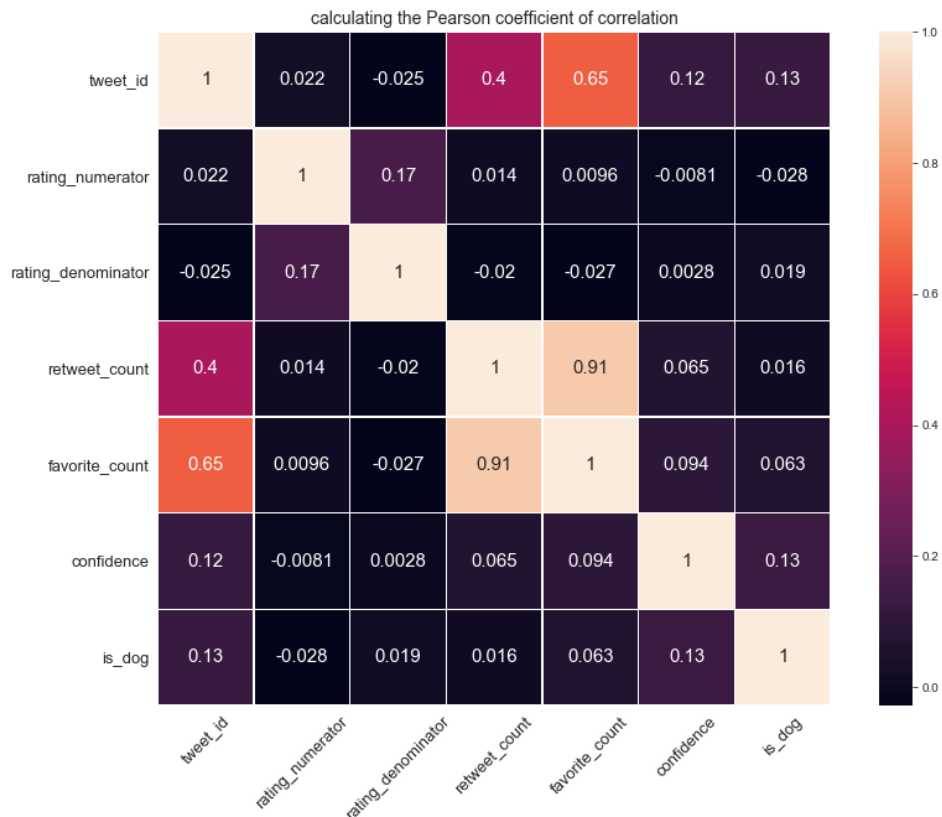# Wrangle and Analyzing Data Act Report
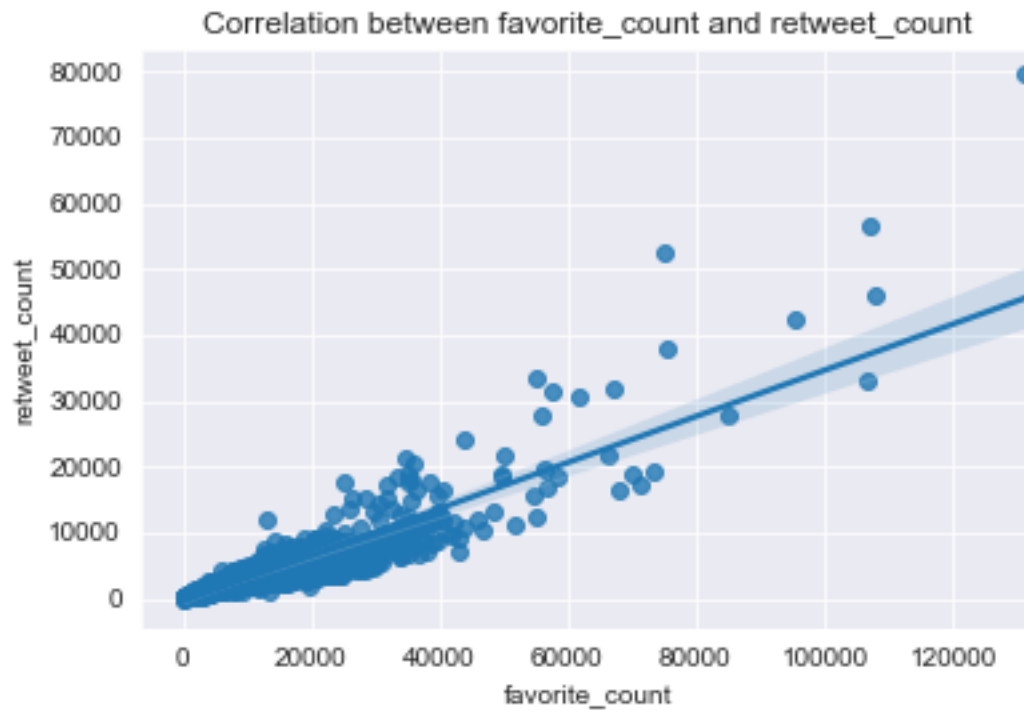
STUDENT: ABDULLAH ABUSABBA

Analysis:

The below heatmap shows the Pearson coefficient correlation between each variable in the data frame and as we can see the retweet and favorite have a correlation of 91%.
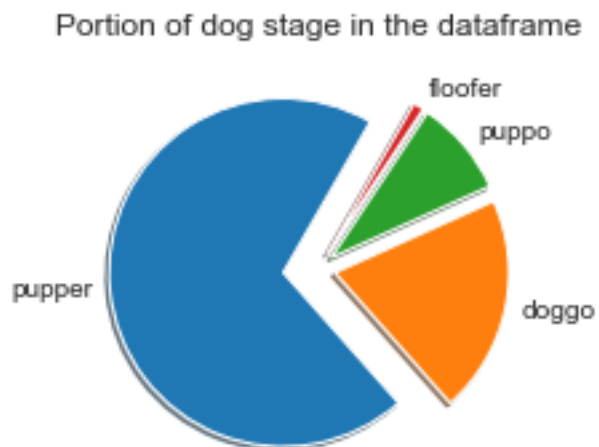

calculating the Pearson coefficient of correlation

	tweet_id	rating_numerator	rating_denominator	retweet_count	favorite_count	confidence	is_dog
tweet_id	1	0.022	-0.025	0.4	0.65	0.12	0.13
rating_numerator	0.022	1	0.17	0.014	0.0096	-0.0081	-0.028
rating_denominator	-0.025	0.17	1	-0.02	-0.027	0.0028	0.019
retweet_count	0.4	0.014	-0.02	1	0.91	0.065	0.016
favorite_count	0.65	0.0096	-0.027	0.91	1	0.094	0.063
confidence	0.12	-0.0081	0.0028	0.065	0.094	1	0.13
is_dog	0.13	-0.028	0.019	0.016	0.063	0.13	1

Insight-1)

For each increase in retweet, favorite count increases as well.

Correlation between favorite_count and retweet_count

Insight-2)

Most tweet pictures have a pupper dog stage in we WeRateDogs Twitter Data. Not all rows have dog stage but the count that we have for each are as following:

```
pupper      242
doggo        74
puppo        30
floofer       4
```

Portion of dog stage in the dataframe

Insight-3)

Most true image prediction are shown in the below figure, as we can see golden retriever is the most predicted image with 132 which is all golden retriever images.

most true dog image predictions