WRANGLE AND ANALYZE DATA
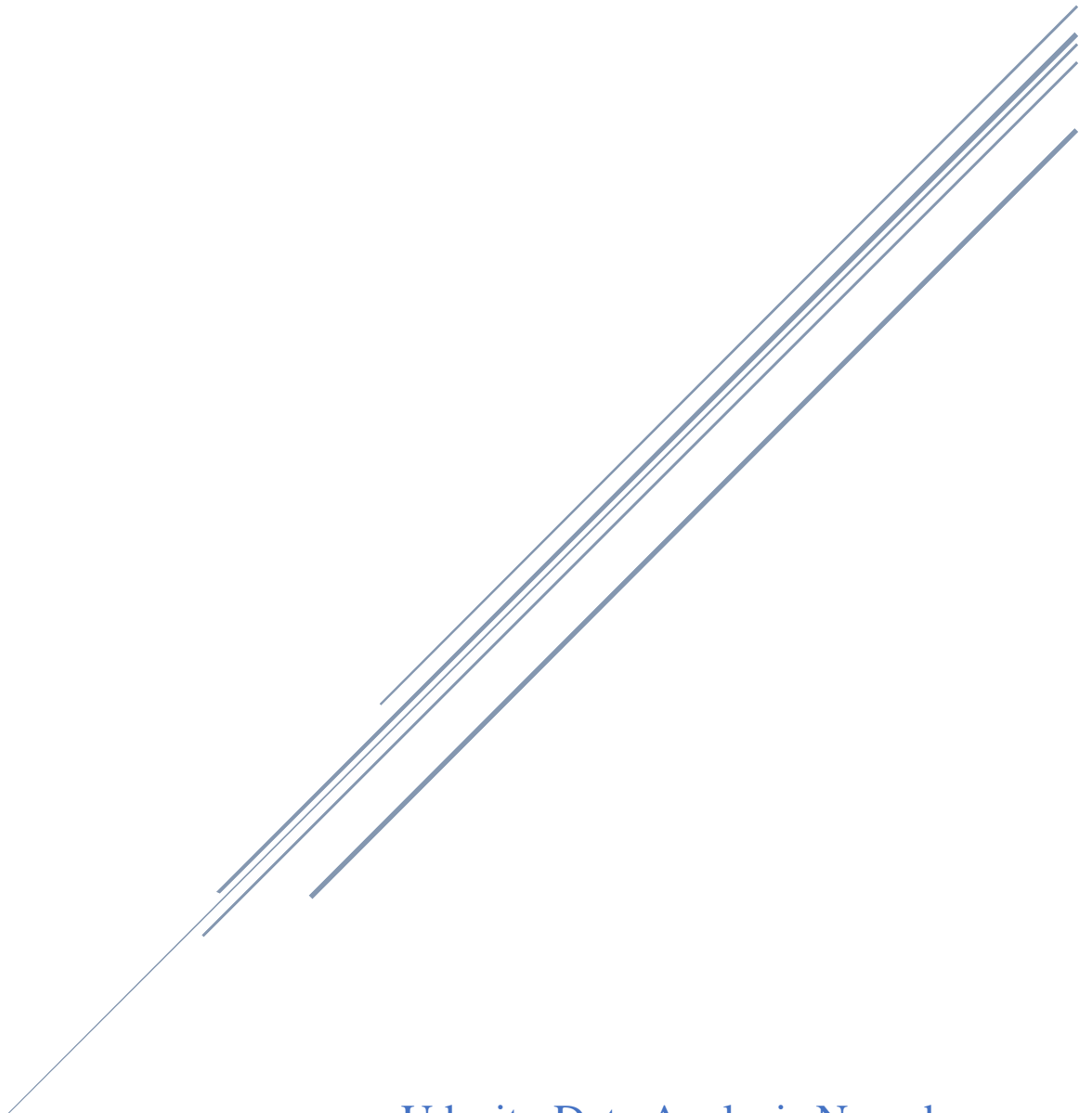
Act Report

Introduction

Real-world data is rarely accurate. I will collect data in a variety of forms and sources using Python and its libraries, evaluate its quality and tidiness, and then clean it. We call this "data wrangling." I'll keep track of my data wrangling activities in a Jupyter Notebook and present them using Python analyses and visualizations. 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.
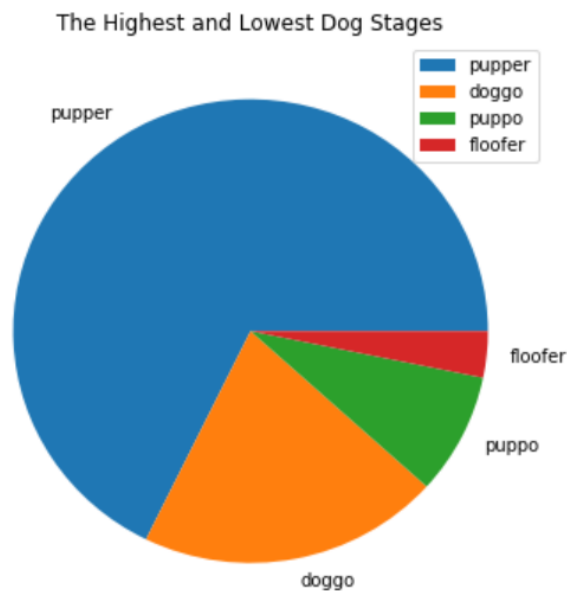
