

# Report on Data Wrangling – We Rate Dogs

-By Shefali Luley

## Documentation of Analysis and Insights into final data

### Introduction

Real-world data rarely comes clean. Using Python and its libraries, I will gather data from a variety of sources and in a variety of formats, assess its quality and tidiness, then clean it. This is called data wrangling.

I will document wrangling efforts in a Jupyter Notebook, plus showcase them through analyses and visualizations using Python (and its libraries) and/or SQL.

The dataset that I 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.

### Steps involved :

- **Gathering**
- **Assessing**
- **Cleaning**
- **Insights and Visualization**

### Insights and Visualization:

Stored the cleaned dataframe in csv file named twitter\_archive\_master.csv

#### Visualizing and Analyzing Data: ¶

```
In [66]: #Import the cleaned data
df = pd.read_csv('twitter_archive_master.csv')
```

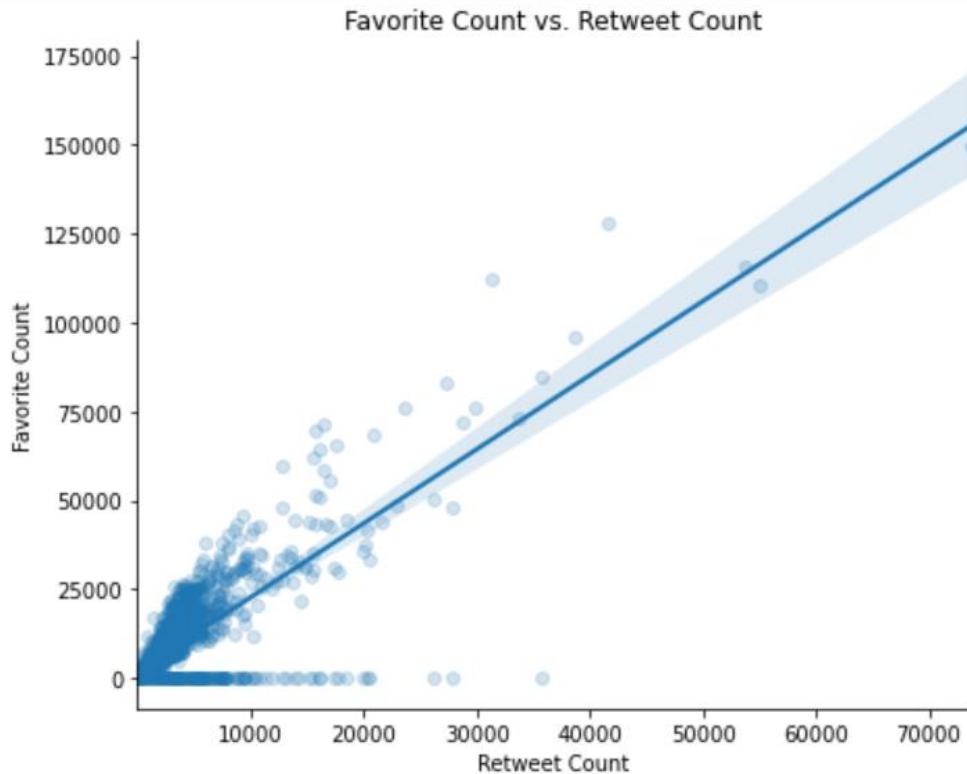
```
In [67]: df.head()
```

```
Out[67]:
```

	Unnamed: 0	tweet_id	timestamp	source	text	expanded_urls	rating_number
0	0	89242064355336193	2017-08-01 16:23:56	<a href="http://twitter.com/download/iphone" r...	This is Phineas. He's a mystical boy. Only eve...	https://twitter.com/dog_rates/status/892420643...	
1	1	892177421306343426	2017-08-01 00:17:27	<a href="http://twitter.com/download/iphone" r...	This is Tilly. She's just checking pup on you....	https://twitter.com/dog_rates/status/892177421...	
2	2	891815181378084864	2017-07-31 00:18:03	<a href="http://twitter.com/download/iphone" r...	This is Archie. He is a rare Norwegian Pouncin...	https://twitter.com/dog_rates/status/891815181...	

