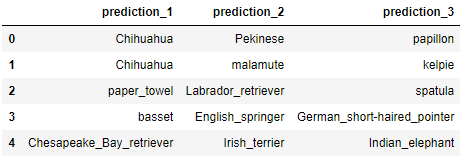
Analysing data

I asked 6 questions at the beginning of my analyses in my Notebook, then I tried to provide an answer for them.

Some very interesting insights were found and presented in the notebook. I have also tested a hypothesis on the dataset and did some statistical analysis. The results of these analyses are presented in the notebook with comprehensive description and plots.

Some of the findings are also listed below:

* As we can see from the tables below, Golden Retriever is the most frequently appeared Breed in Prediction 1, it has appeared 139 times in prediction 1 column.



* The tope 10 most popular names are listed below:
  + a 46
  + Lucy 10
  + Cooper 10
  + Charlie 10
  + Oliver 9
  + Tucker 9
  + Penny 8
  + Sadie 7
  + Daisy 7
  + the 7

Obviously, the names “a” and “the” are not real names. It could be another small project within the data wrangling phase to find out what the cause of it was.

* The table below shows the statistical facts about “rating”:
* count 1686.000000
* mean 1.089195
* std 0.250791
* min 0.000000
* 25% 1.000000
* 50% 1.100000
* 75% 1.200000
* max 7.500000
* Because the p-value in our t-test analysis was less than 0.05 and indeed it was far below that number. So, **statistically there was a significant difference between the number of retweets based on whether or not the dog had a dog stage label.**