Wrangling and Analyze Data

Image via Boston Magazine

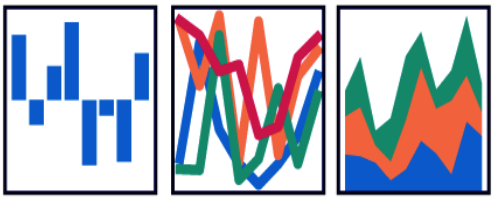
Introduction

Real-world data rarely comes clean. Using Python and its libraries, you 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. You will document your wrangling efforts in a Jupyter Notebook, plus showcase them through analyses and visualizations using Python (and its libraries) and/or SQL.

The dataset that you 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. 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.

pandas

$$y_i t = \beta' x_{it} + \mu_i + \epsilon_{it}$$



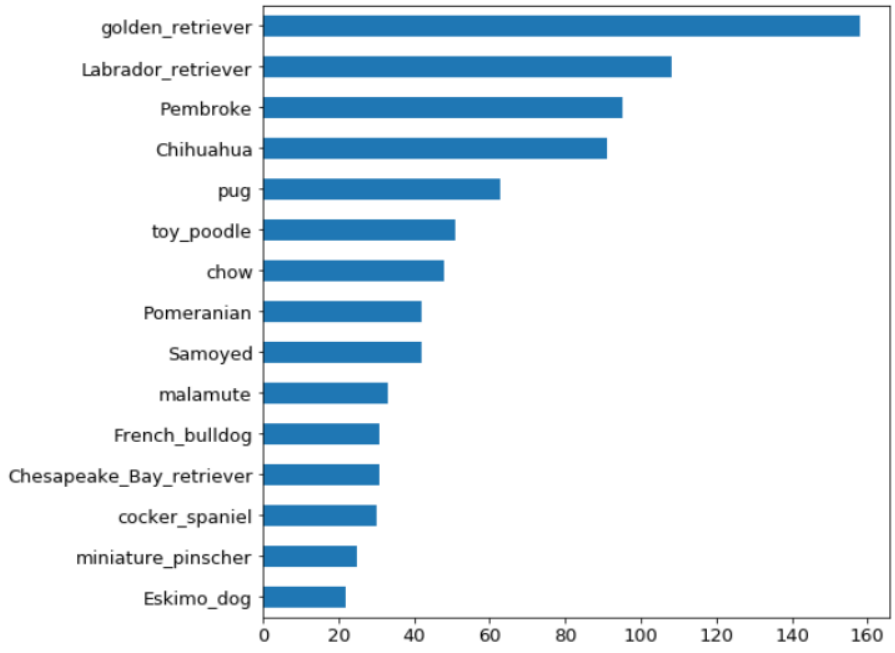
Data Visualization Highlights:

After gathered and assessing and cleaned the data it's time to visualization the importance of the tweet datasets through four questions:

1. What are the most type of dog frequency?
2. what is the most source of tweets ?
3. is there a correlation between favorite numbers vs retweet numbers?
4. What are the most dog stage of dog frequency**?**

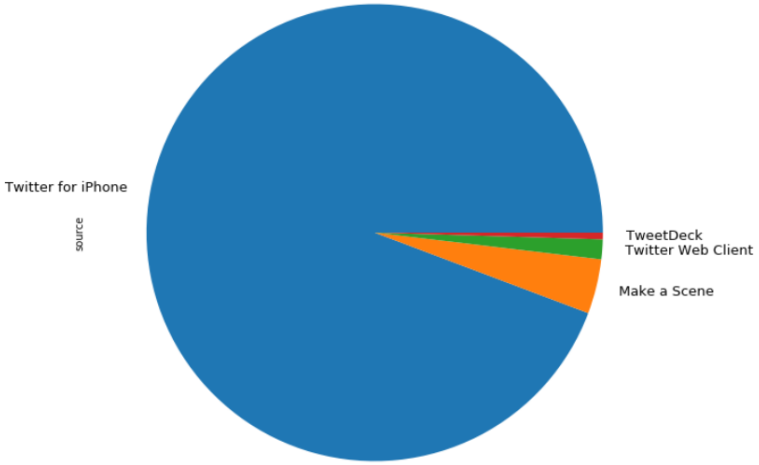
1. What are the most type of dog frequency?

Show Graph 1 the most type of dogs gets the highest frequency and the most frequency dog type was for "**golden retriever**" and the second to **Labrador Retriever** and the third one to **Pembroke**.



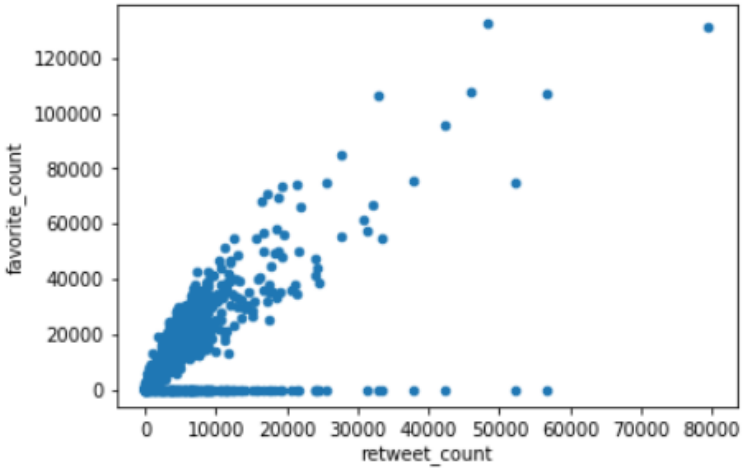
2. what is the most source of tweets ?

Show Graph 2 the most source of the tweet and appears the lower source was by **tweet Deck** then through **Twitter Web Client** after that by **Make A Scene** and the higher source came through twitter for **iPhone**.



3. is there a correlation between favorite numbers vs retweet numbers?

Show Graph 3 the correlation between retweet count and favorite count of tweet and appear as **positive** correlation between retweet count and favorite count of tweets by WeRateDogs twitter account



4. What are the most dog stage of dog frequency?

Show Graph 4 the most stage of dogs gets the highest frequency and the most frequency stage of dog was for "**pupper** " and the second to **doggo** and the third to **puppo** and the lowest to **floofer**.