## Insight 1

### Visualizing the retweet counts, and favorite counts comparison over time.



### Observation :

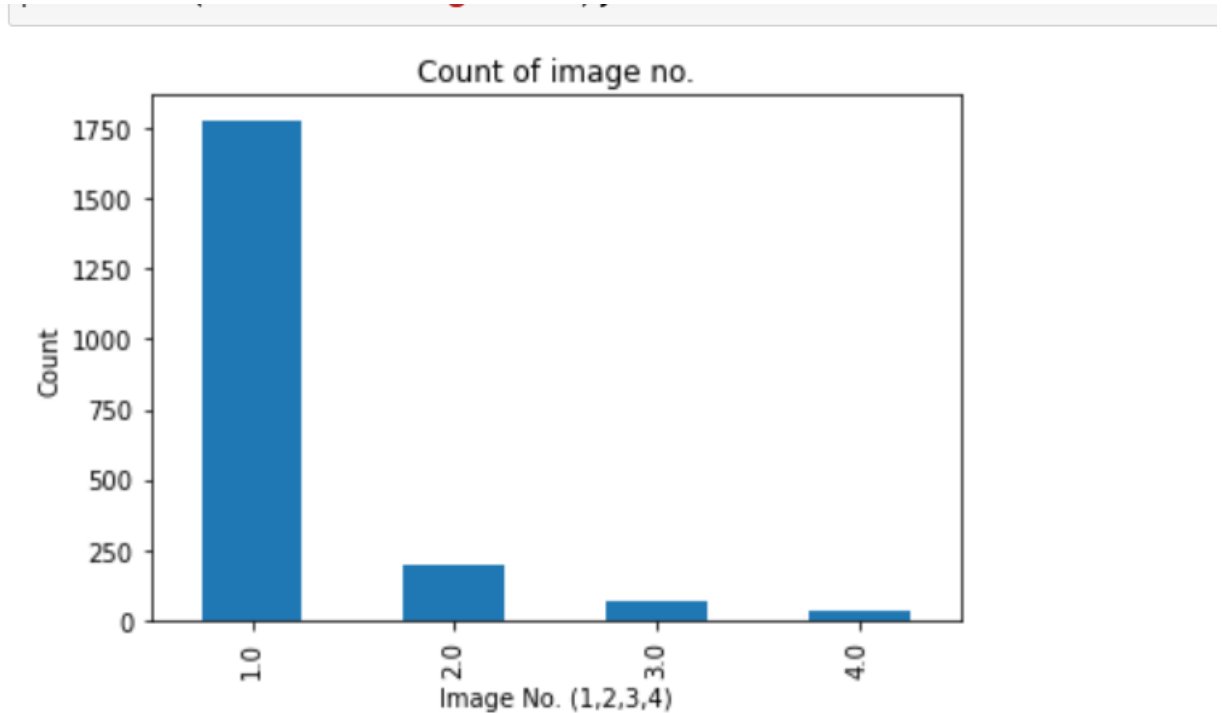
This is a positive correlation, since favorite counts are correlated with retweet counts

### Favorite Count vs Retweet Count

There is a positive correlation between the “like”(favorite) counts, and how much the post was retweeted. The owner needs to understand the correlation of the WerateDogs Twitter account to understand when determining method to increase traffic on the page. A team of data analyst could recommend previous posts with either a high tweet count or high favorite count so that page owner could model future posts.

## Insight 2

### Visualize the count of image number



## Observation

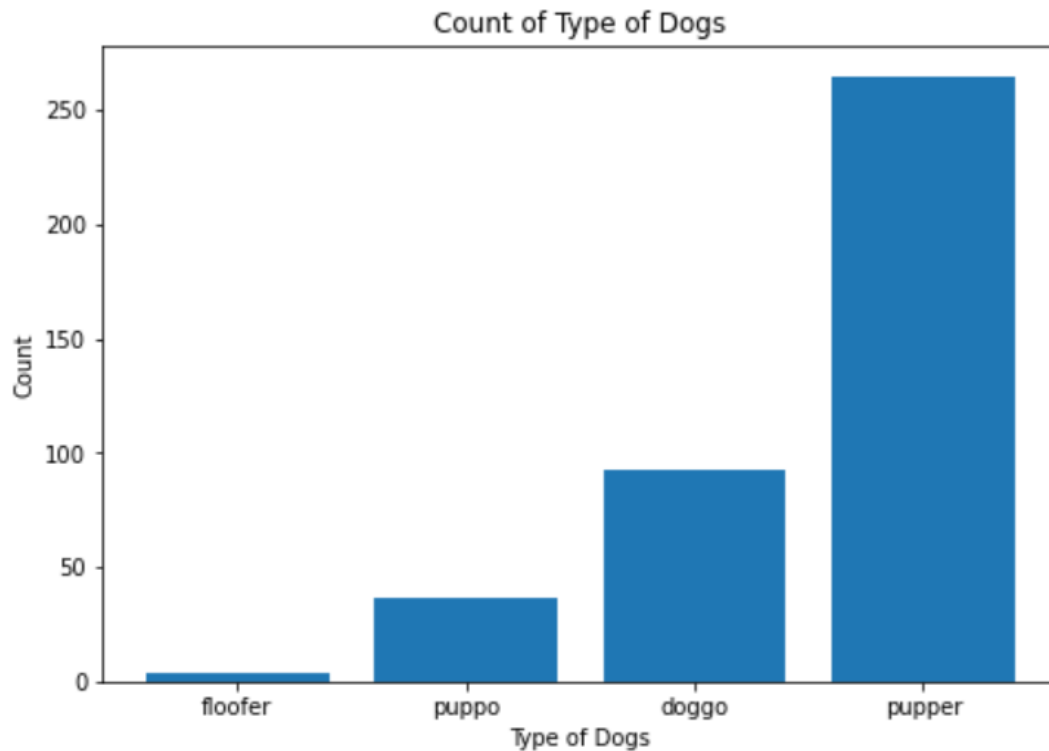
The count of Image number 1 is the highest compared to others

### Count of Image number

From the above representationn, we can observe that the highest count is for image no. 1, while the lowest is for image no. 4 . With the help of this analysis , the owner of WeRateDogs Twitter account will be able to make changes accordingly for more efficiency.

## Insight 3

### Visualize the most common dog type



### Observation :

We can clearly observe that the most famous dog type is " PUPPER" followed by "Doggo" .

### Count of Types of Dogs

The highest count for the types of Dogs is pupper followed by doggo in the second position. The page owner could use this information to create targeted marketing efforts for different types of dogs that aren't popular and can also use the dog type which is already higher to gain more popularity.

## **Conclusion**

**The write up above offer a straight look at the data wrangling process. There is so much more that can be done in this dataset.**