Act Report

Hello everyone, who is reading this report today we will discuss the insights that I got from the dataset that was talking about the twitter archive for 'WeRateDogs' page that contain the dog names , dog breeds , the rating of each dog, dog stage , etc. And that made me ask the following insights

1- Which dog breed get the most rating on this page
2- What was the time that people used to post their dog images the most
3- What is the type of captions that got more high ratings

Let's talk first about the first insight 'Which dog breed get the most rating on this page'

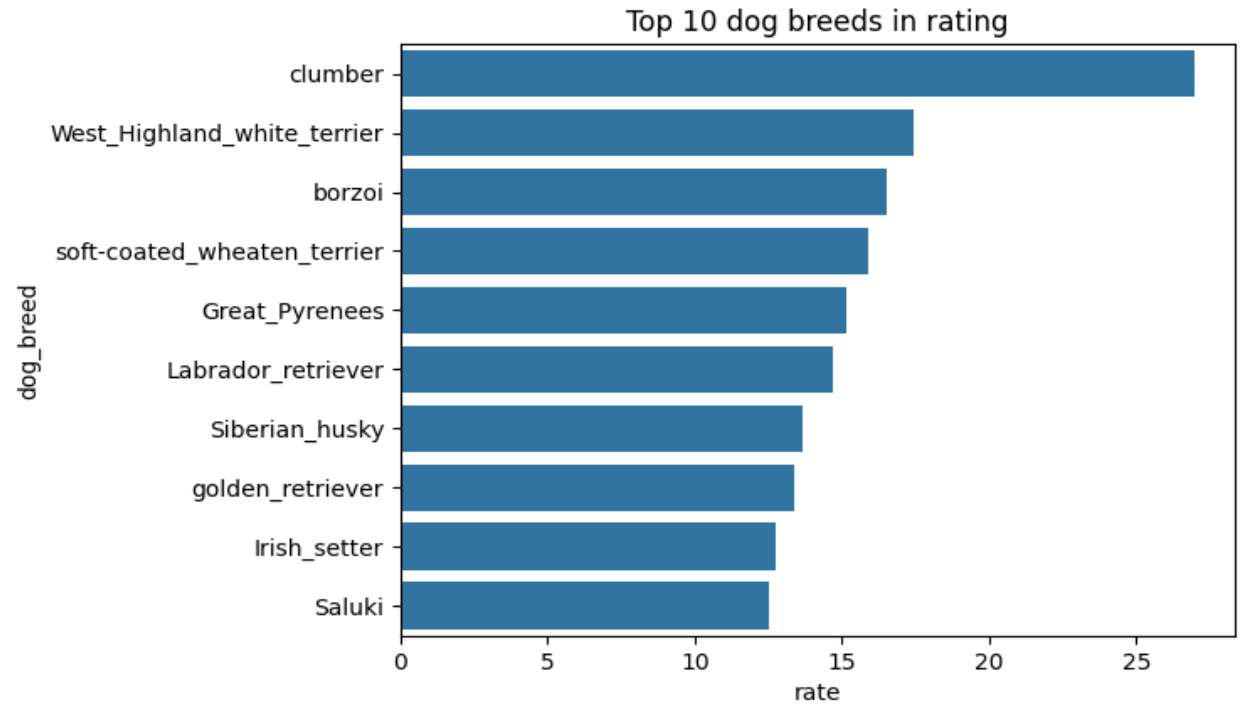
I wanted to know if there are dog breeds that look cuter or more beautiful than other.

First I got the average rating of each breed which happened using this code

```
# This code will give me the average rating for each breed

avg_rating = twitter_archive_master.groupby('dog_breed')['rate'].mean().reset_index()
```

then I made a bare chart from this variable to see the results below:



This bar chart tells us that clumber dog had the on average more than 25/10, which is quit obvious this dog breed look very cute

Act Report

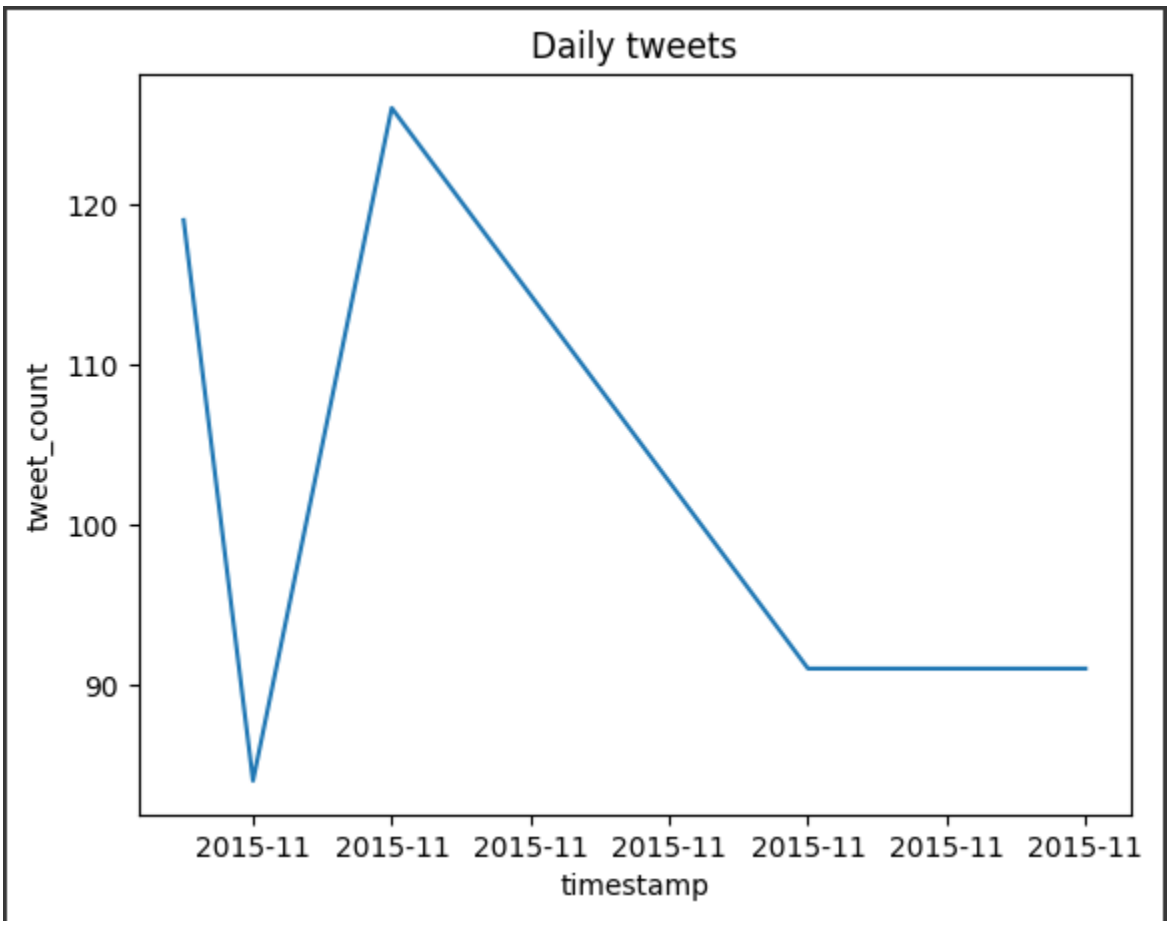
Let's move now to the second insight ‘What was the time that people used to post their dog images the most’

I wanted to know if there was an era that people used to post their dogs photos the most

To get this result first I had to have the number of tweets per day then I sorted them to the most 5 times as shown in the code below.

```
# first I will creat a dataframe that contain all timestamp with the tweet count in this time
daily_tweets = twitter_archive_master.groupby(
    twitter_archive_master['timestamp'].dt.date).size().reset_index(name='tweet_count')
top_5_times = daily_tweets.sort_values('tweet_count' , ascending = False).head(5)
```

Then I made a line plot to show the result of this as below:



This tells us that the most time that people used to post their dog’s image was in the around the beginning of November 2015

Act Report

Now lastly let me talk about the third and the last insight I found in this dataset ‘What is the type of captions that got more high ratings’

I wanted to know if type of captions on the tweet affects the rating of the dog

First, I used TextBlob library so I can make a function that detect the type of the captions and separate it to a new column in the dataframe as in the code below:

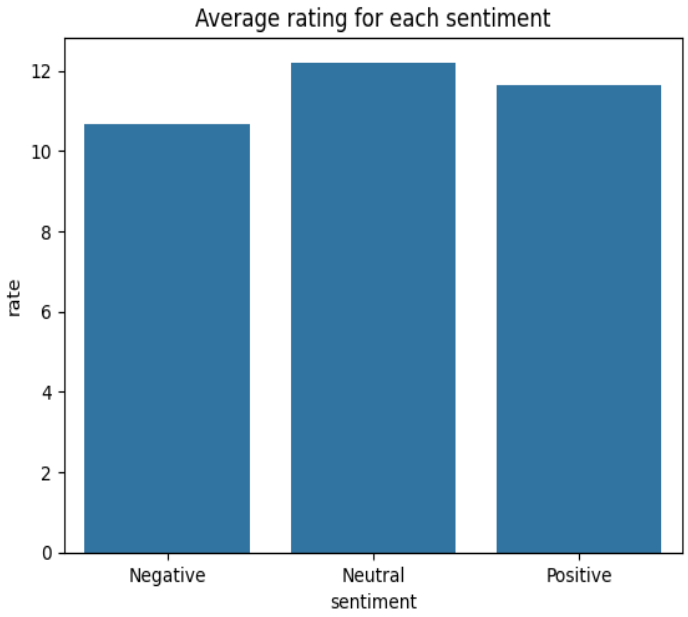
```
def get_sentiment(text):
    blob = TextBlob(text)
    polarity = blob.sentiment.polarity
    if polarity > 0:
        return 'Positive'
    elif polarity < 0:
        return 'Negative'
    else:
        return 'Neutral'

twitter_archive_master['sentiment'] = twitter_archive_master['text'].apply(get_sentiment)
```

Then I had to get the average rating for each sentiment as below:

```
avg_rating_for_sentiment = twitter_archive_master.groupby('sentiment')['rate'].mean().reset_index()
```

Then I made a bar chart form it



That tells us that Neutral captions almost get the most of high rates.