act_report

June 1, 2022

0.1 WeRateDogs - Insights report

- In this report I will share some insights based of the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog.

In [6]:
```
import pandas as pd

import matplotlib.pyplot as plt
```

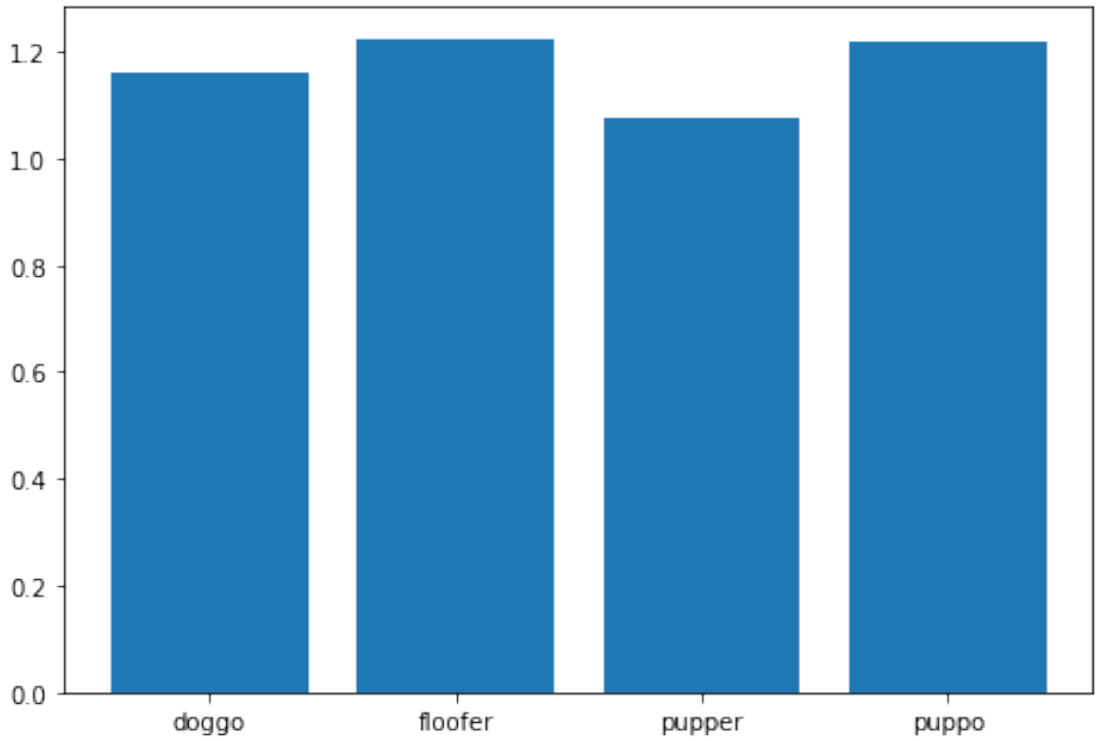
In [7]:
```
df_merge=pd.read_csv('twitter_archive_master.csv')
```

0.1.1 Visualization - What's the most liked category of dogs?

To find out what's the most liked categosy of dogs I used the rating system from weratedog . Basically the tweets are rated with a custom rating system having a rating numerator and a rating denominator . For my investigation I calculated the rating numerator/rating denominator and used it as a "likeiness" metric.

In [8]:
```
doggo_mean = (df_merge.query('dog_stage == "doggo"')['rating_numerator']/df_merge.query(
puppo_mean = (df_merge.query('dog_stage == "puppo"')['rating_numerator']/df_merge.query(
pupper_mean = (df_merge.query('dog_stage == "pupper"')['rating_numerator']/df_merge.quer
floofer_mean =( df_merge.query('dog_stage == "floofer"')['rating_numerator']/df_merge.qu

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()
```



