Lizz Mwangi

Data Wrangling

Here's a cheat sheet: the puppers have it!

WeRateDogs is a twitter page that hilariously rates dogs with their funny captions but are just over the top.
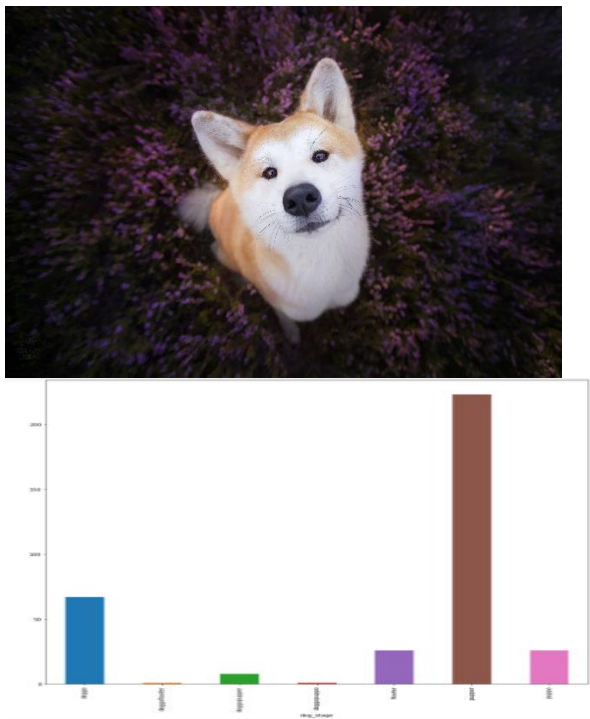
As part of the tasks to complete my second student project at Udacity, I gathered, assessed and cleaned some datasets queried from the this page's Twitter archive.

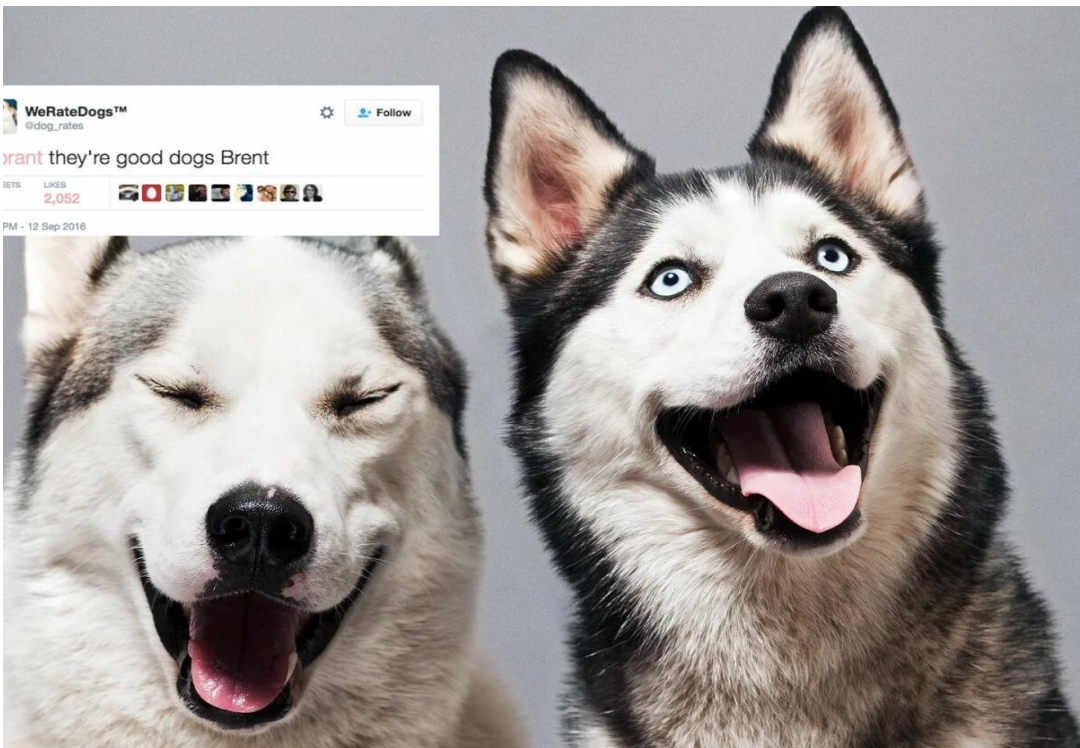Well, these are the 3 main steps of the Data Wrangling Process.

