

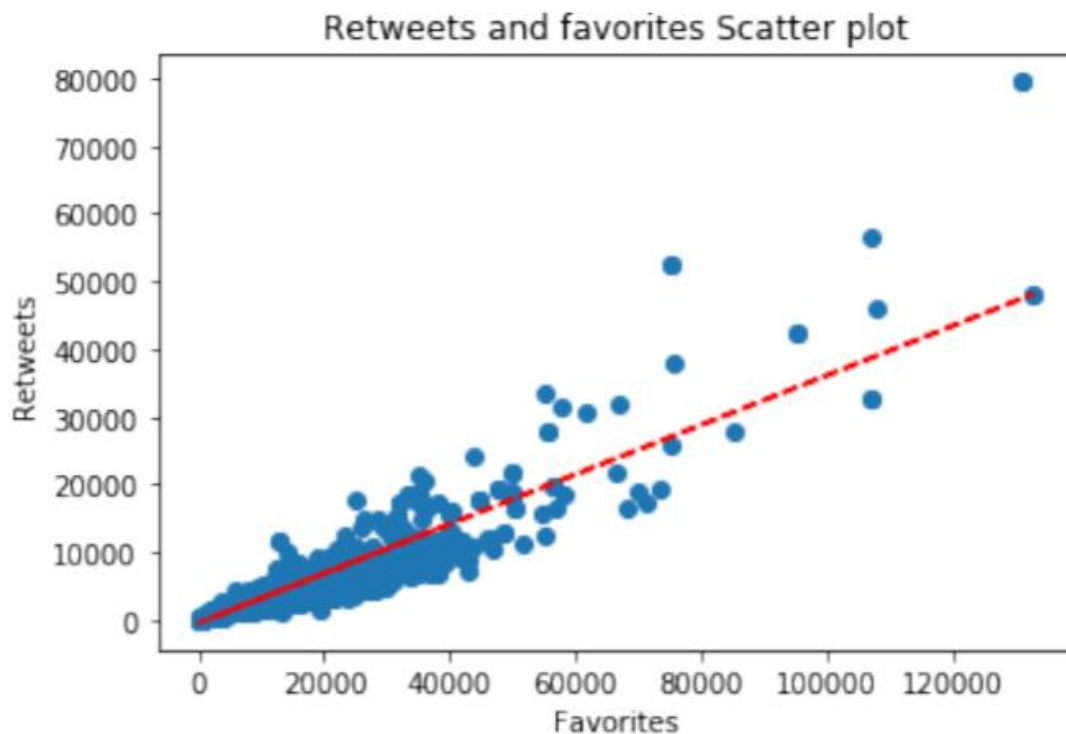
Data Analysis and Visualization

This project aims for implementing data wrangling skills on a dataset collected from multiple resources: API, server, and local.

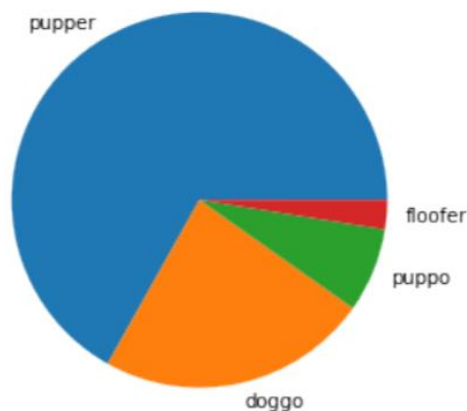
The datasets are about a twitter profile (WeRateDogs) that tweets dogs' pics and rates them with above 100% percent, because they are all good.

This report shall include some of the analysis I did on the dataset.

1. The Retweets vs Favorites scatter plot, which shows the correlation between the favorite counts and the retweet counts.



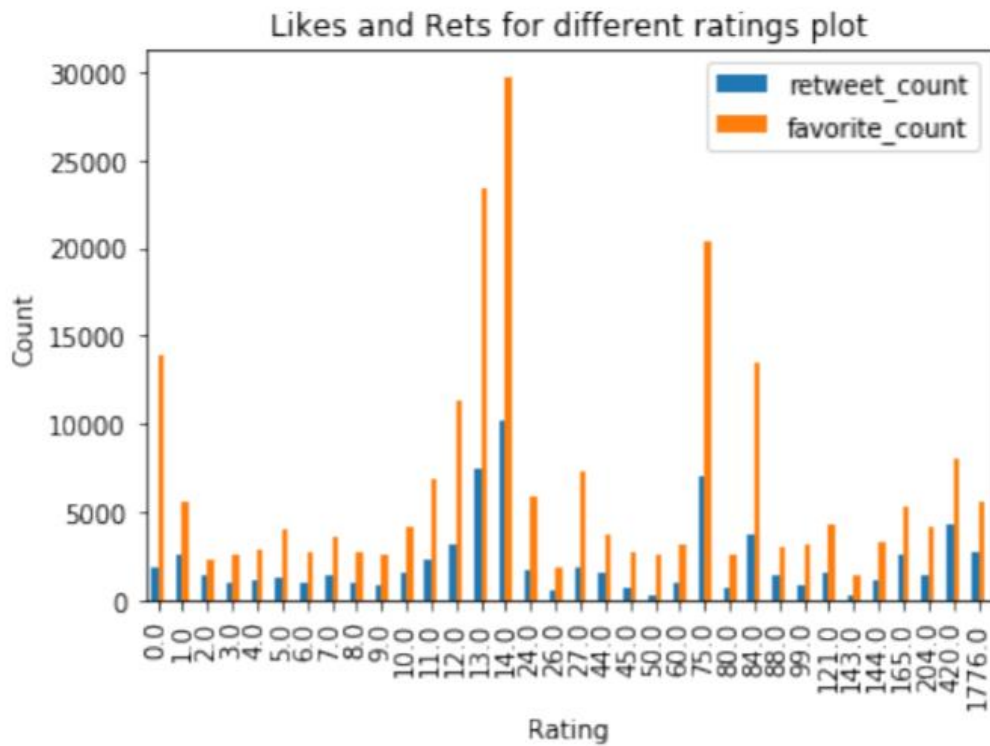
2. A pie chart illustrating the percentages of dogs' types compared to each other, pupper dogs are dominant over the other types



