



---

# ANALYSIS AND INSIGHTS INTO FINALDATA

---

By: Eng/Muhammed El-Sayed



JANUARY 31, 2021

## **Gather**

**@WeRateDogs** archive that the admin provided to Udacity in the form of csv file which means comma separated values, The archive provides me with tweets information like id, status(deleted or active), text, retweets, etc for about 2500 tweets that I put them in my environment to do operations with in about 40 minutes, also we got image\_predictions.tsv and tsv stands for tap separated values and it contains the pictures urls to test problems that happened when importing this data because we use a programming language solid in dealing with things, with my twitter API that allows me to check and get what I wanted with a known level access.

## **Assess**

**Well, what my course taught me is to find all issues in my imported data, quality issues and tidiness issues, i.e. (Completeness, Validity, Accuracy, Consistency) and also any structural issues solved.**

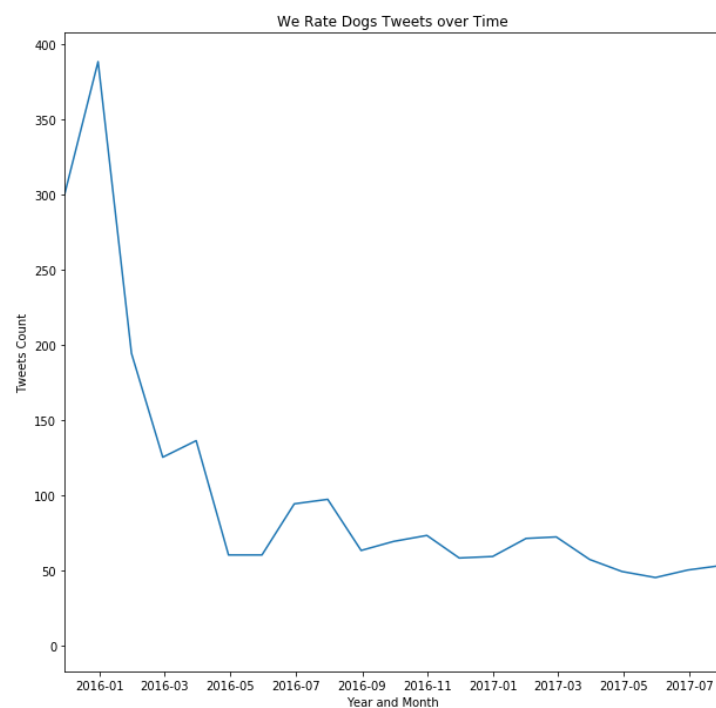
## **Clean**

**Cleaning data is tedious and often iterative. Just when data analyst believe they found all quality and tidiness issue, they often found additional issue arises. The cleaning process involves three steps:**

- 1. Define: Determine exactly what needs to be clean and how.**
- 2. Code: Programmatically clean the code**
- 3. Test: Evaluate the code to ensure the data set was cleaned properly**

