

WeRateDogs Twitter Data: Analyses Report

```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import tweepy
import requests
import re
import time
import os
import json
import seaborn as sns
from tweepy import OAuthHandler

%matplotlib inline

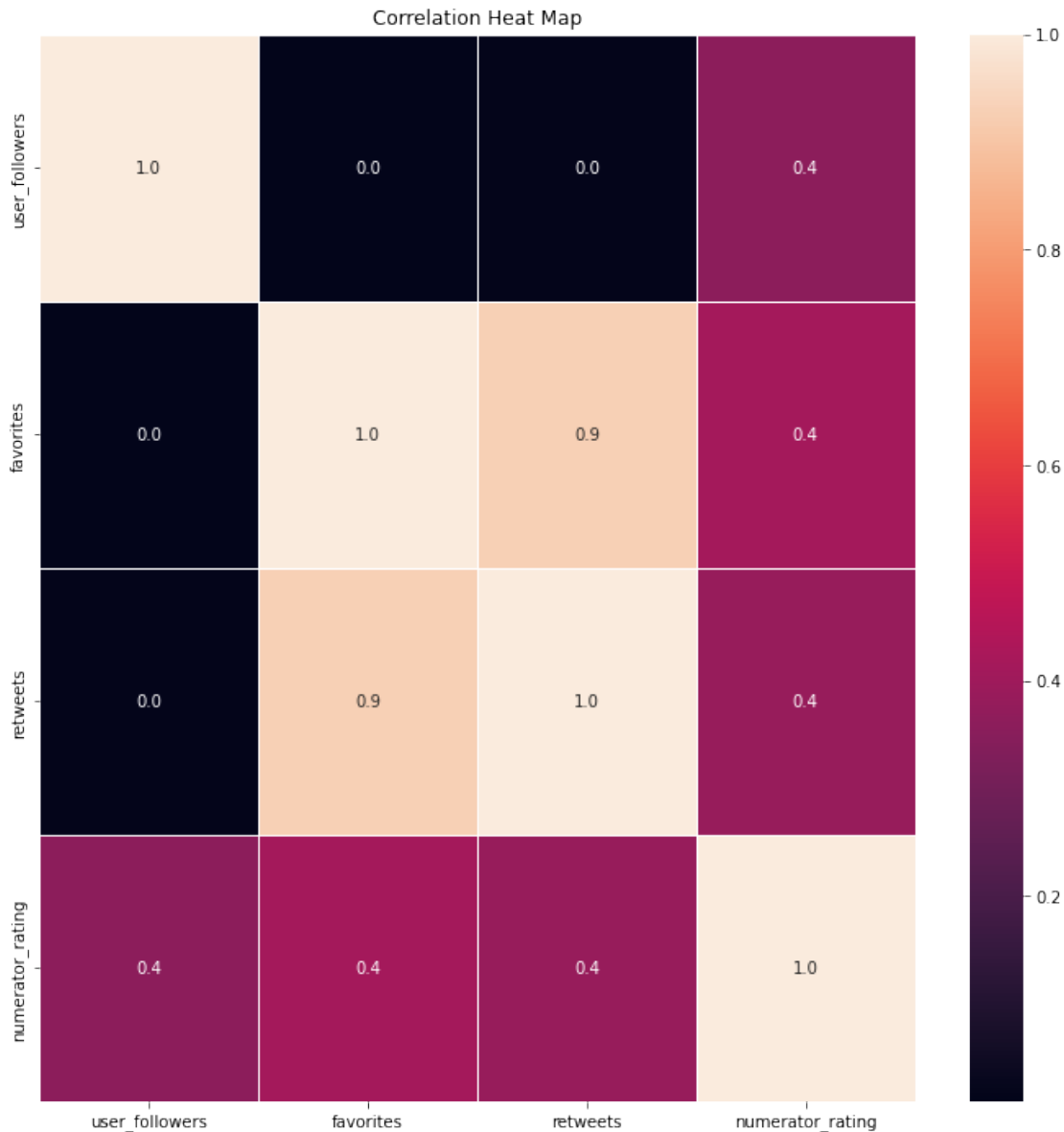
df_enhanced_clean = pd.read_csv('twitter_archive_master.csv')

df_enhanced_clean['tweet_id'] =
df_enhanced_clean['tweet_id'].astype(str)
df_enhanced_clean['timestamp'] =
pd.to_datetime(df_enhanced_clean.timestamp)
df_enhanced_clean['text'] = df_enhanced_clean['text'].astype(str)
df_enhanced_clean['source'] =
df_enhanced_clean['source'].astype('category')
df_enhanced_clean['favorites'] =
df_enhanced_clean['favorites'].astype(int)
df_enhanced_clean['retweets'] =
df_enhanced_clean['retweets'].astype(int)
df_enhanced_clean['user_followers'] =
df_enhanced_clean['user_followers'].astype(int)
df_enhanced_clean['name'] = df_enhanced_clean['name'].astype(str)
df_enhanced_clean['numerator_rating'] =
df_enhanced_clean['numerator_rating'].astype(float)
df_enhanced_clean['denominator_rating'] =
df_enhanced_clean['denominator_rating'].astype(float)

Heatmap Correlation
# To come up with a correlation map

f,ax = plt.subplots(figsize=(12, 12))
sns.heatmap(df_enhanced_clean[['source','user_followers','favorites','retweets',
                                'numerator_rating']].corr(), annot=True, linewidths=.8,
fmt= '.1f',ax=ax)
plt.title('Correlation Heat Map')

Text(0.5, 1.0, 'Correlation Heat Map')
```



Column Chart

Column chart showing the most used twitter source

To load the dataset

```
df_master = pd.read_csv('twitter_archive_master.csv')
```

```
source = df_master['source'].value_counts()
```

To plot the column chart

```
source
```

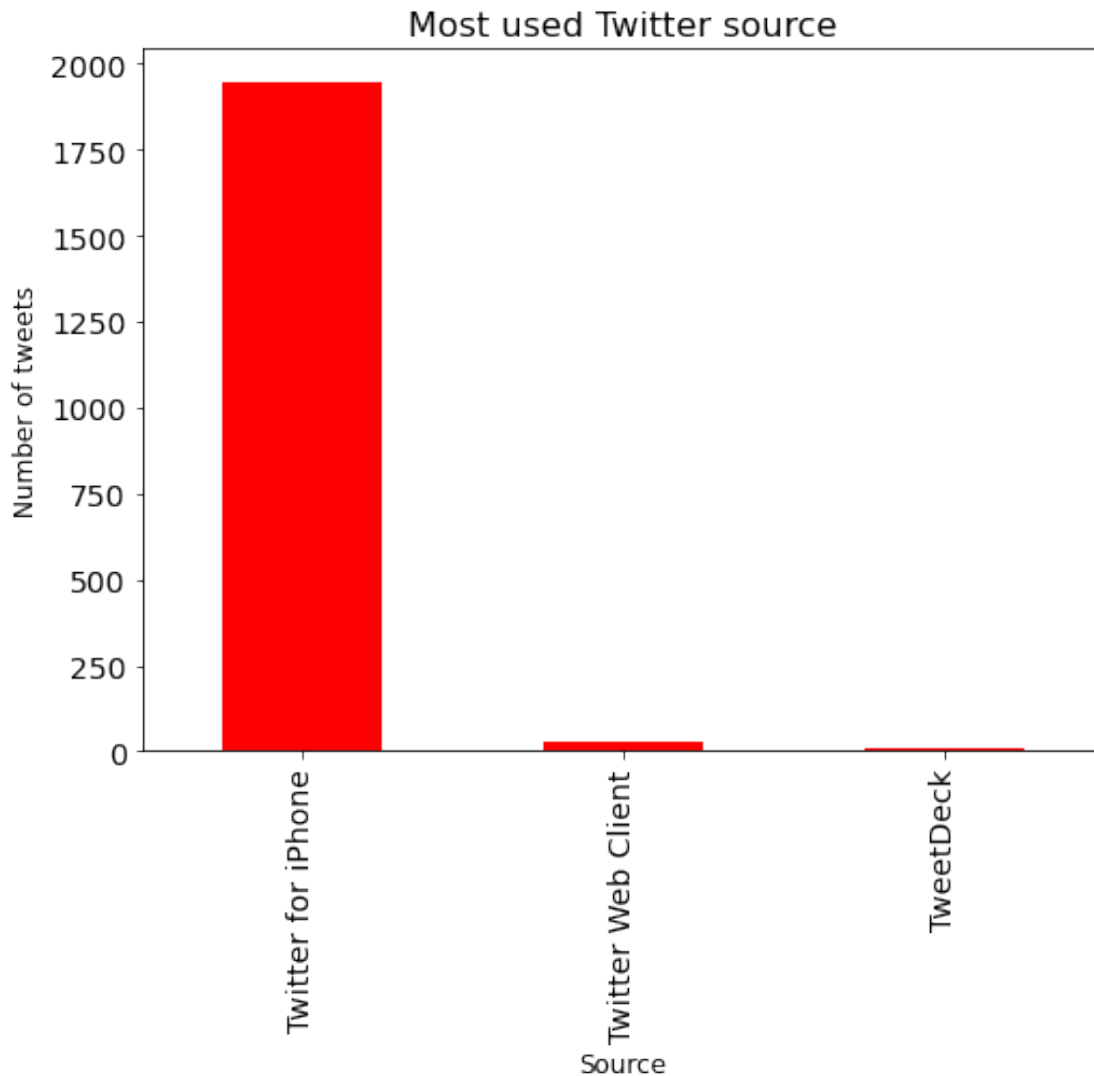
```
g_bar = source.plot.bar(color = 'red', fontsize = 14)
```

```

#figure size(width, height)
g_bar.figure.set_size_inches(8, 6);

#Add labels
plt.title('Most used Twitter source', color = 'black', fontsize =
'16')
plt.xlabel('Source', color = 'black', fontsize = '12')
plt.ylabel('Number of tweets', color = 'black', fontsize = '12');

