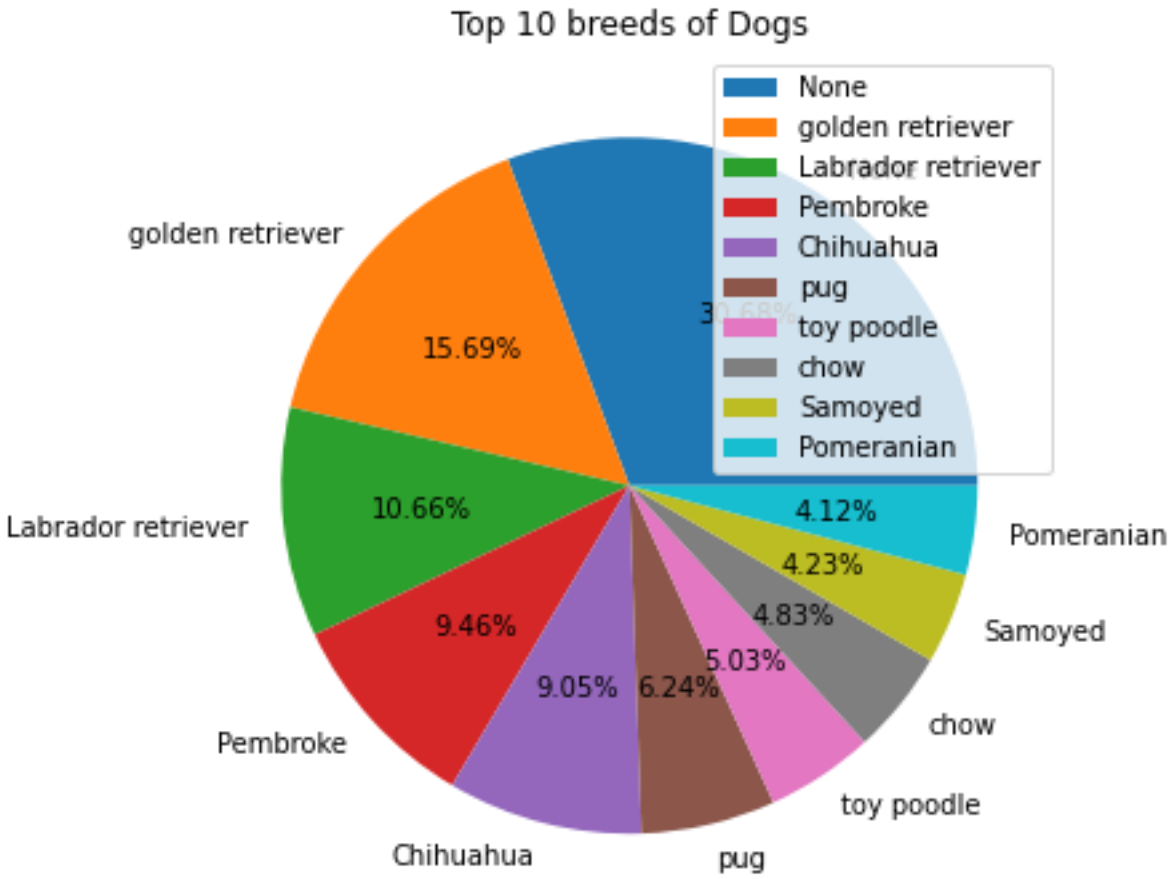
REPORT: act_report

The WeRateDogs Twitter data was in three different formats: CSV, tsv, and txt. Each contains（ 2356, 17), (2075, 12), and (2354, 3) rows and columns respectively.

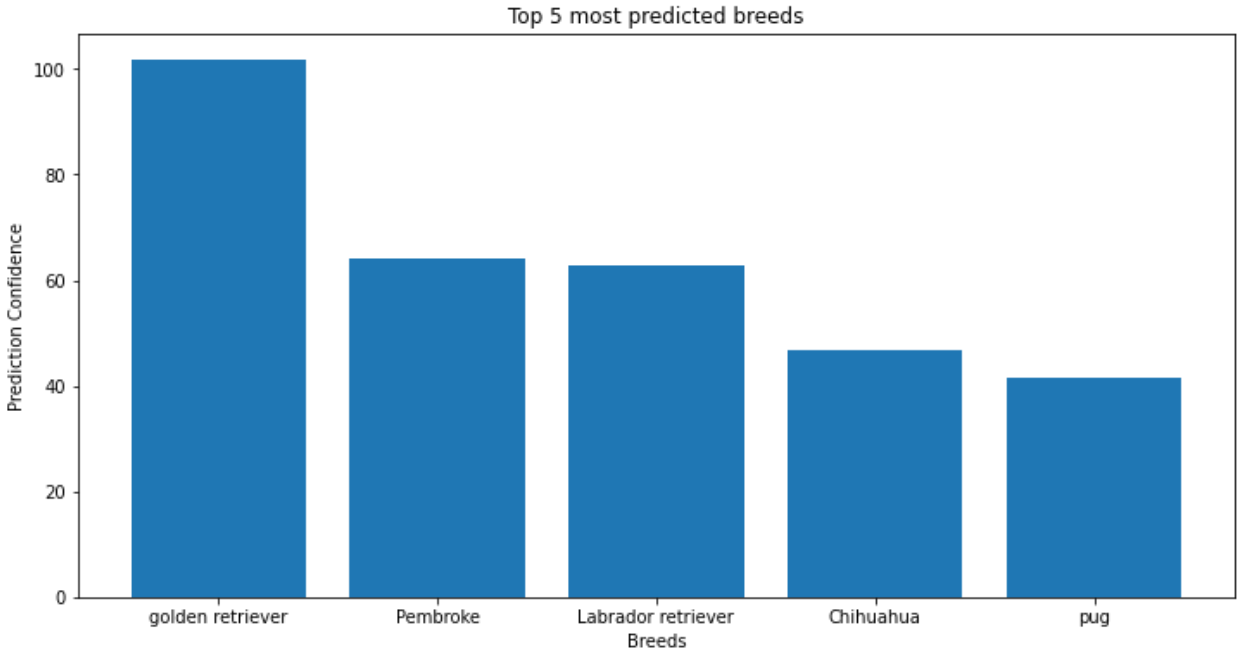
In this report, I will be visualizing some keys insight about my data. First will look at the percent age of each breed in the data, I will tilt it to the top 10 dog breeds using a pie chart. Secondly, I w ill look at how the algorithm (confidence) predicted each breed of dog, I will tilt it to the top 10 b reeds using a bar chart. Lastly, I will look the dog breed with the highest rating, then tilt it to the t op 5 breeds with the highest rating using a bar chart to visualize it.

Viz 1:



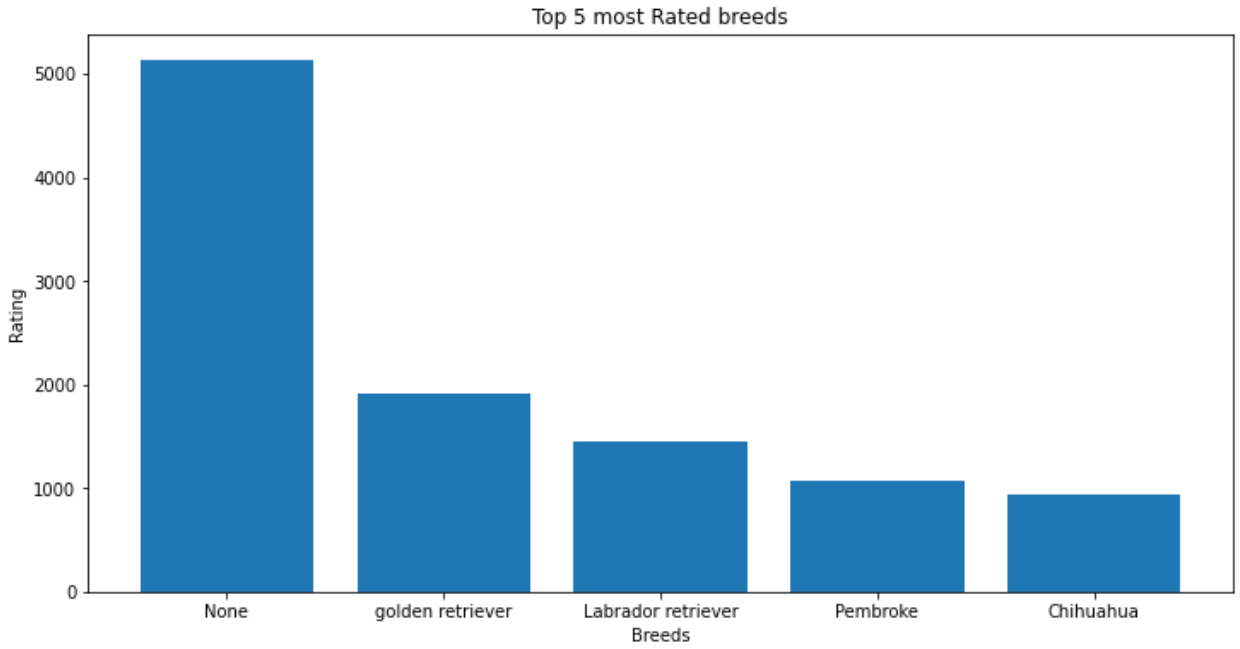
We have 114 breeds of dogs in our dataset, none value inclusive. In my visualization above, I tried to filter out the top 10 breeds in the data using a pandas visualization library matplotlib. The followings are the top 10 breeds and their overall percentage values: None - 30.68%, golden retriever – 15.69%, Labrador retriever – 10.66%, Pembroke – 9.64%, Chihuahua – 9.05%, Pug – 6.24%, toy poodle – 5.03%, chow – 4.83%, Samoyed – 4.23% and Pomeranian – 4.12%.

Viz 2:



Also, I looked at how well the algorithm predicted for each of the breeds which is the confidence. From the chart above, I looked at the top 5 most predicted breeds from the algorithm and it shows that golden retriever performed so well, followed by **Pembroke, Labrador retriever breeds, Chihuahua,** and **pug**.

Viz 3:



From the above chart, the top 5 most rated breeds are **None**, **golden retriever**, **Labrador retriever**, **Pembroke**, and **Chihuahua**.