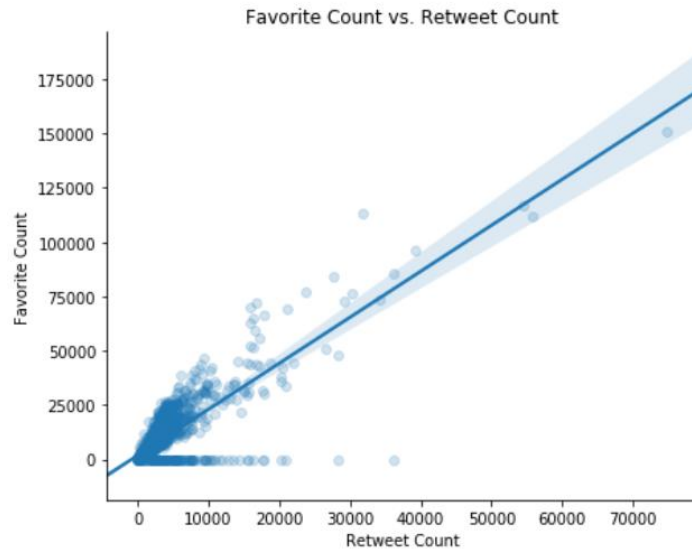
## Analysis and Visualization

### Tweets over time:



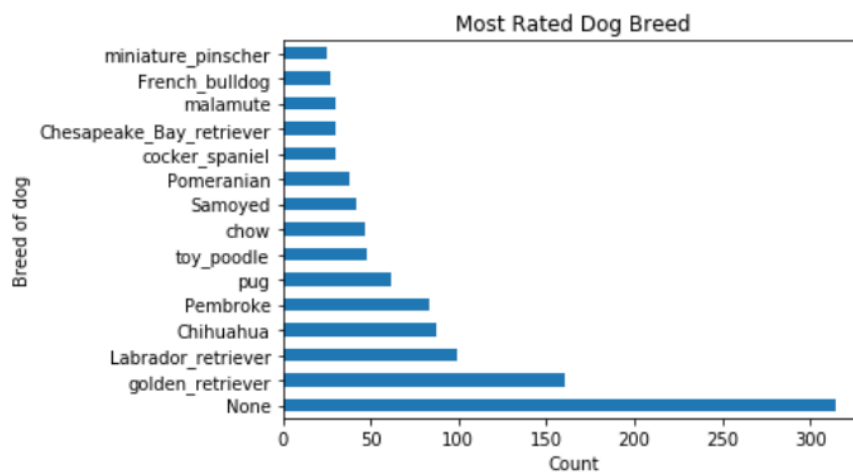
**It obviously decreasing over time.**

### Favorite vs Tweet count:



**As we see in the diagram below it's a positive correlation between favorite count and retweets count.**

## Dog Breed Popularity



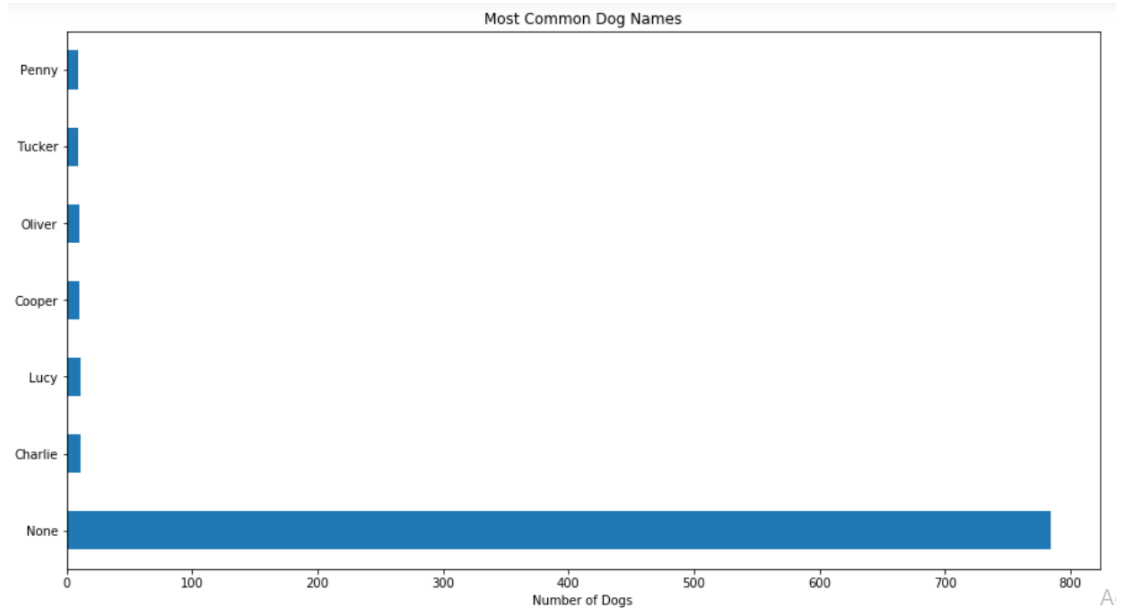
**So the most dog breed are:**

**1.golden retriever**

**2.Labrador\_retriever**

**3.Chihuahua**

## Dog Names



**So most common names are:**

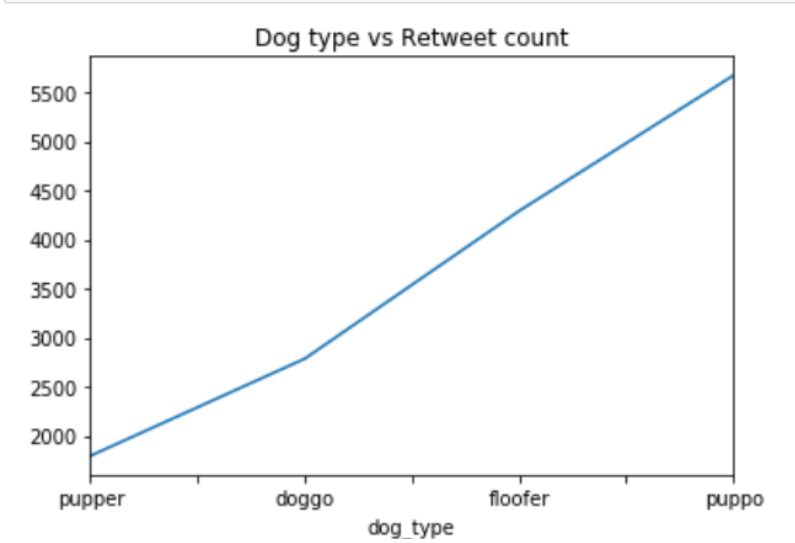
**1-Lucy**

**2-Charlie**

**3-Oliver**

**4-Cooper**

# Dog Types



**Most dog retweet count is puppo.**

---