Act Report

Introduction:

This report shows the analysis part of the wrangled data previously. We will discuss three insights related the master dataframe collected.

## 1) On average, what dog stage/type considered the best based on We rate dogs rating?

As shown below we can see the difference is considered very small between all dog type/stage. By numbers we can see that **puppo and floofer** are considered the best dog type/stage based on we rate dogs twitter account rating.

```python
doggo_mean = master_df.query('dog_stage == "doggo"')['rating_numerator'].mean()
puppo_mean = master_df.query('dog_stage == "puppo"')['rating_numerator'].mean()
pupper_mean = master_df.query('dog_stage == "pupper"')['rating_numerator'].mean()
floofer_mean = master_df.query('dog_stage == "floofer"')['rating_numerator'].mean()

print(doggo_mean)
print(puppo_mean)
print(pupper_mean)
print(floofer_mean)
```

```
11.88888888888889
12.0
10.6381592039801
12.0
```

```python
fig = plt.figure()
ax = fig.add_axes([0,0,1,1])
dog_type = ['doggo', 'puppo' ,'pupper', 'floofer']
rating_avg = [doggo_mean,puppo_mean,pupper_mean,floofer_mean]
ax.bar(dog_type,rating_avg)
plt.show()
```
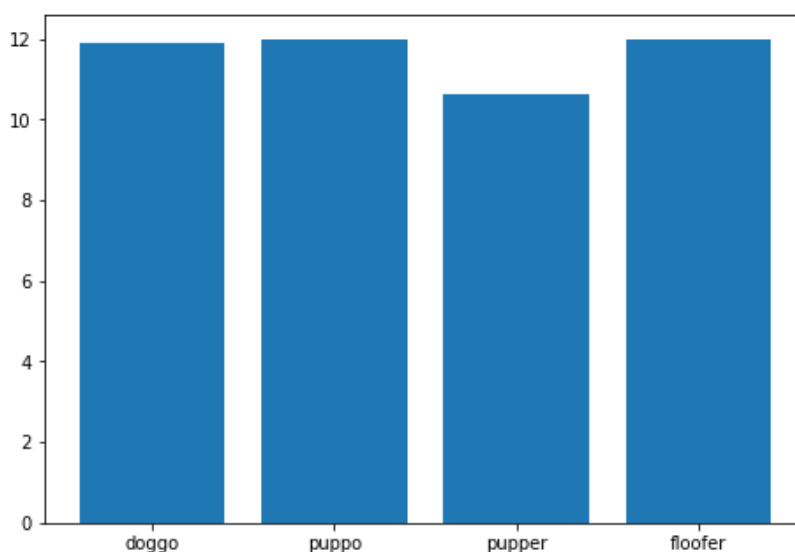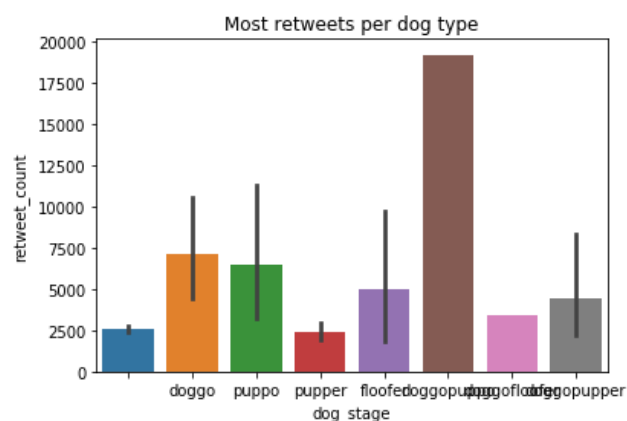

2) based on dog type, who has the most retweets?

After collecting tweet retweet count and merge them with master dataframe. We applied the seaborn library to visualize our results. Based on bar chart given below, we can answer this question by saying that **pupper** is considered the dog with highest number of retweets.

```
dog_stage
                  1668
doggo               63
doggofloofer         1
doggopupper          8
doggopuppo           1
floofer              7
pupper             201
puppo               22
Name: retweet_count, dtype: int64
```

```python
sns.barplot(x="dog_stage", y="retweet_count", data=master_df).set_title("Most retweets per dog type")
```


3) based on dog type, who has the most favorites?

Final insight, based on dog type we can see what dog type get the most favorites? In comparison between the forth types of dogs. As shown in the figure below, we can see that **pupper** is considered the highest favorited dog type/stage.

```
master_df.groupby('dog_stage')['favorite_count'].count()
```

```
: dog_stage
                   1668
   doggo            63
   doggofloofer      1
   doggopupper       8
   doggopuppo        1
   floofer           7
   pupper          201
   puppo            22
   Name: favorite_count, dtype: int64
```

```
sns.barplot(x="dog_stage", y="favorite_count", data=master_df ).set_title("Most favorites per dog type");
```
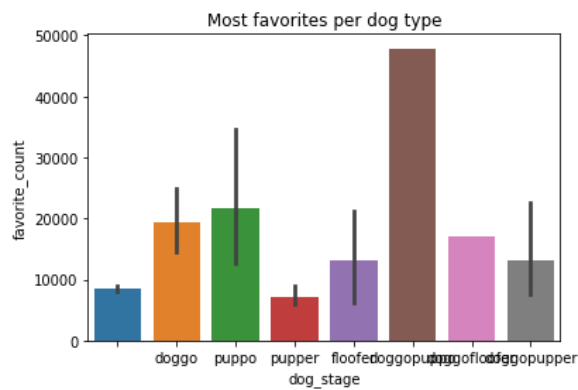

Conclusion:

At the end I would like to share some limitations in this project I personally faced:

1- There are free APIs such as Wikipedia we can use to get data from it.
2- I could use beautiful soup for example to get tweets data, but I thought that solution is not the best for me and you in order to retrieve the data.

Thank you at the end of this project it is hard but I have enjoyed each part of it.