# Exploring data and visualization

After gathering different datasets, merging them and cleaning some of their content, exploring data is needed to confirm some observation. Some questions arose. APIs for visualizations were used to answer them.

Below the questions will be listed with their respective answers and charts.

1. Which dogs' breeds (between the top 10 listed one) in second prediction have the highest rating?

As there are several breeds listed in the dataset, I opted to analyze only the most common breeds.

Then, the ratings of these breeds were visualized in the bar chart. The top 2 rated most common dogs are Chihuahua and French bulldog.
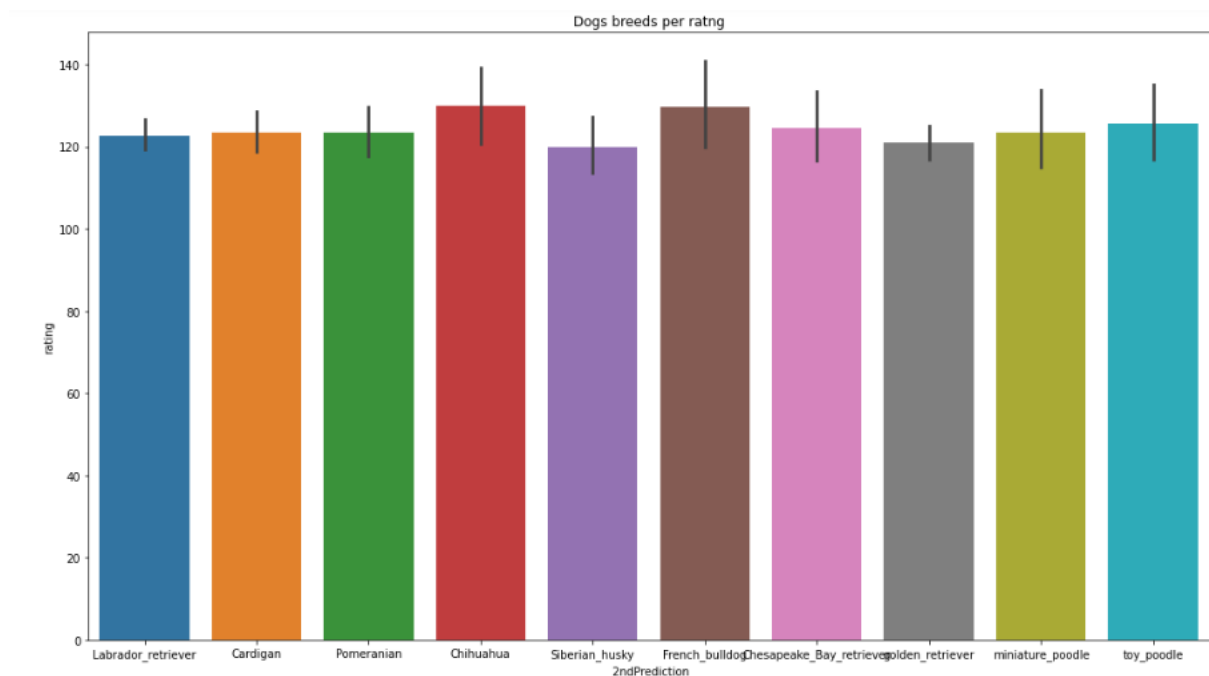

2. Which classification has the highest rating?

To answer this question, I drew a bar chart. The classification with the highest rating is 'Pupper'. It is also the most common one in the dataset per an earlier analysis (during dataset assessment).

classification per ratng
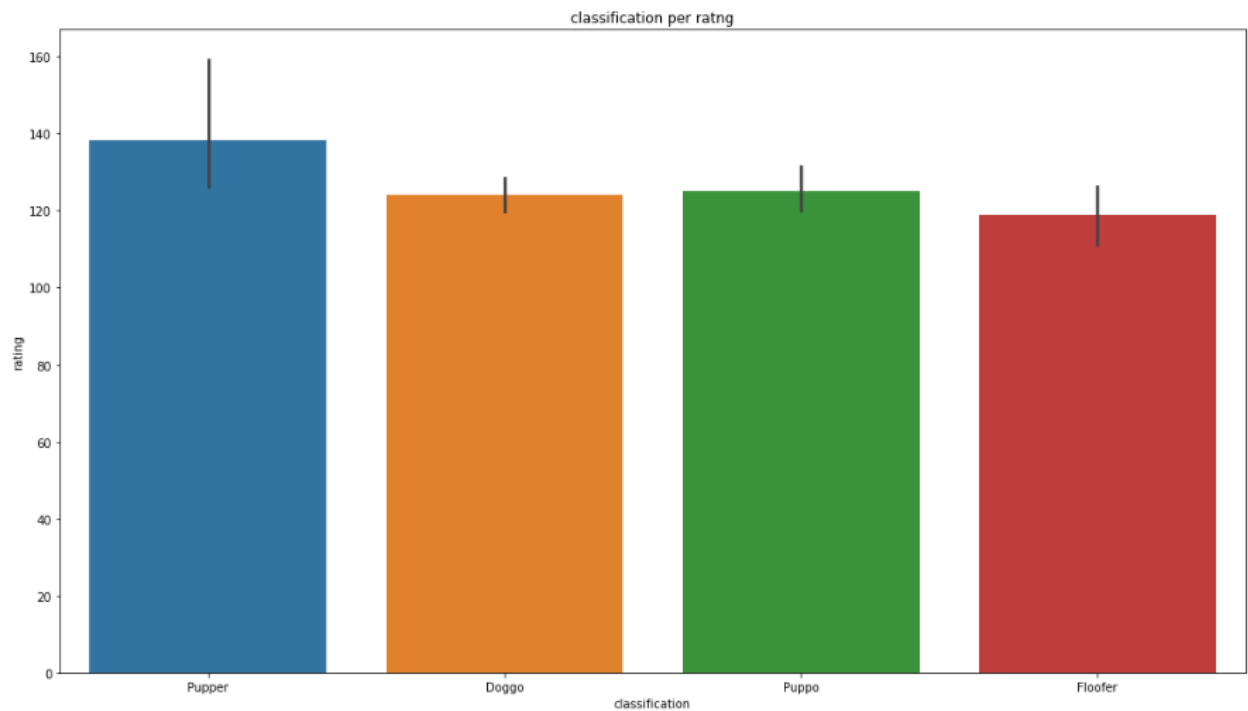
3. Which most common dogs' names are used

There are 2006 different names in the dataset. Drawing a chat with all these values will not be clear.

Hence, I opted to check the top ten values. Surprisingly, the list of the names shall be cleaned too. After cleaning, the top ten list, a pie chart was drawn. The most common name is Cooper. Then, we find three names with the same usage percentage. They are: Tucker, Oliver and Charlie. The least common name between Kota.

Most common dogs names
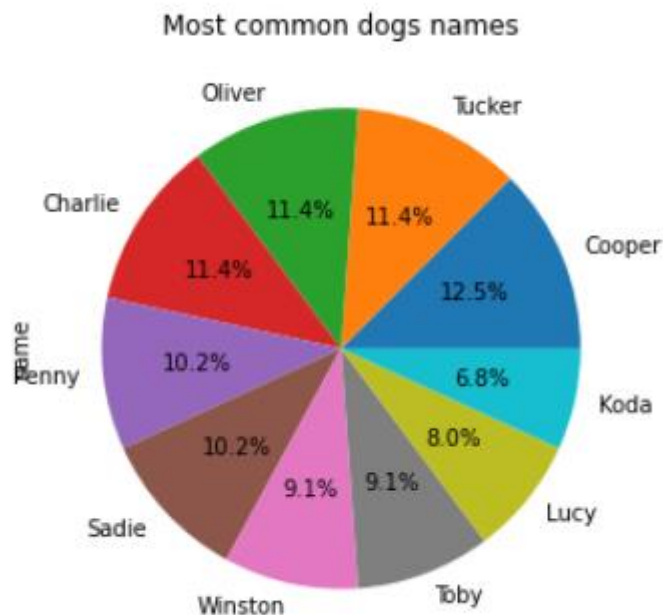
There are other questions that can be answered to explore deeper the dataset and draw conclusions.

While working on these questions, I found out that data analysis steps are not exactly sequential. We may return back to assessment or cleaning steps which is the case when answering the third question related to dogs' names. Also this draws the importance of data wrangling within data analysis.