act_report

June 27, 2022

0.0.1 STUDENT INFOMATION

- **Full Name:** Angel Newportright
- **Email:** angeltucker007@gmail.com
- **Country:** Nigeria

0.1 Report: act_report

- Create a **250-word-minimum written report** called "act_report.pdf" or "act_report.html" that communicates the insights and displays the visualization(s) produced from your wrangled data. This is to be framed as an external document, like a blog post or magazine article, for example.

0.2 ACT REPORT

The dataset for this project is the tweet archive of Twitter user @dog_rates, also known as WeRateDogs.

About The Ratings: These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

For effective reporting, we need to import the libraries that are needed

Libraries Import

[10]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

1. Wrangling the data required data conversion from one type to another.

For example, datetime column from the `twitter archive data` was indicated as an object, it was converted to `datetime` type. Tweet_id column was in `integer` format and was converted to `string` datatype

2. Wrangling the data required assessment of data tidyness, such as structural data problems

For example, some columns were droped that were considered irrelevant to our analysis, and others were merged to form a new table

0.2.1 Insights

For us to be able to view the visuals of our data we need to read the master data and plot the charts

[22]:
```
tweet_data = pd.read_csv('twitter_archive_master.csv', sep=';')

# Converting tweet_id to string from new dataframe
tweet_data.tweet_id = str(tweet_data.tweet_id)
```

Getting Summary Statistics of the data

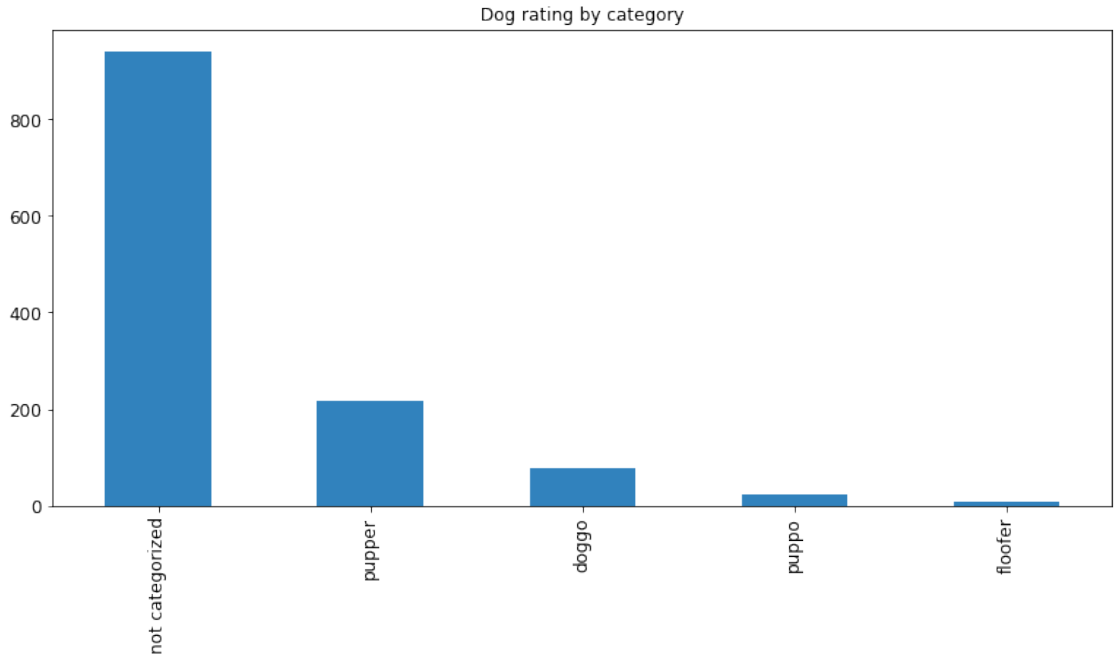
[23]:
```
tweet_data.describe()
```

[23]:

	rating_numerator	retweet_count	favorite_count
count	1268.000000	1268.000000	1268.000000
mean	11.151420	4127.621451	10772.328076
std	1.969565	7795.691771	16374.258849
min	3.000000	83.000000	0.000000
25%	10.000000	885.000000	2660.000000
50%	11.000000	1846.000000	5005.000000
75%	12.000000	3917.000000	12376.000000
max	27.000000	79515.000000	132810.000000

Plotting a bar chart for the category column

[25]:
```
tweet_data.category.value_counts().plot(kind= 'bar', figsize = (13,6),␣
↪fontsize=12, colormap='tab20c')
plt.title('Dog rating by category');
```

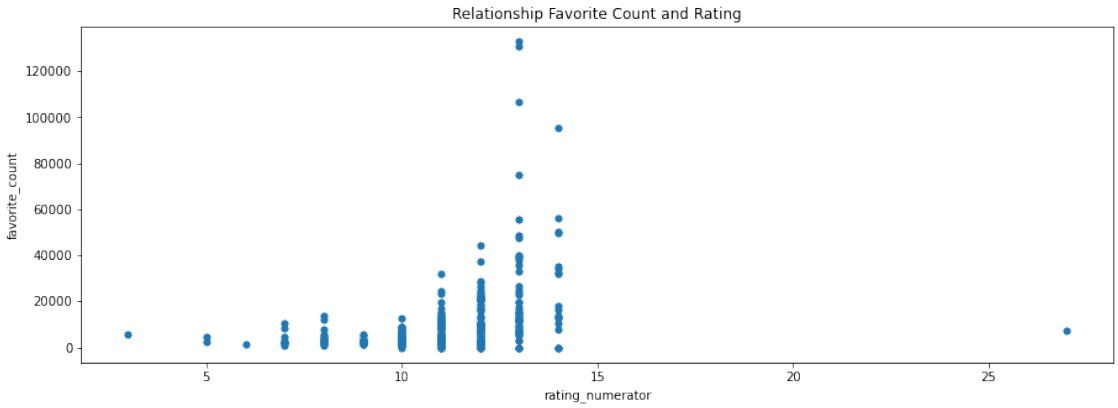


From our bar chart above we can clearly see that while all of the Dogs rated by WeRateDogs may be rated numericaly, majority of the Dogs rated are not categorized

We can also see that majority of the Categorized Dogs fall under the `pupper` category

Plotting Scaterplots to find correlations i. Finding the correlation between the rating a dogs get and the number of favorite it gets

[31]:
```
tweet_data.plot(x= 'rating_numerator',y='favorite_count', kind = 'scatter',␣
↪figsize=(15,5))
plt.title('Relationship Favorite Count and Rating');
```

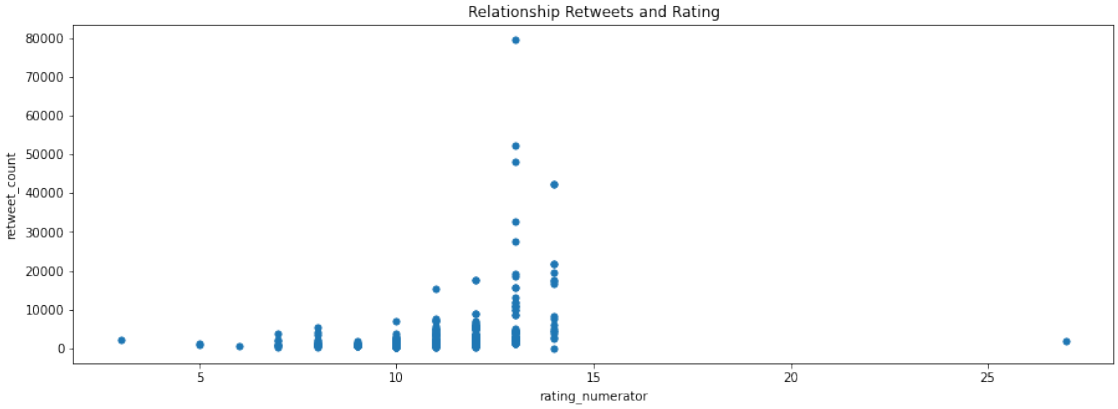


From our scatter plot above we can see that there isn't correlation between the rating a dog gets and the fovorite count it gets.

This suggest that while a dog may be rated high, it doesn't necessarily mean people will give the tweet a like

ii. Finding the correlation between the rating a dogs get and the number of retweets it gets

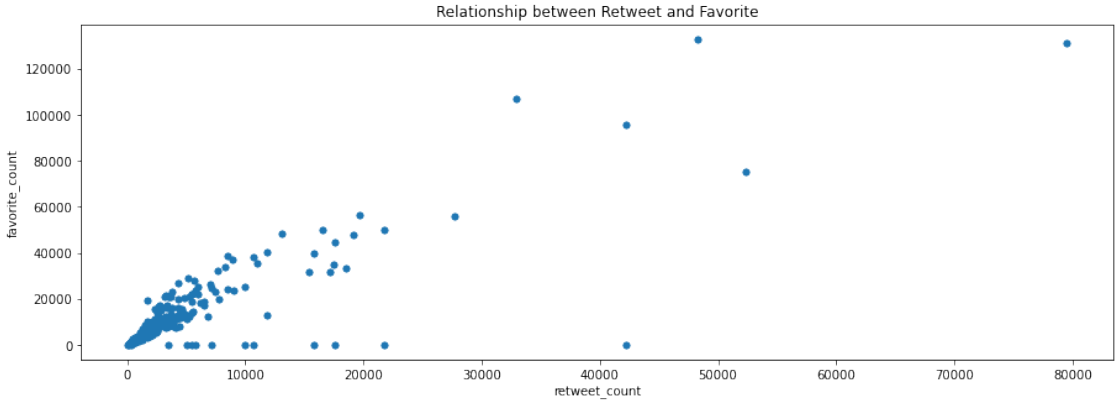
[33]:
```
tweet_data.plot(x= 'rating_numerator',y='retweet_count', kind = 'scatter',␣
↪figsize=(15,5))
plt.title('Relationship between Retweets and Rating');
```



From our Scatterplot above we can see that there is no correlation between the rating a dog get and its retweets

iii. Finding the correlation between the retweets a dogs get and the number of favorite it gets

[34]:
```
tweet_data.plot(x= 'retweet_count',y='favorite_count', kind = 'scatter',␣
↪figsize=(15,5))
plt.title('Relationship between Retweet and Favorite');
```



From the scatter plot above we can see that while there is no correlation between (i)`ratings and retweets`, (ii)`Favorite Count and Rating`. There is positive correlation between the retweet a Dog tweet gets and the Favorite count it also gets