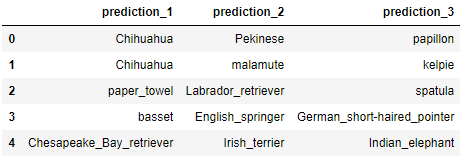
Analysing data

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

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 the dog had a dog stage label.

Below I have brought the plots for time series analysis of retweet counts, Favorites Count, and rating over the entire time period from 2015-11 to 2017-09.

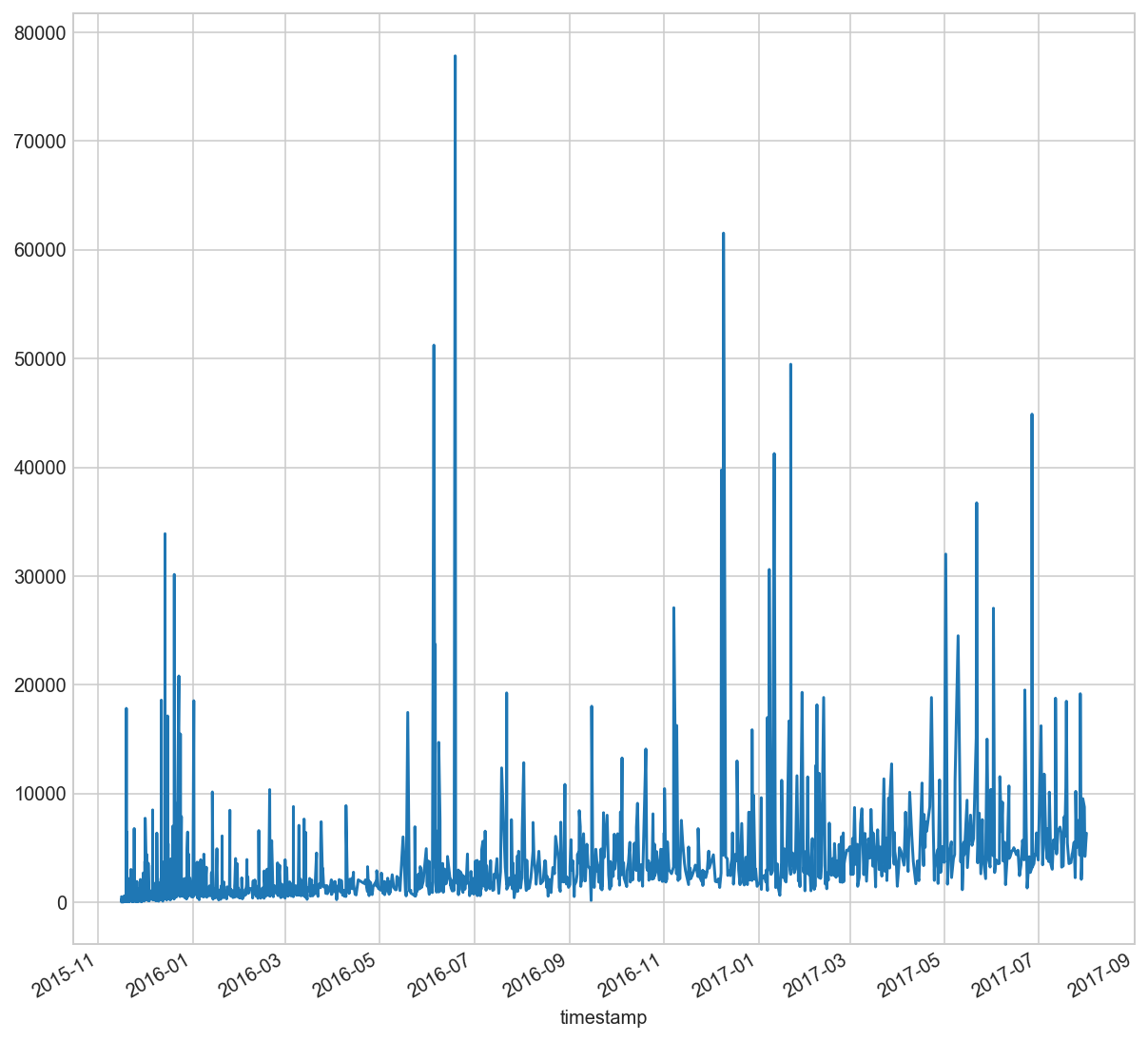