Conclusion: Floofer is the most liked category of dogs closely followed by puppo

0.1.2 What is tweet most added to the favorites ?

The favourite count is a way to measure the success of a tweet so I looked for the tweets with the biggest favourite count to find out what was the tweet most added to te favourites.

```
In [9]: df_merge.sort_values(['favorite_count'], ascending =False ).head(3)
```

```
Out[9]:          tweet_id            timestamp              source  \
        1011  7.474395e+17  2016-06-27 14:40:26                Vine
        395   8.255351e+17  2017-01-29 02:44:34  Twitter for iPhone
        515   8.113868e+17  2016-12-21 01:44:13  Twitter for iPhone

                                                    text  \
        1011  This is Linus. He just wanted to say hello but...
        395   Here's a very loving and accepting puppo. Appe...
        515   This is Craig. That's actually a normal sized ...

                                            expanded_urls  rating_numerator  \
        1011                    https://vine.co/v/5uTVXWvn3Ip              12.0
        395   https://twitter.com/dog_rates/status/825535076...              14.0
        515   https://twitter.com/dog_rates/status/811386762...              11.0
```

```
      rating_denominator   name  contributors  coordinates     ...       \
1011                10.0  Linus           NaN          NaN     ...
395                 10.0   None           NaN          NaN     ...
515                 10.0  Craig           NaN          NaN     ...

                p1   p1_conf p1_dog              p2   p2_conf p2_dog  \
1011  Maltese_Dog  0.780187   True  Dandie_Dinmont  0.074429   True
395         Wombat  0.880257  False            Corn  0.019421  False
515          Teddy  0.098283  False      Toy_Poodle  0.098029   True

                   p3   p3_conf  p3_dog  dog_stage
1011  Norfolk_Terrier  0.033776    True        NaN
395               Pug  0.019044    True      puppo
515   Shopping_Basket  0.077852   False     pupper

[3 rows x 53 columns]
```

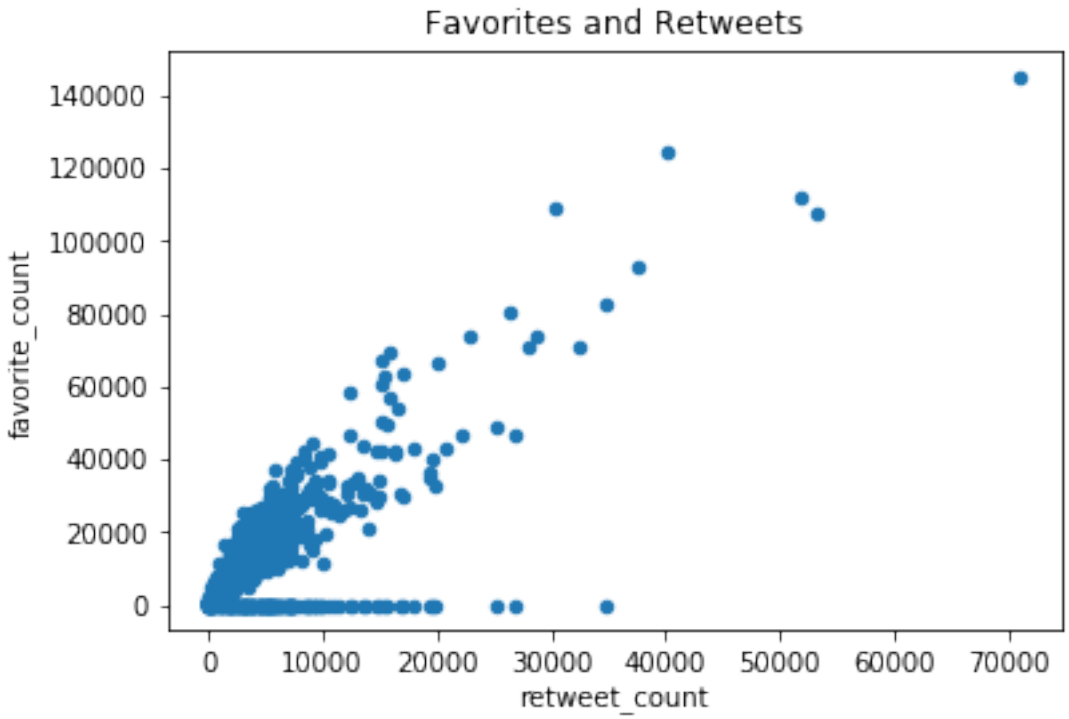
Conclusion: The Linus tweet from 2016 was marked by the most people as their favourite

0.1.3 Favourites vs retweets

There are different ways to express apreciation for a tweet, two of them are adding it to the favourites and share it with your followers (retweet). Are those two related ? A tweet that has a high favourite count will also have a high retweet count ? See conclusion bellow .

```
In [10]: df_merge.plot.scatter(x = 'retweet_count', y = 'favorite_count', title="Favorites and R
```

```
Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x7f3d684a7eb8>
```



Conclusion : There is a correlation between favourite_count and retweet_count

In [ ]: