consistency

And There are three main requirements for tidiness:

1. Each variable forms a column
2. Each observation forms a row
3. Each type of observation unit forms a table

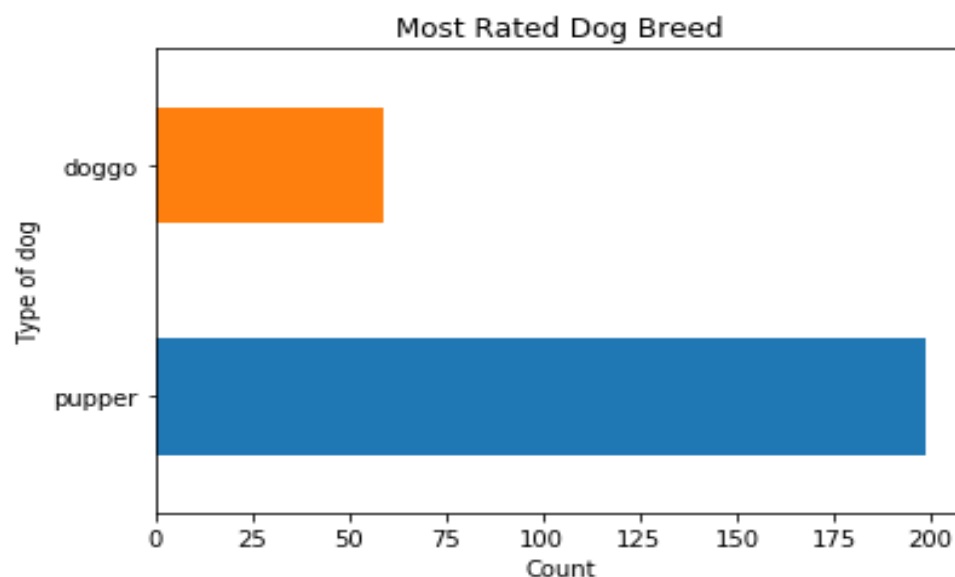
Clean:

- Define
- Code:
- Test:

Analysis and Visualization:

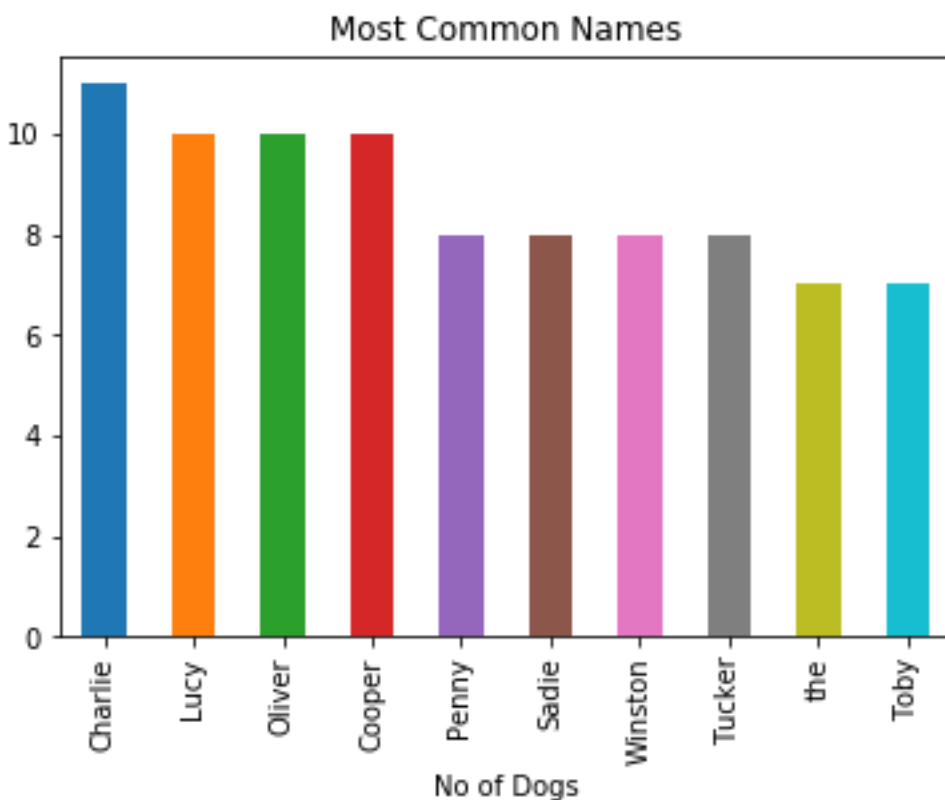
1-Most Rated Dog Breed

In this picture we see that [doggo,pupper] have have values more than 25 and as we see that the [pupper is the most rated dog breed, and this is important to people who will buy dogs.



2-Most Common Names

We see that [Charlie ,Lucy , Oliver , Cooper , Penny , Sadie ,Winston , Tucker , the ,Toby] is the most common names for dogs and as we see that Charlie is the most common name , and this is important to people who will name their dog



3-Tweet Count vs. Retweet Count

There is a direct correlation between favorite (“like”) counts, and the way much a post was retweeted. This correlation is very important for the owner of the WeRateDogs twitter account to know when

determining method to extend users' traffic on the page. a knowledge analysis team could recommend previous posts with either a high retweet counts or high favorite count in order that page owner could model future posts off historically popular posts

