WeRateDogs

Human most beloved friend

Name: Mark George Louiz

The following insights came from the master archive data set with 1971 entries with main columns "tweet id " , "numerator rating" , "dog breed " , " favorite count" and "retweets count" which made with archive data set merge with twitter Api query after cleaning data in both .

Insights :

First: the most likable images with average rating 11 between the rating from 0 to 14 in numerator for each dog in the image and for high numerator rating frequency of ratings become smaller .

Second : image with high ratings of numerator gets high number of retweets and favorites count and with manual and charts investigation may be that because most images with more than one dog specially with two different kind such doggo and puppo .

Third : pupper has the most tweets also theses images get less favorite count and retweets than doggo , puppo and floofer with more manual investigation I notice that pupper gts less favorite count and retweets than the others even if he got the same numerator and denominator ratings similar to doggo , puppo ,floofer .

We can conclude from that images with less frequency dog stage gets more retweets and favorite counts .

Statistics :

```
|: df_archive_master.describe()
```

	retweet_count	favourite_count	rating_numerator	rating_denominator
count	1971.000000	1971.000000	1971.000000	1971.000000
mean	2784.449518	8949.106545	12.223237	10.455099
std	4697.662893	12267.799790	41.634034	6.789092
min	16.000000	81.000000	0.000000	7.000000
25%	628.500000	1997.000000	10.000000	10.000000
50%	1367.000000	4147.000000	11.000000	10.000000
75%	3239.000000	11402.500000	12.000000	10.000000
max	79515.000000	132810.000000	1776.000000	170.000000

```
In [115]: df_archive_master.retweet_count.count()
Out[115]: 1971
```

```
In [116]: df_archive_master.groupby(['rating_denominator','rating_numerator'])['favourite_count'].mean()
```

```
Out[116]: rating_denominator  rating_numerator
          7                   24                   5927.000000
          10                  0                   24574.000000
                              1                    5599.800000
                              2                    2256.777778
                              3                    2395.526316
                              4                    2842.937500
                              5                    3950.424242
                              6                    2716.968750
                              7                    3462.173077
                              8                    2450.905263
                              9                    2536.560000
                              10                   4141.103118
                              11                   6919.831633
                              12                  11343.114350
                              13                  23229.588933
                              14                  29950.939394
                              26                   1879.000000
                              27                   7320.000000
                              50                   2574.000000
                              75                  20296.000000
                              420                  7989.000000
                              1776                 5569.000000
          40                  44                   3748.000000
          50                  45                   2634.000000
                              60                   3201.000000
          70                  84                  13518.000000
          80                  80                   2529.000000
                              88                   2927.000000
          90                  99                   3062.000000
          110                 121                  4225.000000
          120                 144                  3323.000000
          150                 165                  5316.000000
          170                 204                  4196.000000
          Name: favourite_count, dtype: float64
```

```
In [117]: df_archive_master.groupby('dog_breed')['rating_numerator'].mean()
```

```
Out[117]: dog_breed
          doggo           11.888889
          doggofloofer    11.000000
          doggopupper     11.000000
          doggopuppo      13.000000
          floofer         12.000000
          pupper          10.716418
          puppo           12.000000
          Name: rating_numerator, dtype: float64
```

```
In [118]: df_archive_master.groupby('dog_breed')['rating_numerator'].count()
```

```
Out[118]: dog_breed
          doggo            63
          doggofloofer      1
          doggopupper       8
          doggopuppo        1
          floofer           7
          pupper          201
          puppo            22
          Name: rating_numerator, dtype: int64
```

```
In [129]: df_archive_master.groupby('dog_breed')['rating_numerator'].sum()
```

```
Out[129]: dog_breed
          doggo            749
          doggofloofer      11
          doggopupper       88
          doggopuppo        13
          floofer           84
          pupper          2154
          puppo            264
          Name: rating_numerator, dtype: int64
```

```
In [119]: df_archive_master.groupby('dog_breed')['retweet_count'].sum()
```

```
Out[119]: dog_breed
          doggo           448919
          doggofloofer      3433
          doggopupper      35178
          doggopuppo       19196
          floofer          34781
          pupper          478883
          puppo           142427
          Name: retweet_count, dtype: int64
```

```
In [132]: df_archive_master.groupby('rating_numerator')['retweet_count'].count()
```

```
Out[132]: rating_numerator
          0            1
          1            5
          2            9
          3           19
          4           16
          5           33
          6           32
          7           52
          8           95
          9          150
          10         417
          11         392
          12         446
          13         253
          14          33
          24           1
          26           1
          27           1
          44           1
          45           1
          50           1
          60           1
          75           1
          80           1
          84           1
          88           1
          99           1
          121          1
          144          1
          165          1
          204          1
          420          1
          1776         1
          Name: retweet count, dtype: int64
```

```
In [120]: df_archive_master.groupby('dog_breed')['favourite_count'].sum()
```

```
Out[120]: dog_breed
          doggo           1219452
          doggofloofer      17169
          doggopupper      105759
          doggopuppo        47844
          floofer           92442
          pupper          1457356
          puppo            474806
          Name: favourite_count, dtype: int64
```

visualization:

```
In [126]: df_archive_master.groupby(['dog_breed'])['favourite_count'].mean().plot.bar(stacked=True)
          plt.title('dog breed favourite tweets numbers')
          plt.xlabel('dog breed')
          plt.ylabel('tweets likes numbers')
```

```
Out[126]: Text(0, 0.5, 'tweets likes numbers')
```
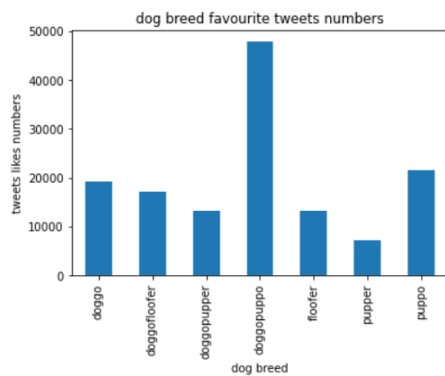
```
In [127]: df_archive_master.groupby(['dog_breed'])['rating_numerator'].mean().plot.bar(stacked=True)
          plt.title('dog breed rating')
          plt.xlabel('dog breed')
          plt.ylabel('rating numerator')

Out[127]: Text(0, 0.5, 'rating numerator')
```
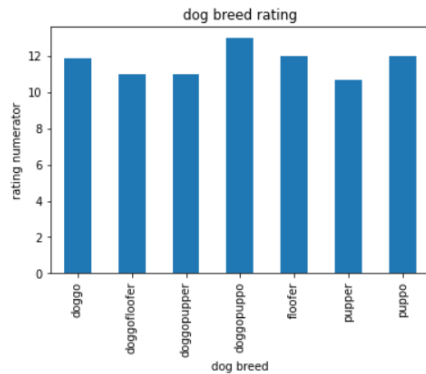

Final conclusions :

The tweets id with less frequency tweets gets more favorite counts and retweets which means the got better quality content we further manual conclusion with tweets that may depend on the quality of the image the dogs position .