These steps are followed to have a clean and tidy dataset that can be analysed and visualized.

The end goal, you ask? To identify trends and come up with insights which are then used to make decisions.

An image of a dog at doggo dog stage.

This graph above shows the counts of the dogs at every dog stage.
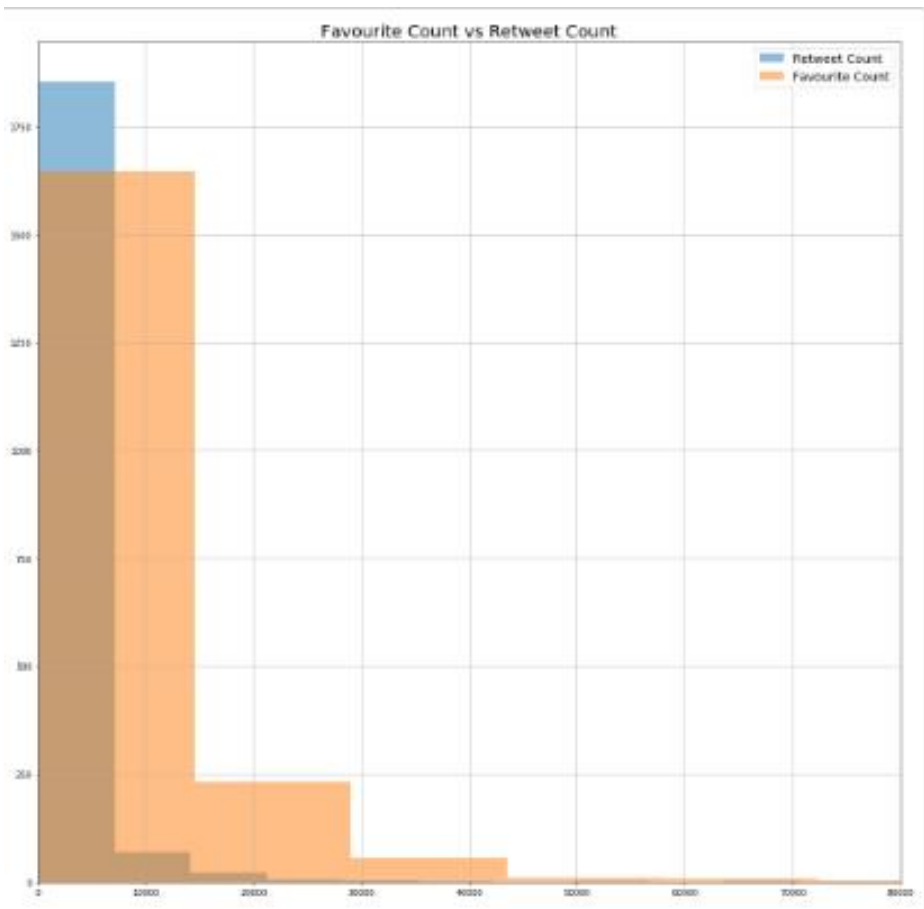Dogs in dog stage pupper are the ones that are mostly tweeted about.

An image from the WeRateDogs Twitter page.

INSIGHTS

Here are 3 key these to note from this dataset:

1. The most retweeted tweets also have a very high favourite count. They are clearly quite popular.
2. The most retweeted dog stage is doggo
3. The more favourited a tweet is, the more likely it is to be retweeted again and again.

Here is Bar Graph showing how Favourite Count is associated with Retweet count.

Something else that I found interesting was the rating of the dogs. The numerator is most of the times bigger than the denominator which is mostly 10.

The reason the give is, 'They're good dogs Brent!'