

Data Wrangling Act Report

This report summarizes the findings of the data wrangling project. The report content will be as follows.

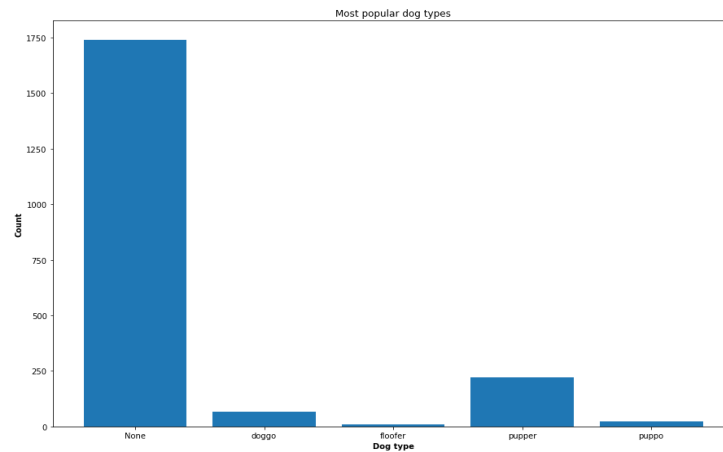
In the project, we obtain data from different sources; such as twitter enhanced data from the Udacity platform, Image prediction data wrangle from online and tweet data query from twitter API.

All the data were gathered, assessed to detect any quality and structural issues in the gathered data set and cleaned all the issues discovered.

Insights gained in the analysis of the cleaned data.

Visualization of some of the data was done to give some insight into the new, cleaned and merged data set.

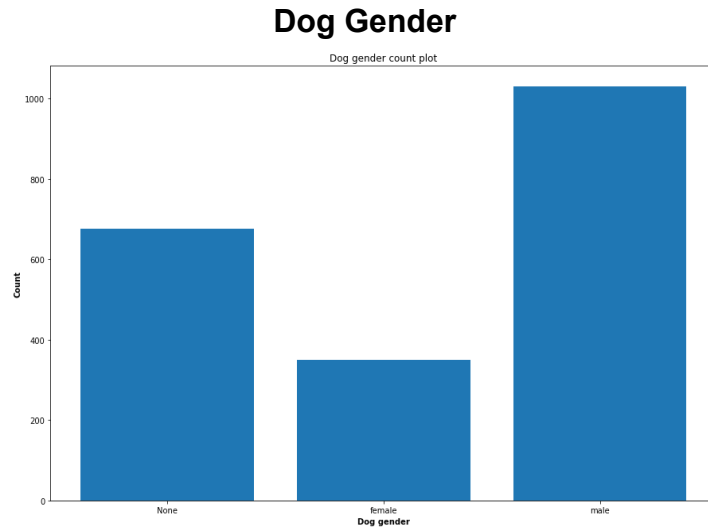
Visualizing the dog life stage



From the visualization above, we can observe that there exist four different life stages with unidentified life stages making it five(5). Below are the figures.

life_stage	
None	1739
doggo	65
floofer	8
pupper	221
puppo	24

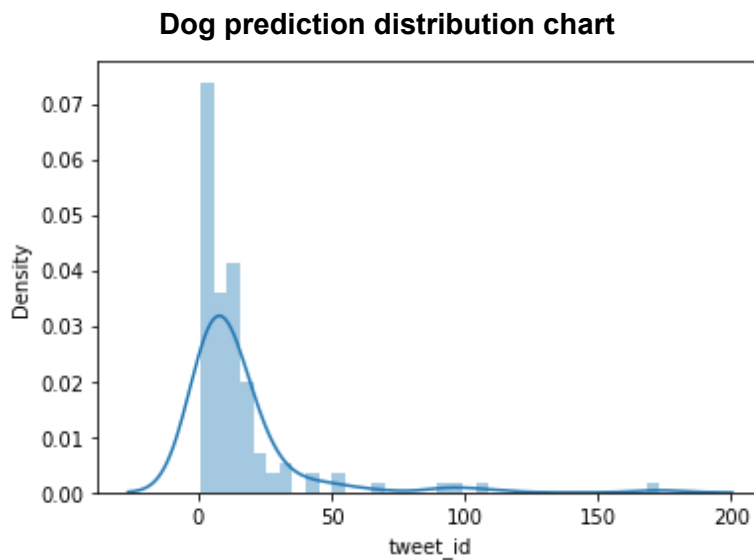
Visualizing the dog gender also has 3 bars inclusive of None which belongs to the dog type with unidentified sex.



From the visualization above, we can observe that there exist three different dog genders with unidentified gender making it four(4)). Below are the figures.

dog_gender	
None	677
female	350
male	1030

The distribution below shows the distribution of prediction of the final dataset. We can observe from the chart below that the dog prediction which is an integer is normally distributed.



In conclusion, Data wrangling makes data analysis very effective. So it is pertinent to always do wrangling before any analysis.