Author: *Marcello Victorino*

Data Analysis

After Wrangling the data, it is finally time to analyze the final cleaned dataset to extract insights.

The research questions driving this investigation are as following:

1. What are the top 5 most common dog breeds?
2. What are the top 5 most liked dog breeds?
3. What are the top 5 most popular/common dog names?
4. Is there a specific time associated with higher creation of tweets/ratings?

Q1) Most Common Dog Breeds

	breeds ⬍	percentage ⬍
golden_retriever	105	0.14
Pembroke	59	0.08
Labrador_retriever	54	0.07
Chihuahua	40	0.05
pug	38	0.05

The Top 5 most common breeds represent almost 40% of the dataset, having the beautiful Golden Retriever as the undisputed first place with 105 tweets submitted!

Q2) Most Liked Dog Breeds

Instead of simply looking at the number of times a specific breed has been rated, this time we investigate the breeds most retweeted and favorited.

In this particular case I decided to adopt the *median* instead of the mean, in order to avoid the influence of extreme values (like the swimming labrador retriever tweet that is clearly an outlier based on the number of retweets and favorites).

Author: Marcello Victorino

breed	median popularity	count popularity
black-and-tan_coonhound	39941.0	1
Newfoundland	27450.0	1
Greater_Swiss_Mountain_dog	25508.0	1
Afghan_hound	21276.0	1
Cardigan	20917.0	9
Leonberg	19562.0	2
bloodhound	19391.0	2
Great_Pyrenees	16990.0	8
Border_terrier	16594.0	2
Staffordshire_bullterrier	16287.0	7

Although I am a big Labrador and Golden retriever fan, this approach seems more accurate to me... and it leads to an ugly truth:

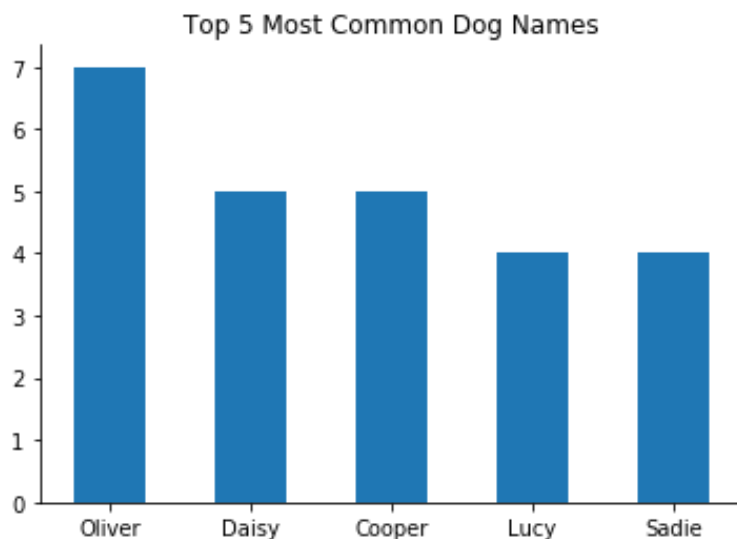The majority of people love **huge-clumsy-goofy-looking dogs**! 😂

Particularly, the Top 4 most liked dog breeds are:

1. Black and Tan Coonhound
2. Newfoundland
3. Greater Swiss Mountain Dog
4. Afghan Hound
5. Having a **Small-funny-looking** Cardigan in the Fifth Place - well deserved!