Analyzing and Visualizing Data

In this step, I will analyze and visualize the wrangled WeRateDogs data, I have done before by asking some questions as following:

- What is the highest dog stage and lowest dog stage?
- What are the five highest breeds of dogs?
- Is there a relationship between favorite count and retweet count for breeds of dogs?
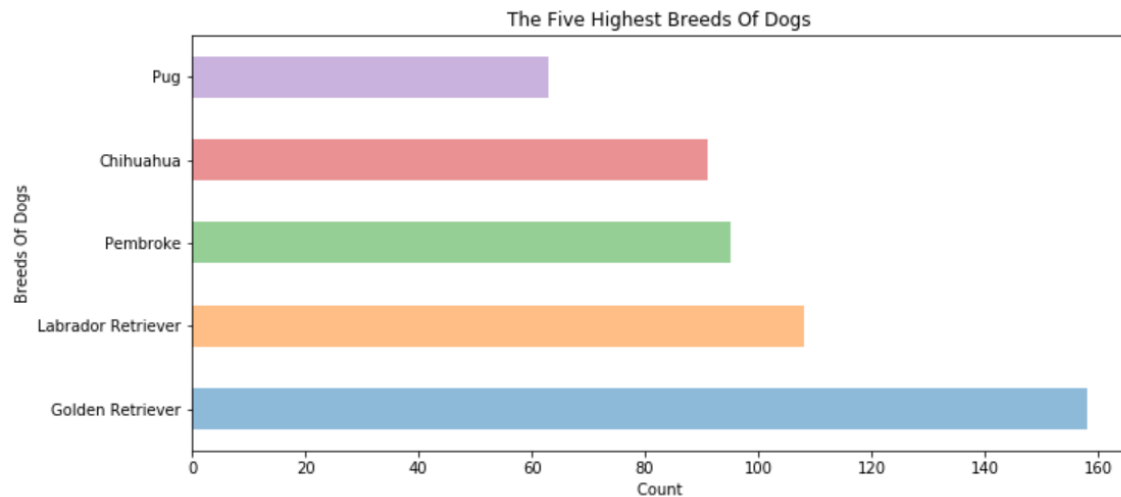- What is the most common numerator?

Next, I will explain the answers to these questions

1- What is the highest dog stage and lowest dog stage?

The Highest and Lowest Dog Stages

So, as shown in the above pie chart:
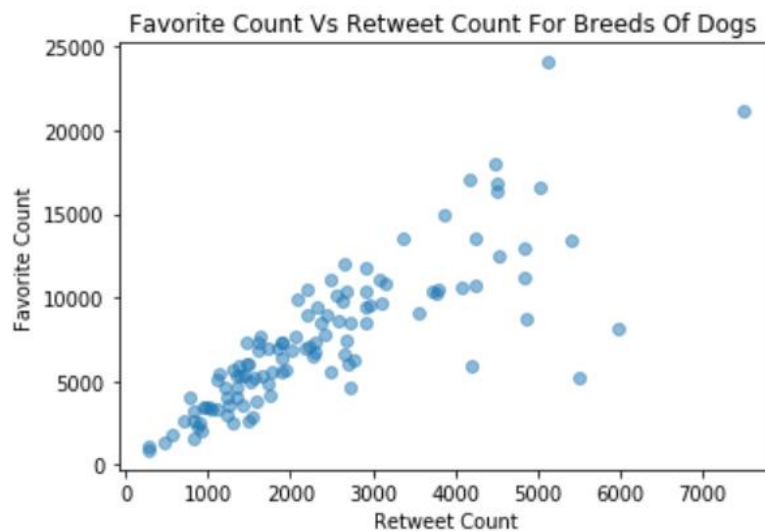The highest dog stage is pupper and the lowest dog stage is floofer.

2- What are the five highest breeds of dogs?

The Five Highest Breeds Of Dogs

So ,as shown in the above barh chart:
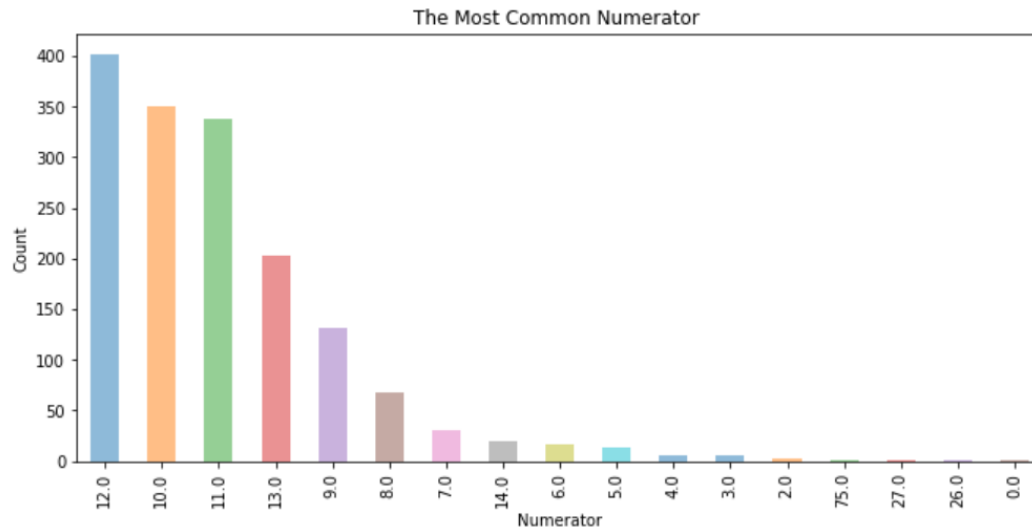The five highest breeds of dogs are: Golden Retriever, Labrador Retriever, Pembroke, Chihuahua, and Pug.

3- Is there a relationship between favorite count and retweet count for breeds of dogs?

Favorite Count Vs Retweet Count For Breeds Of Dogs

So, as shown in the above scatter plot chart:
There is a positive relationship between favorite count and retweet count for breeds of dogs.

The Most Common Numerator

So, as shown in the above bar chart:
The most common numerator is 12 followed by 10 then 11.

Conclusion

In this project we wrangling and analyze the WeRateDogs data, by the following: gathering data, assessing data, cleaning data, storing data, analyzing and visualizing data. At the end of our wrangling and analyze the data, we got the following insights:

1. The highest dog stage is pupper.
2. The lowest dog stage is floofer.
3. The five highest breeds of dogs are:
   - Golden Retriever
   - Labrador Retriever.
   - Pembroke.
   - Chihuahua.
   - Pug.
4. There is a positive relationship between favorite count and retweet count for breeds of dogs.
5. The most common numerator is 12 followed by 10 then 11.