

Act Report

So, at the end of the whole process of wrangling data where the final output we have a table with information from different sources, already clean in csv format, now we have to do the most interesting part and, after all, the objective of all data analytics. This stage is where creativity plays a great role and it is very useful to have an approach with lateral thinking. Although we can reach this stage with definite questions, when seeing the clean and organized dataset we can come up with very useful insights.

I opened the csv file as a panda's data-frame and when looking at the information I got different questions, which are:

- What are the breeds of the 3 best tweets with the most favorites?
- What are the most common dogs' names?
- What are the breeds with most retweets?

The first question I could easily answer by doing a sort by 'Favorite' column in descending order and limiting the results to the first 3. Then filter this result by breeds, so I have the next insight:

These are the breeds for the top 3 favorites tweets for users:

- 1) Lakeland Terrier with 132810 fav.
- 2) Labrador Retriever with 131075 fav.
- 3) English Springer with 107956 fav.

To reach the result of the most common dogs names, I simply counts the name in the 'Name' column, so I have this result:

The most common dogs names are:

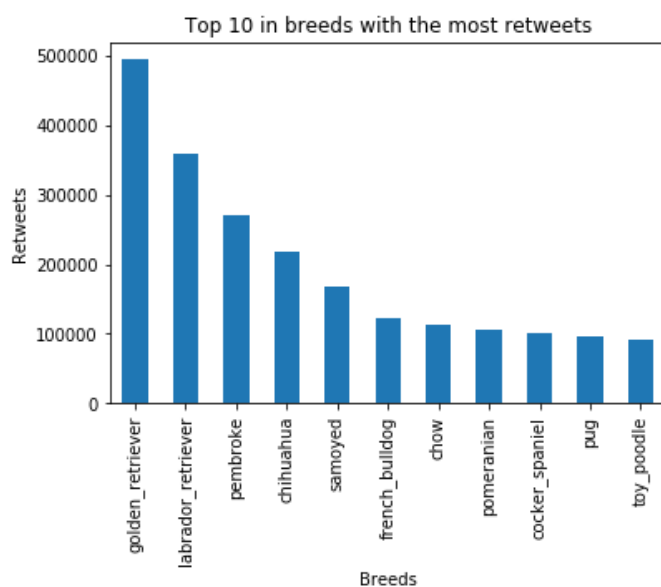
- Charlie
- Oliver
- Lucy
- Cooper
- Tucker
- Penny

- Sadie
- Winston
- Lola
- Toby

In the last question I group by 'p1' column that is the first probably of the breed, then I sum the retweets for each group by breed for later sort the results in descending order and limit them to 10. These are the interesting result:

What are the breeds with most retweets?

- Golden Retriever
- Labrador Retriever
- Pembroke
- Chihuahua
- Samoyed
- French Bulldog
- Chow
- Pomeranian
- Cocker Spaniel
- Pug
- Toy Poodle



Analyzing the graphics, we can see that golden and Labrador retriever have more chances of having more retweets.