Report of Anaylsis

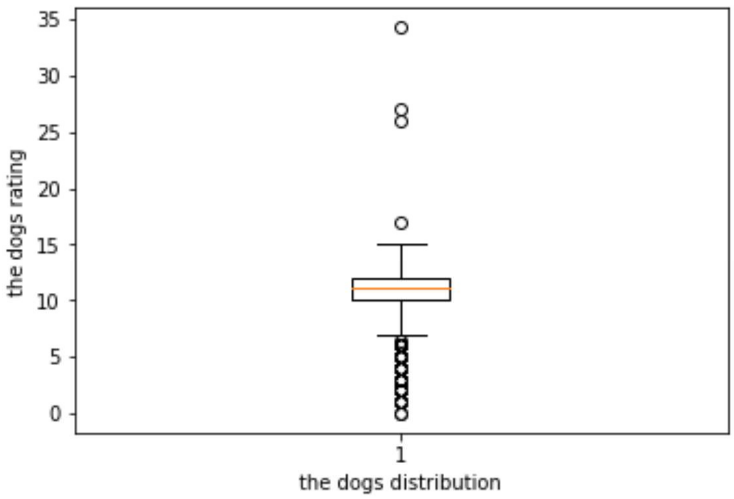
The Final dataframe (csv file) consists of 14 variables that is :

- Tweet_id: the id of the tweet that can be seen in the URL of each tweet.
- Times_stamp: the time and date of the tweet.
- Text: the content of the tweet.
- Name: the name of the dog.
- Breed: the breed of the dog.
- Stage: the current stage [*doggo, pupper, puppo, and floof(er)*]
- Rating: out of 10 rating and it could be more actually which makes this rating vendor more popular.
- Jpg_url: image of the dog
- Favorite_count: count of likes and favorites
- Retweet_count: count of retweets.
- Following_count: number of friends of the user.
- Application_type: whether it's an Iphone-Android-Web app.
- Tweet_url: the tweet URL.

Now I have a couple of questions towards this dataset that is :

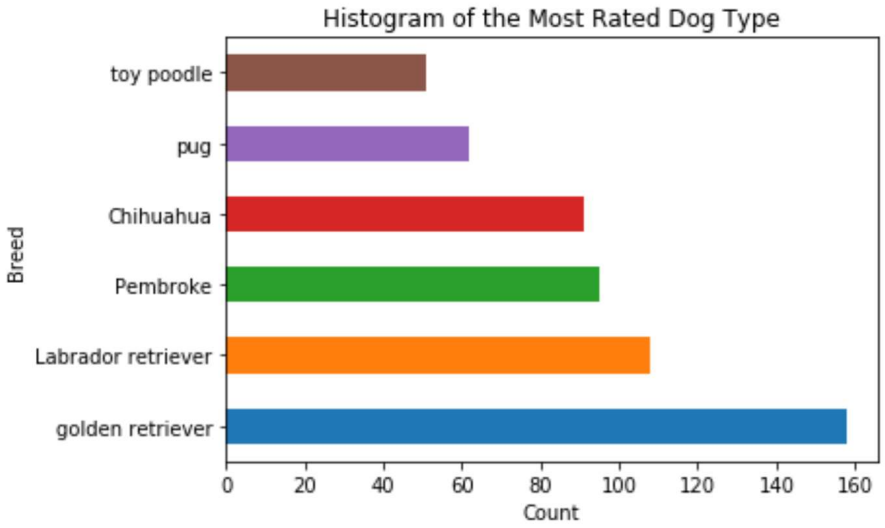
1- What is the distribution of the ratings? What is the average of the ratings? Probably 9 or 10? Lets plot and get summaries!?

After using Matplotlib I got the following BoxPlot.



And using Pandas I got that the average of ratings is **12.**

2- What is the most rated breed of dogs?



Using Matplotlib Python Package and drawing a Horizontal Bar Chart, I got that the most common breed is the Golden Retriever.

Let's dig deeper and get some information about that breed.

3- What are the summary stats of the Golden retriever.

Using Pandas package I got the following.

- There were 158 of Golden Retriever in our dataset that has around 2K of records.
- the golden retriever is the most common dog with average rating equals to 11.7.
- the maximum rating was 34/10 which means that it's a so lovely dog, isn't it?

4- let's look for some relationships, the maximum rating number associated with which dog?

Using Pandas library and it's search functionality we got that it's called Sam and guess what!

It's a Golden Retriever dog. Here is his pic.