Evette Trevino

Professor Steve Kozakoff

Data Visualization - C939

May 15, 2023

Data Wrangling Project: Act Report

This report is intended to communicate the three major insights I found during the analysis process of the during the Data Wrangling process using datasets from a twitter account. The account is We_Rate_Dogs or user name @dog_rates. It is a playful account the spotlights different dogs with fun facts and a rating. During the Analyze and Visualize process of Data Wrangling, I came to three main conclusions. The first is finding the dog with the highest ratings, next is the discovery of popular dog names and last is identifying the most popular dog breeds from these data sets.

Finding the dog with the highest ratings was a bit challenging. In mathematics and science, we usually don't leave a number as an improper fraction, or with the numerator (top number) larger than the denominator (bottom number). This is what made this particular twitter account interesting. The users rate the dogs and almost always provide a rating that is higher than the limit. Many dogs received 15/10. I combined all ratings from all users then grouped them by dog. Then I saw that the dog names Atticus, received total rating far above the rest. His total rating was 893, while the second-place dog received a total of 75. Atticus was a clear winner with the votes. I created a bar chart to display 15 of the highest rated dog names.

The next discovery was of popular dog names. I pulled all data and performed a count by dog names using Pandas. I found Charlie and Lucy were equally popular with being used eleven times each. I then decided to use a WordCloud to aide in visualization of these finding, along

with the 15 other top dog names. The WordCloud I found as the best visual aid as I am not able to measure it against a Y axis and it enlarges the fonts for the more popular names. Below is the output of the reference WordCloud visual.

Last, I identified the most popular dog breeds. I sorted this data set by the breed column. I used twenty breeds that were the most common or repeated in my data set. I then displayed the top breeds in a simple pie chart.

Overall, this was a great project as the subject of dogs is fascinating. Placing visualizations was the finishing touches and made the project fun.