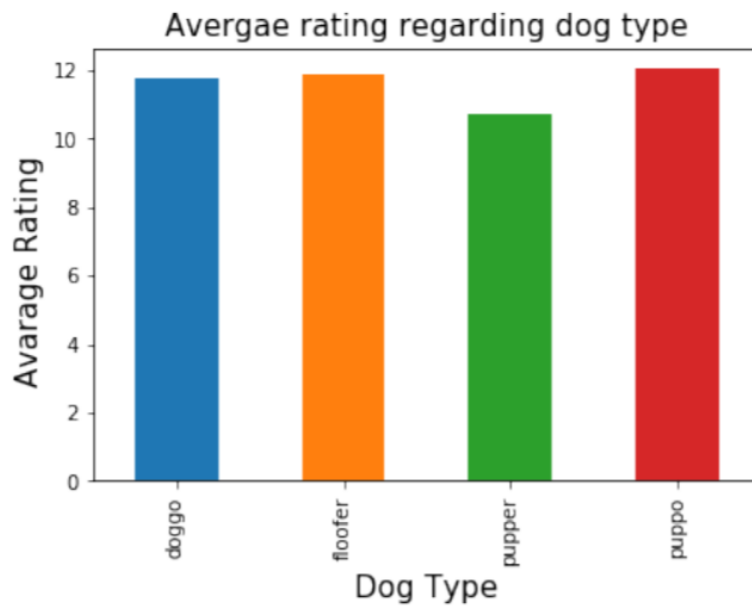
- Dataset description , explaining the min and max with other important statistics about the rating, favorite and retweet counts.

	favorite_count	retweet_count	rating
count	2311.000000	2311.000000	2311.000000
mean	9234.764171	2917.178711	12.118563
std	12909.085456	5153.877835	38.554451
min	81.000000	16.000000	0.000000
25%	2153.000000	666.000000	10.000000
50%	4257.000000	1415.000000	11.000000
75%	11510.500000	3265.500000	12.000000
max	132810.000000	79515.000000	1776.000000

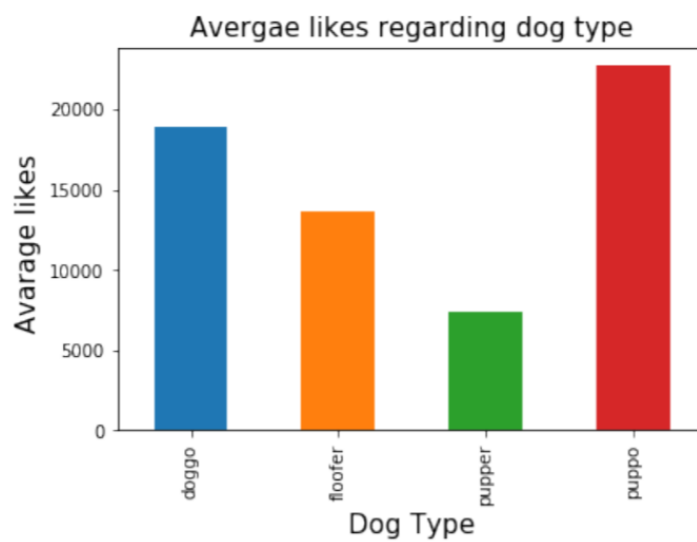
- Plot illustrating the relation between the favorite and retweet counts regarding the rating. Indeed there exists some extreme values like 1776 rating, but the one with the max favorites has only 14/10 rating. Also 13/10 and 75/10 rated dogs are having a big fan base.



5. Plot comparing the dog types regarding the average rating. The puppo appears to be very close with the floofer and doggo.



6. Plot comparing the dog types regarding the average favorite counts. The puppo has the most average likes compared to other types, then comes the doggo.



Some important notes : -

- **The most liked dog tweet award :**



- **The most retweeted dog tweet award :**



- **The highest dog tweet rating award :**



Performing the data visualization is extremely important for data analysis ; not only it helps the analyst to draw important conclusions, but also helps him with data cleaning verification, as it might come in handy to uncover some issues regarding the data.