Figure 1 - Retweet Count over time

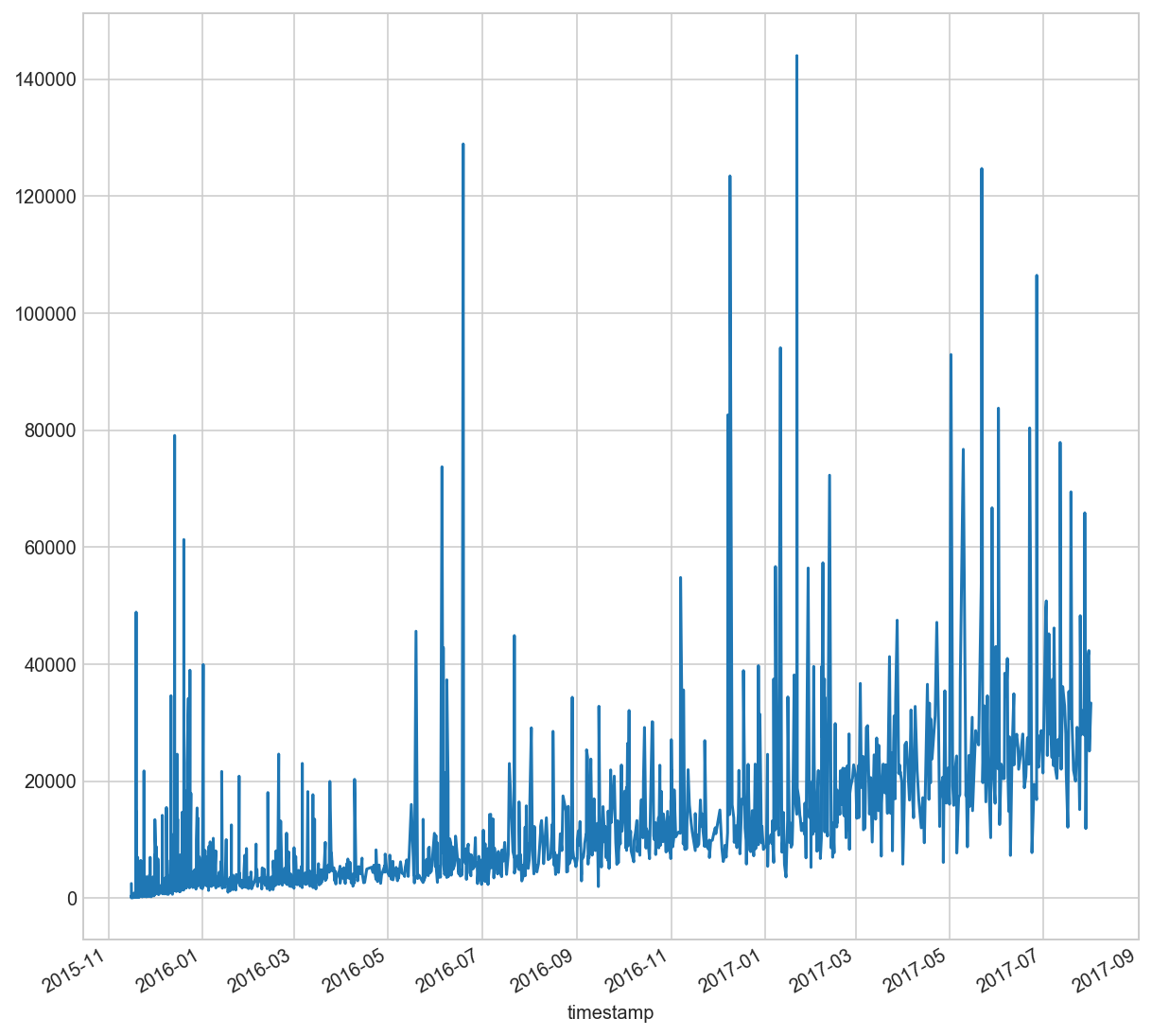
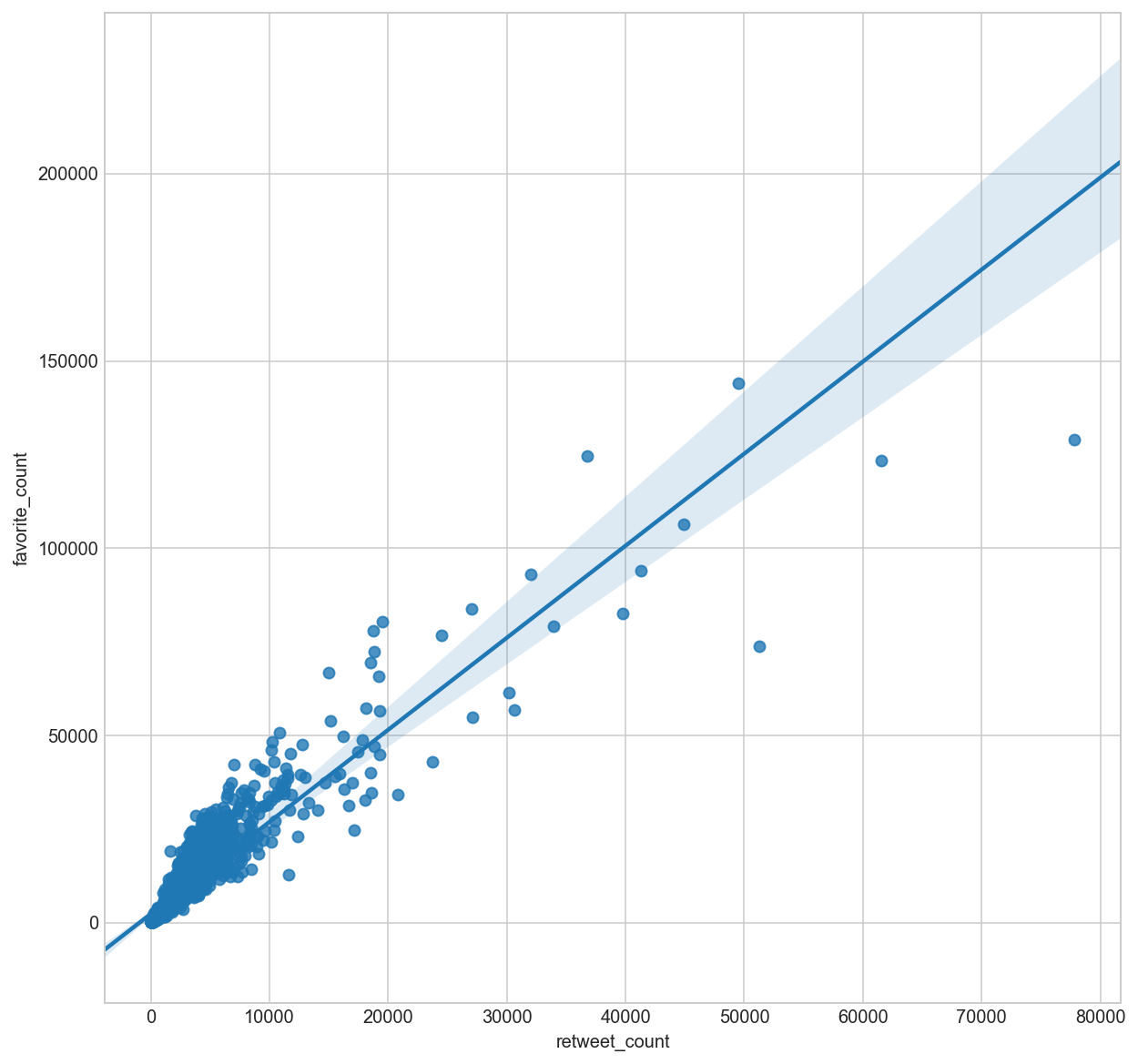


Figure 2 - favourite Count over time

A scatter plot of favourite count against retweet count is produced (see below). This plot shows that there is a very strong positive correlation between the two variable which is what we expect as well. There are some outliers in this plot that represent the ones which have receive extremely high retweets and favourites (likes). I was interested to see the images of these dogs, so below, I have brought them as well.



The highest retweet count was 77844 times, and the image of the dog is below:



The highest favourite count was 144041 times, and the image of the dog is below:

