

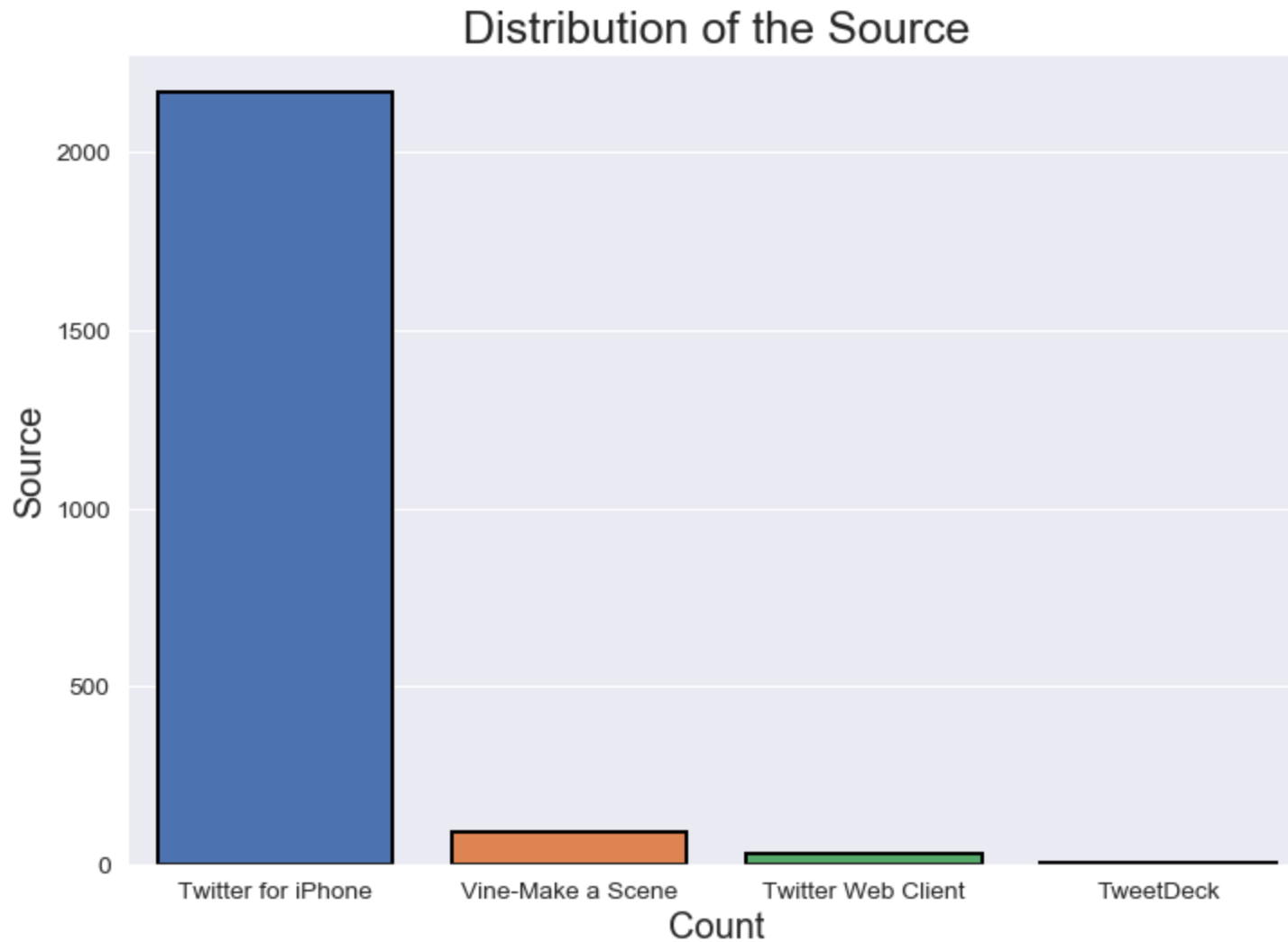
WeRateDogs Twitter Archive Insights

Introduction

The dataset that we will be wrangling, 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. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage. The Twitter archive is great, but it only contains very basic tweet information so we are going to gather the data using variety of sources, assess the data and clean the data

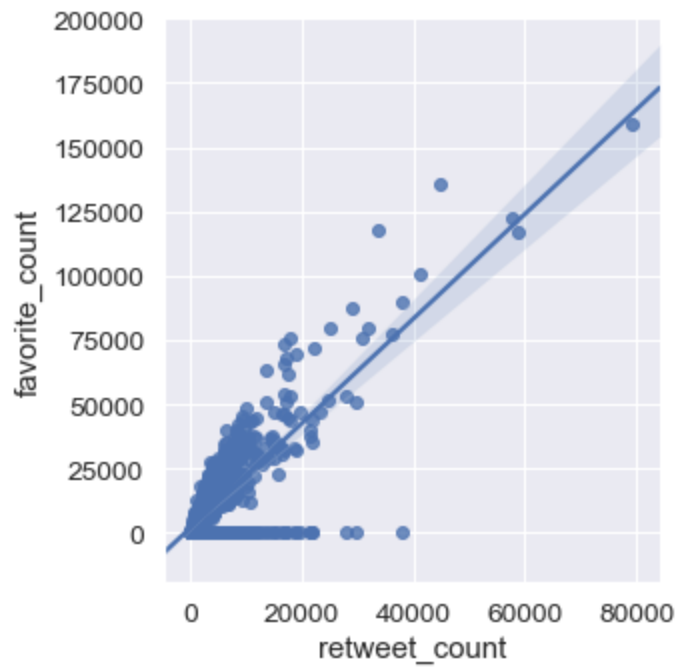
Insight 1: Distribution of the Source

Interestingly majority of the source of the data is from Twitter for iphone, all other sources does not have major part in data gathering.



