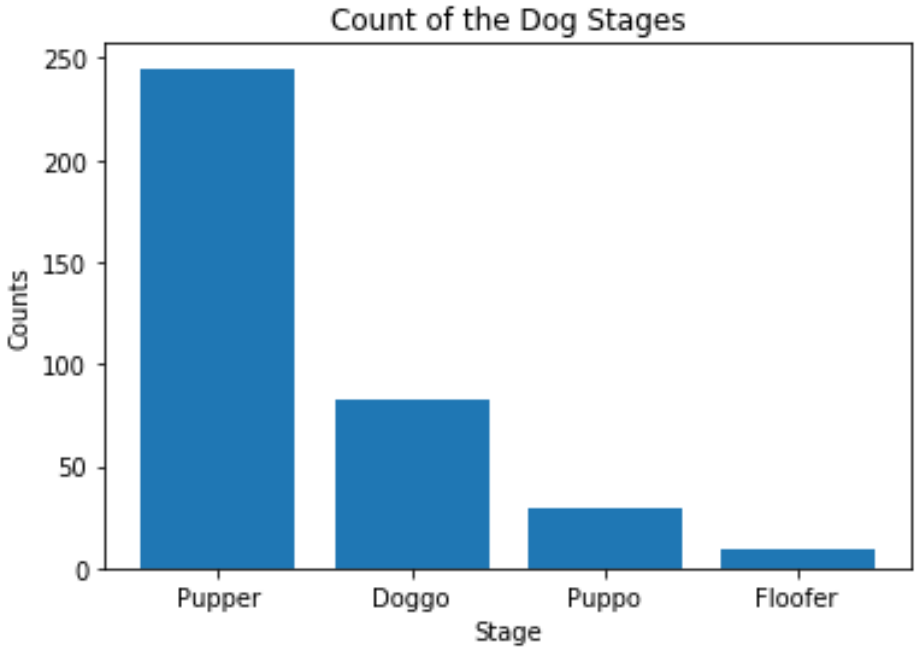
Analysis and Visualization

The first insight I've conveyed about the percentage of counts for each dog stage from the first data frame created called df_clean.

So I started with getting the counts for each specific dog stage and the total of them after that the percentage of each stage by dividing each count over the total.

I've neglect the mix dog stages counts and in the total counts too.

Using **matplotlib** library I could get the bar chart that visualized the counts for dog stages as shown:



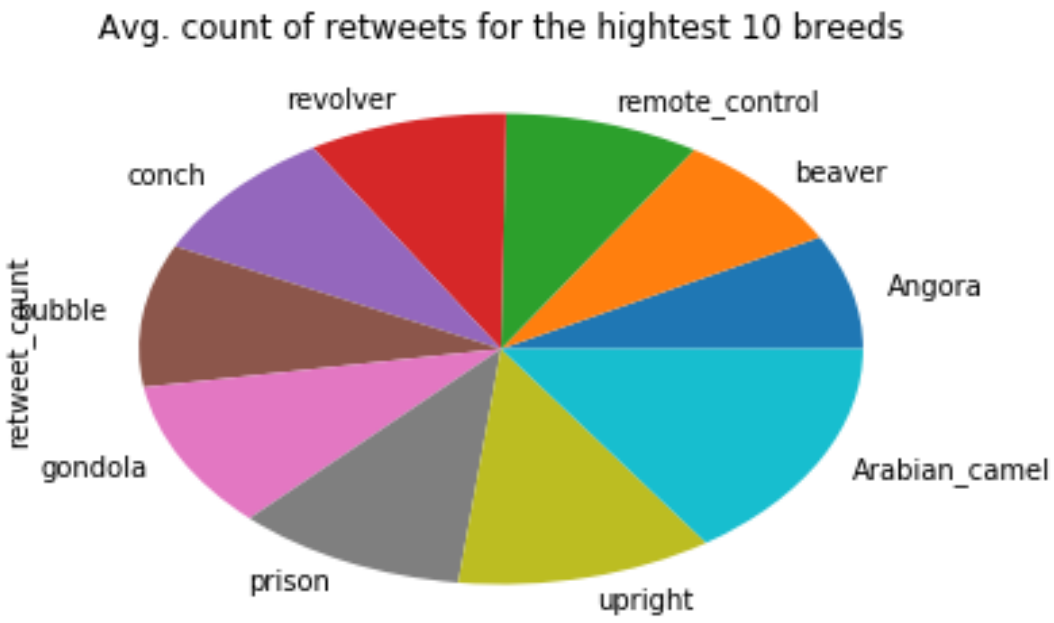
So from this analysis we can know the most popular dog stage through our dataset is the Pupper stage as it has the highest percentage over other dog stages.

