act_report

September 11, 2022

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.

in "prediction1_conf", the data shows us how confident the algorithm is . using the describe method, "prediction1_conf" has a higher mean, max and min value than "prediction2_conf" which has higher values than "prediction3conf". We can see that the "prediction1_conf" median is 0.588 (58.8%) which is not high. the prediction is not very accurate.

In the "retweet_count" the median is equal to 1366.5 and the mean is equal to 2784.47 this explains that the distribution is skewed to the right and the data contains outliers as we can see the min, max , std values especially the max which is 79515

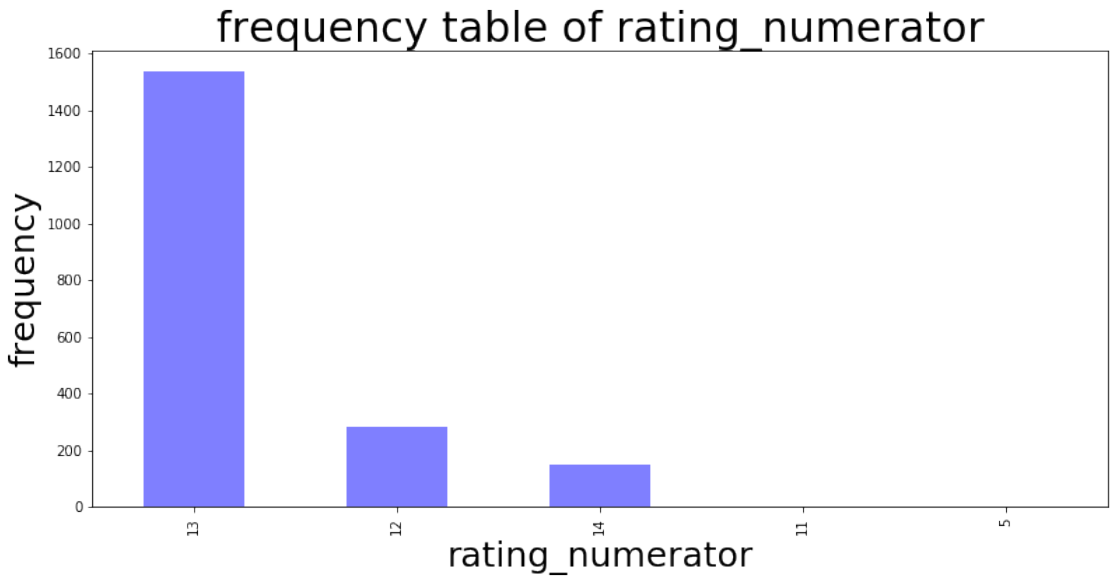
in the correlation table, we can see that there is a high positive correlation between "favourite_count" and "retweet_count". that mean if one increase by 1 count the other increases by 0.913 . between "prediction1_conf" and "prediction2_conf" there is negative correlation : -0.708 which means if the first prediction is high in confidence the other is affected. which is very logical , if the first prediction guess the breed of the dog the third prediction is likely to be not accurate

as we can see in the bar chart the rating with the highest frequency is by far value '13' most of tweet rates a are 13 with over 1500 occurences, then 12 and 14 with respectively 282 and 150

```
In [1]: import pandas as pd
        import matplotlib.pyplot as plt
```

```
In [6]: df = pd.read_csv("twitter_archive_master.csv")
        rating = df["rating_numerator"].value_counts()
        rating.plot(kind = "bar", x = "rating_numerator" ,  color = "blue", alpha = 0.5,figsize
        plt.xlabel("rating_numerator",fontsize = 25)
        plt.ylabel("frequency" , fontsize = 25)
        plt.title("frequency table of rating_numerator" , fontsize = 30)
```

```
Out[6]: Text(0.5,1,'frequency table of rating_numerator')
```


frequency table of rating_numerator
frequency
rating_numerator
1600
1400
1200
1000
800
600
400
200
0
13
12
14
11
5

In [ ]: