Data-Analysis-NanoDegree-Udacity-Project-4

Project-done-by-Cherif.Arsanious

The report is the first page only, the remaining of the pages are just the code used

Analysis and Visualizations

After gathering, assessing, cleaning and storing our data. It is time now to do some analysis and visua
our data about the twitter account WeRateDogs.

By running the describe method, I observed the following thing:

- Favorites (or like) column has a minimum of 72 likes and it goes to tweet_id 666102155909144
lowest retweet of 11 retweets only. The dog is english_setter and actually it did get a high rating
8400 likes. It has a maximum of 157953 and it goes to to tweet_id 744234799360020481 and a
retweets number of 78695. The highest likes went to a labrador_retriever in doggo stage with a
- The highest rating in our wrangled data is 14/10 and it belongs to 33 tweets. The lowest rating v
tweets_id of the the lowest and hightest ratings are found in the code cells). The average rating
on the rating_numerator I found that 12/10 is the mode rating with a frequency of 444 tweets. T
rating are pembroke(4) french_bulldog(2),and golden_retriever(2)

By running the corr() method on the twitter_archive_master data frame, I observed:

- Weak postive pearson correlation between retweets count and rating_numerator with a value 0.
the account admin doesn't guarantee that the followers will retweet these posts and the other w
- Moderate postive pearson correlation between favourites count and rating_numerator with a val
dog by the account admin doesn't guarantee that the followers will like these posts and the othe

By plotting a scatter plot between retweets count and favourites counts:

- There is a strong positive pearson correlation between favourites count and retweets counts wit
indicates only a relational but says not about causation, I can imply that one caused the other.

```
1 #Using describe() method to analyze our data
2 twitter_archive_master.describe()
```

	tweet_id	favorites	retweets	rating_numerator	rating_denominator
count	1.941000e+03	1941.000000	1941.000000	1941.000000	1941.0
mean	7.359180e+17	8399.435343	2492.583205	10.537352	10.0
std	6.762265e+16	12372.113161	4468.137245	2.162537	0.0
min	6.660209e+17	72.000000	11.000000	1.000000	10.0
25%	6.757816e+17	1785.000000	556.000000	10.000000	10.0
50%	7.084699e+17	3799.000000	1204.000000	11.000000	10.0
75%	7.881506e+17	10451.000000	2850.000000	12.000000	10.0
max	8.924206e+17	157953.000000	78695.000000	14.000000	10.0

```
1 twitter_archive_master.query('favorites==72')
```

	tweet_id	favorites	retweets	dog_name	dog_breed	stage	rating_nu
1924	666102155909144576	72	11	NaN	english_setter	NaN	

```
1 twitter_archive_master.query('favorites==157953')
```

	tweet_id	favorites	retweets	dog_name	dog_breed	stage	rating_
756	744234799360020481	157953	78695	NaN	labrador_retriever	doggo	

```
1 list(twitter_archive_master.query('rating_numerator==14')['tweet_id'])
```

```
[890240255349198849,
 887517139158093824,
 884441805382717440,
 881536004380872706,
 878057613040115712,
 870063196459192321,
 868880397819494401,
 864873206498414592,
 860184849394610176,
 856282028240666624,
 854120357044912130,
 852226086759018497,
 851464819735769094,
 841439858740625411,
 832273440279240704,
 831911600680497154,
 828650029636317184,
 828381636999917570,
 825535076884762624,
 822462944365645825,
 821407182352777218,
 820314633777061888,
 819006400881917954,
 819004803107983360,
 813812741911748608,
 807621403335917568,
 794205286408003585,
 778408200802557953,
 774314403806253056,
 762035686371364864,
 755206590534418437,
 742465774154047488,
 685547936038666240]
```

```
1 twitter_archive_master.query('rating_numerator==14')['dog_breed'].value_counts()
```

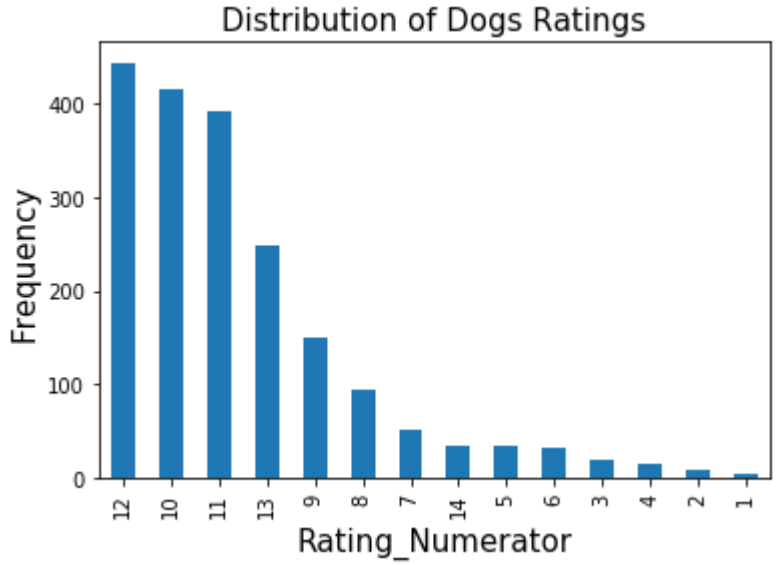
```
pembroke                  4
french_bulldog            2
golden_retriever          2
pomeranian                1
chihuahua                 1
gordon_setter             1
samoyed                   1
bloodhound                1
black-and-tan_coonhound   1
eskimo_dog                1
rottweiler                1
standard_poodle           1
bedlington_terrier        1
irish_setter              1
Name: dog_breed, dtype: int64
```

```
1 twitter_archive_master.rating_numerator.value_counts()
```

```
12    444
10    416
11    392
13    249
9     149
8      95
7      51
14     33
5      33
6      32
3      19
4      15
2       9
1       4
Name: rating_numerator, dtype: int64
```

```
1 twitter_archive_master.rating_numerator.value_counts().plot(kind='bar')
2 plt.xlabel('Rating_Numerator',fontsize=15)
3 plt.ylabel('Frequency',fontsize=15)
4 plt.title('Distribution of Dogs Ratings',fontsize=15)
```

```
Text(0.5, 1.0, 'Distribution of Dogs Ratings')
```

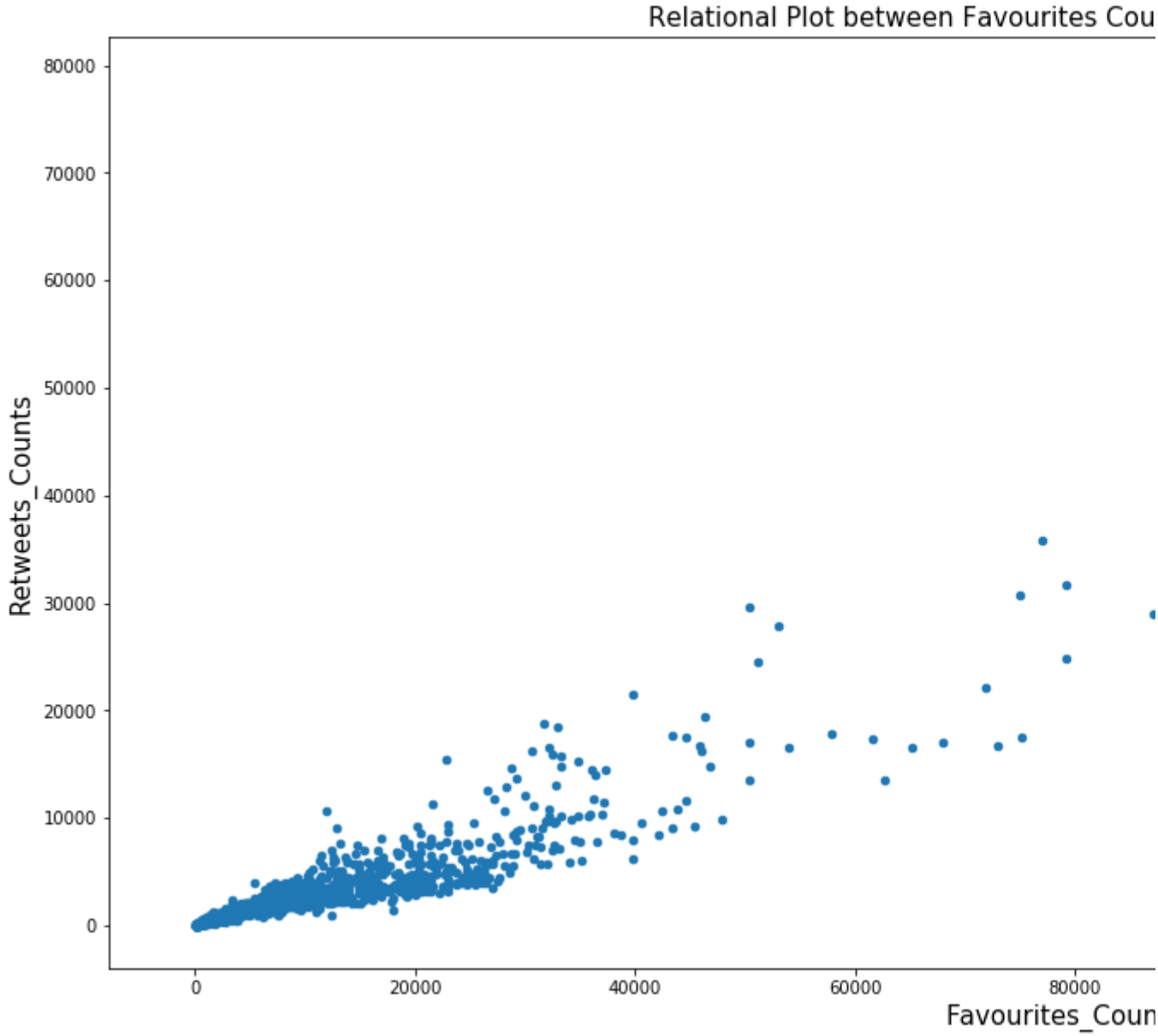



```
1 twitter_archive_master.corr()
```

	tweet_id	favorites	retweets	rating_numerator	rating_denominat
tweet_id	1.000000	0.613850	0.386979	0.537975	Na
favorites	0.613850	1.000000	0.929102	0.395391	Na
retweets	0.386979	0.929102	1.000000	0.297646	Na
rating_numerator	0.537975	0.395391	0.297646	1.000000	Na
rating_denominator	NaN	NaN	NaN	NaN	Na
img_num	0.213148	0.126056	0.101735	0.202592	Na
prediction_confidence	0.100789	0.071119	0.049980	0.102845	Na

```
1 twitter_archive_master.plot(x='favorites',y='retweets',kind='scatter',figsize=(20,10));
2 plt.xlabel('Favourites_Counts',fontsize=15)
3 plt.ylabel('Retweets_Counts',fontsize=15)
4 plt.title('Relational Plot between Favourites Counts and Retweets Counts',fontsize=15)
```

```
/usr/local/lib/python3.6/dist-packages/pandas/plotting/_matplotlib/core.py:420: Futur
        To accept the future behavior, pass 'dtype=object'.
        To keep the old behavior, pass 'dtype="datetime64[ns]"'.
  numeric_data[col] = np.asarray(numeric_data[col])
Text(0.5, 1.0, 'Relational Plot between Favourites Counts and Retweets Counts')
```




```
1 twitter_archive_master[['favorites','retweets']].corr()
```

	favorites	retweets
favorites	1.000000	0.929102
retweets	0.929102	1.000000

```
1 twitter_archive_master.plot(x='timestamp_x',y='rating_numerator',kind='line',figsize=(1
2 plt.xlabel('Time',fontsize=15)
3 plt.ylabel('Rating',fontsize=15)
```

```
4 plt.title('Relation bet. Time and Rating of Dogs',fontsize=15)
```

```
Text(0.5, 1.0, 'Relation bet. Time and Rating of Dogs')
```