The second insight I've conveyed about the breed of dogs that had the highest average retweets.

So I started with collection the data I would need in one data frame by merging the first data frame called df_clean with the second one called df2_clean to get the breed column in the same data frame with the retweet counts and favorite counts, it could be applied by the same way of retweet counts, and user counts too.

After merging, breed list was created to get the highest ten breeds in average retweet counts.

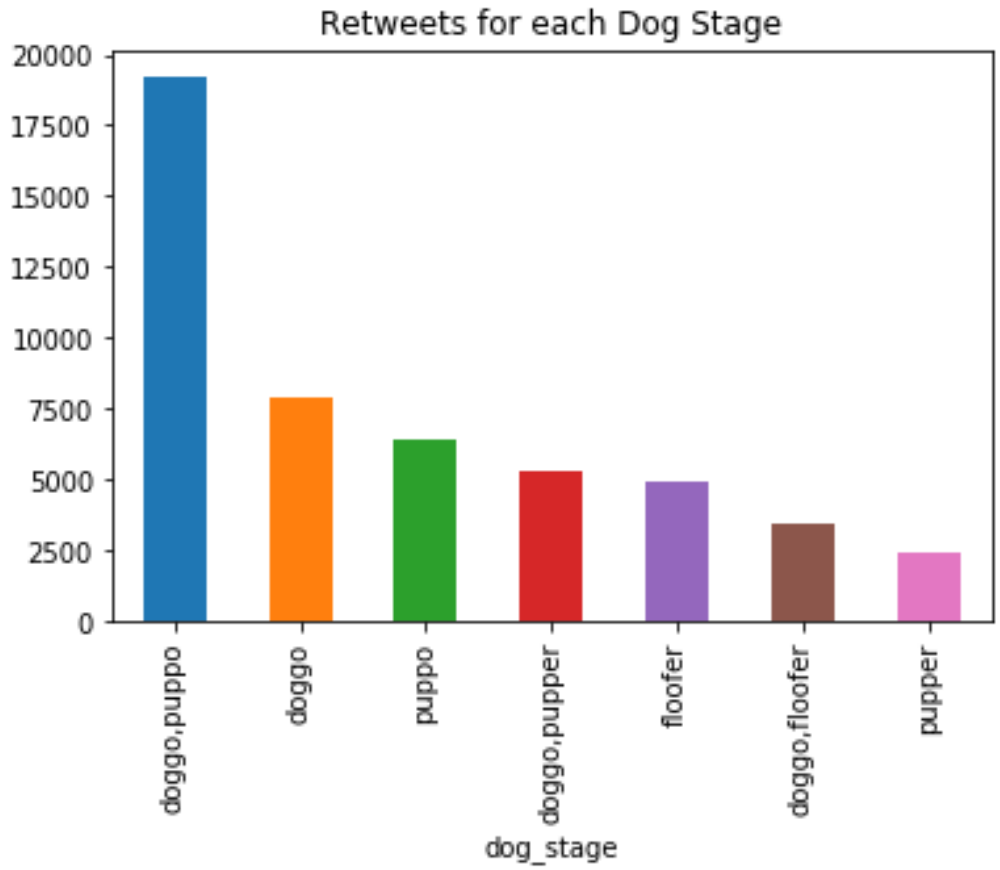
Using **matplotlib** library I could get the pie chart that visualized the average retweet counts for dog breeds, only the highest ten breeds, as shown:



The second insight I've conveyed to access the dog stages in terms of their average retweet counts too.

Using the previous merged data frame called dfmerged, or the first data frame called df_clean alternatively, the relation between the dog stages and the average of retweet counts could be gotten easily.

Using **matplotlib** library I could get the bar chart that visualized the average retweet counts for dog stages as shown:



So that were insights and visualizations produced through my wrangle notebook, I hope that was accurate and clarified reported.