

# Data insights

My first insight was looking at the ratings of the different dogs. For this I used the `.value_counts()` function.

```
1 twitterAnalysis['rating_numerator'].value_counts()

12    483
10    433
11    413
13    283
9     152
8      98
7      51
14     38
5      34
6      32
3      19
4      15
2       9
1       4
0       1
```

This tells us that the dogs were most commonly rated 12/10, and the top 5 most common ratings being 12, 10, 11, 13, and 9. I then wanted to find out what percentage of dogs were given the highest rating so far of 14/10.

```
1 (twitterAnalysis[twitterAnalysis['rating_numerator'] == 14].shape[0]
2 /twitterAnalysis.shape[0])*100

1.8401937046004841
```

With the code above, I found out that just under 2% of all dogs were given the highest rating of 14, making this rating quite a rarity.

Once I had done this, I wanted to look at the dogs name and see what the most common names were.

```
1 twitterAnalysis['name'].value_counts()[:20]

None    587
a        54
Oliver   10
Lucy     10
Cooper   10
Charlie  10
Penny     9
Tucker    9
Lola       8
Winston   8
the        8
Sadie      8
Daisy      7
Toby       7
Koda       6
Oscar      6
Jax        6
Stanley    6
Bo         6
Bailey     6
```

Hmm, that's weird, the top 2 most common names are 'None' and 'a', I highly doubt that many people named their dogs this. After looking into some of the tweets that matched a name of 'None' I realised that the dog's name wasn't actually 'None' but rather the owner had not sent the dog's name as part of the tweet. That doesn't explain the name 'a' though. Well, as it turns out, the formula for collecting the dog's names follows a pattern similar to 'This is...' or 'Meet...'. Therefore, if the tweet started off with 'This is a golden retriever...', the name would be placed as 'a'. Regardless, looking at the most common names after these show quite common names such as 'Cooper' and 'Lucy'.

After this, I looked at the different dog types (doggo, floofer, pupper, and puppo), and use `.value_counts()` to see which category was the most common, and it turns out that 'pupper' is the most common dog type.

```
1 twitterAnalysis['doggo'].value_counts()
```

```
None      1983
doggo      82
Name: doggo, dtype: int64
```

```
1 twitterAnalysis['floofer'].value_counts()
```

```
None      2055
floofer    10
Name: floofer, dtype: int64
```

```
1 twitterAnalysis['pupper'].value_counts()
```

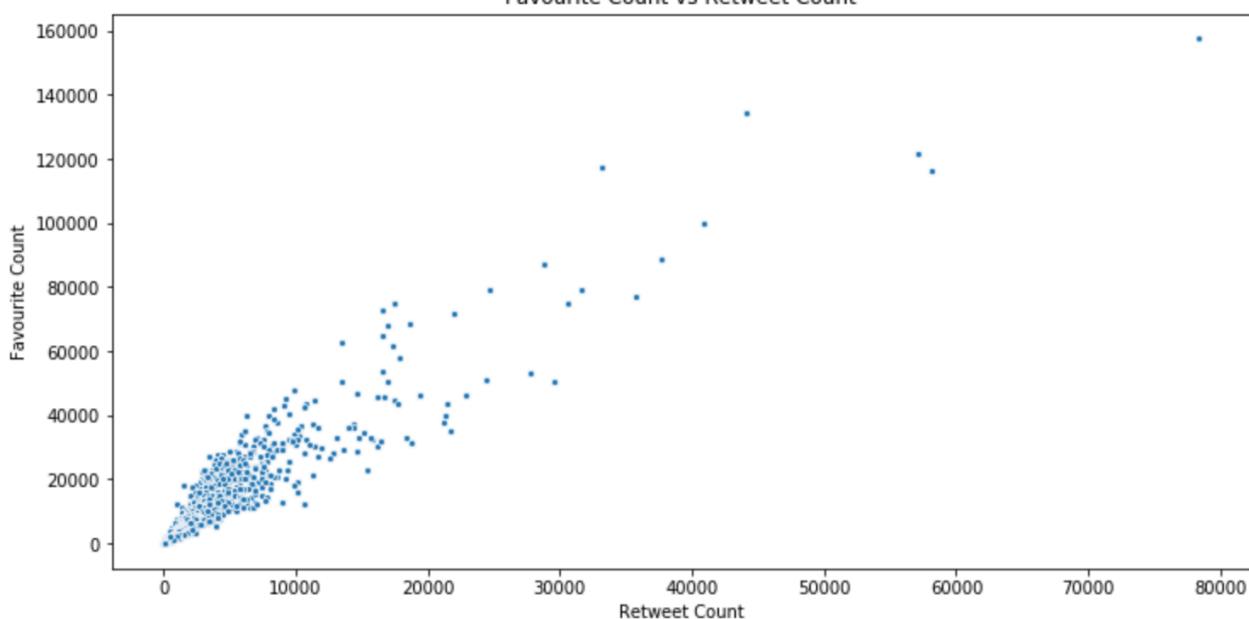
```
None      1837
pupper    228
Name: pupper, dtype: int64
```

```
1 twitterAnalysis['puppo'].value_counts()
```

```
None      2041
puppo      24
Name: puppo, dtype: int64
```

I then made a scatter graph using Seaborn to see the correlation between retweet count and favourite count and saw that there was quite a strong positive correlation between the two. I then calculated the correlation coefficient, and got a very strong coefficient of 0.926. Most of the tweets were gathered under the 10000 retweet count and 20000 favourite count, with a tweet often receiving double the amount of favourites as retweets.

Favourite Count vs Retweet Count



```
1 twitterAnalysis.corr()['retweetCount']['favouriteCount']
```

```
0.9262776964099702
```