Q3) Most Popular Dog Names
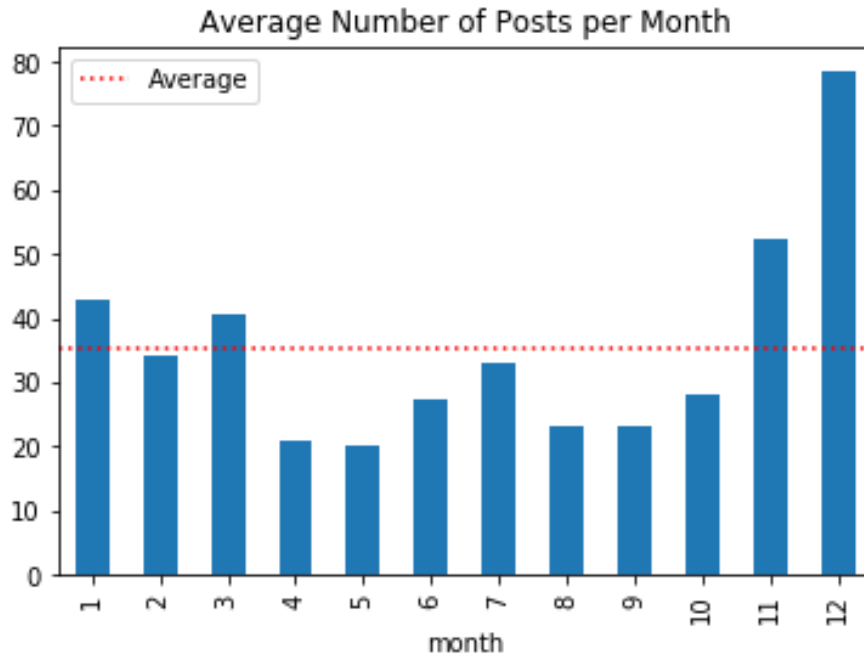
Looking at the count of values in the column 'name', we find the Top 5 most common dog names:

1) Oliver
2) Daisy
3) Cooper
4) Lucy
5) Sadie

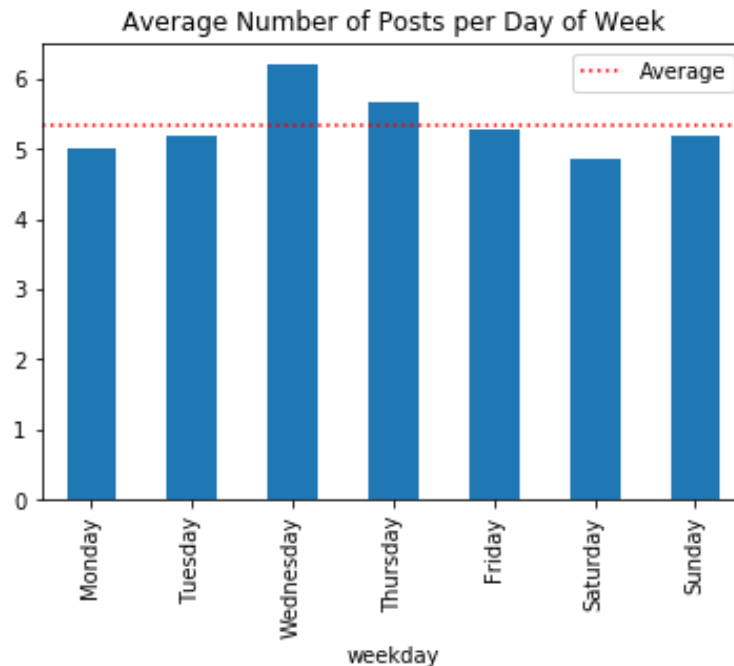Top 5 Most Common Dog Names

Author: Marcello Victorino

Q4) Most Typical Tweet Date/Time

Grouping the number of tweets by month and year, I calculate the average amount of tweets posted per month (considering data from 2015 – 2017):
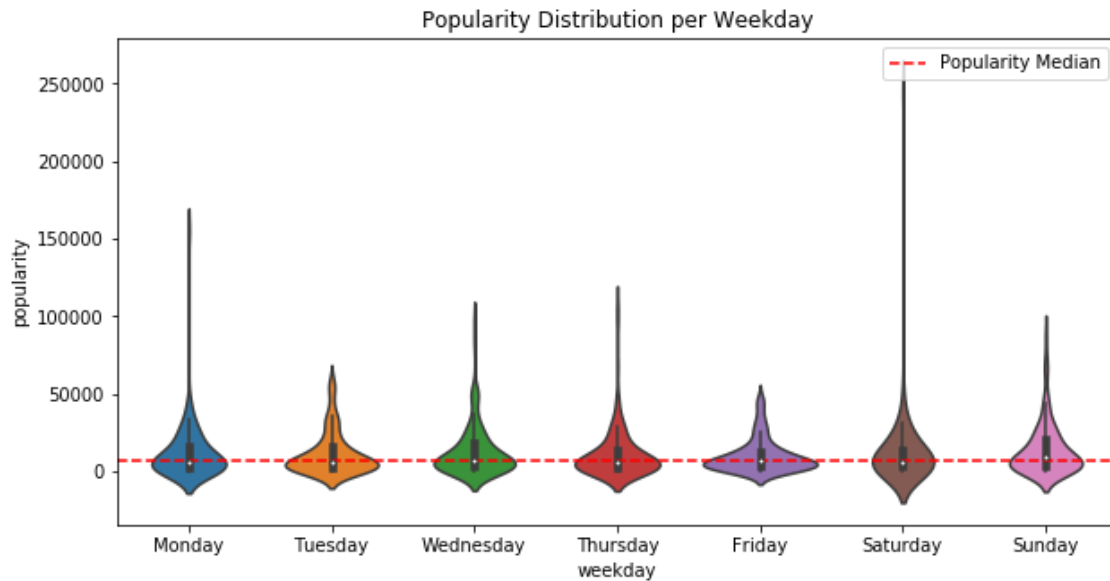

Looks like the busiest months are from **November through March**, where the amount of posts is above the average of 35 post per month.
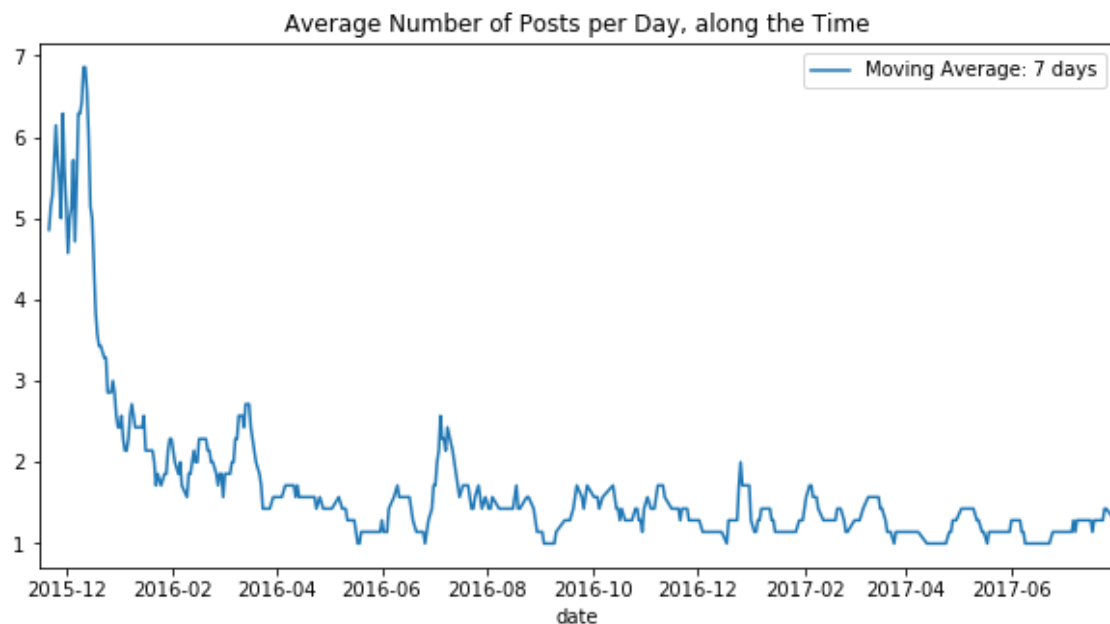

Interestingly, the busiest days are not the weekend! As a matter of fact, **Wednesday and Thursday** are typically the week days with above the average number of posts.

Author: Marcello Victorino

Popularity Distribution per Weekday

The popularity doesn't seem to be influenced by the day of the week. They all present very similar distribution, having their median close to the overall popularity median (6,600). That is, the sum of retweets and favorites.

Average Number of Posts per Day, along the Time

Lastly, plotting the average number of posts per day, it is possible to see the diminishing trend along the time.

Some possible explanations would be that, perhaps, the *WeRateDogs* is losing their initial momentum and losing popularity; or the author behind the channel decided to limit the ratings to about one per day.