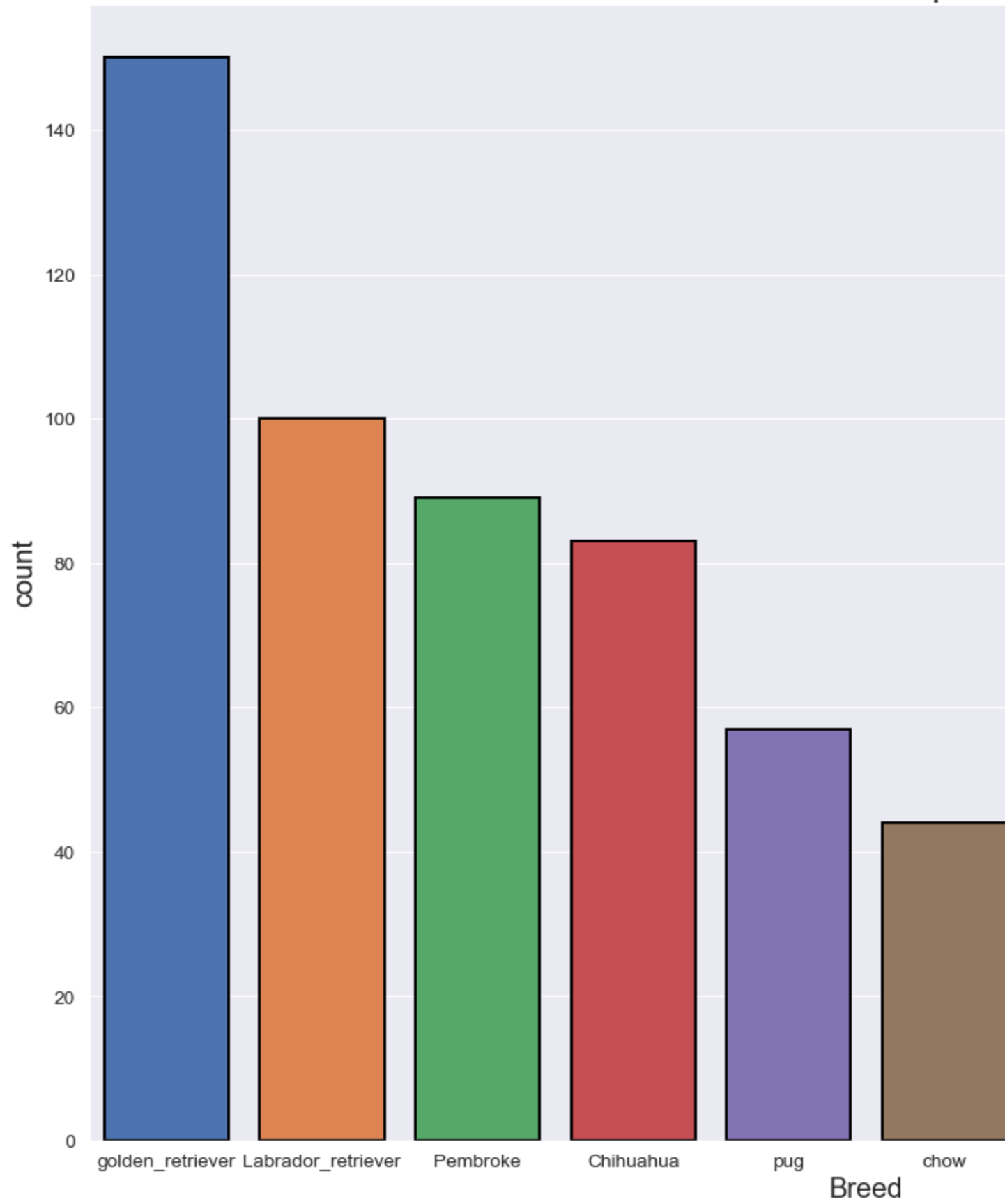
Insight 2: Relationship between retweet count and favorite count

There seems to be a very strong positive correlation between retweet count and favorite count with r_s (spearman coefficient) of 0.707. Strength of association between variables can be expressed in terms of spearman coefficient and the value is between +1 and -1. Positive correlation indicates that if one variable increases other variable also increases and closer the value to 1 the more strong the relationship is. In our case spearman coefficient is 0.707 which is positive and closer to one which indicates that retweet count and favorite count are strongly related.



Insight 3: Visualizing first probability

Distribution of first proba



When we look the visualization for the most predicted dog not surprisingly its Golden retriever followed by Lab.