```



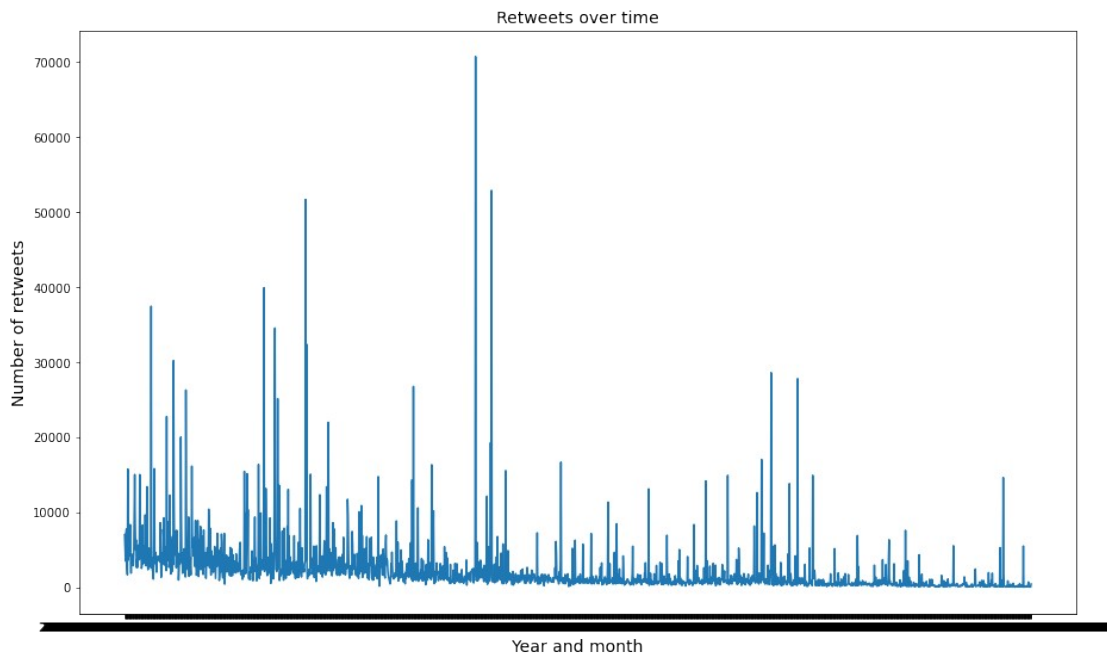
Retweets overtime (Line Chart)

```

# To plot the line chart
sns.set_context()
plt.subplots(figsize=(15, 9))
plt.plot(df_master.timestamp, df_master.retweets)
plt.title('Retweets over time', color = 'black', fontsize = '14')

```

```
plt.xlabel('Year and month', color = 'black', fontsize = '14')
plt.ylabel('Number of retweets', color = 'black', fontsize = '14');
```

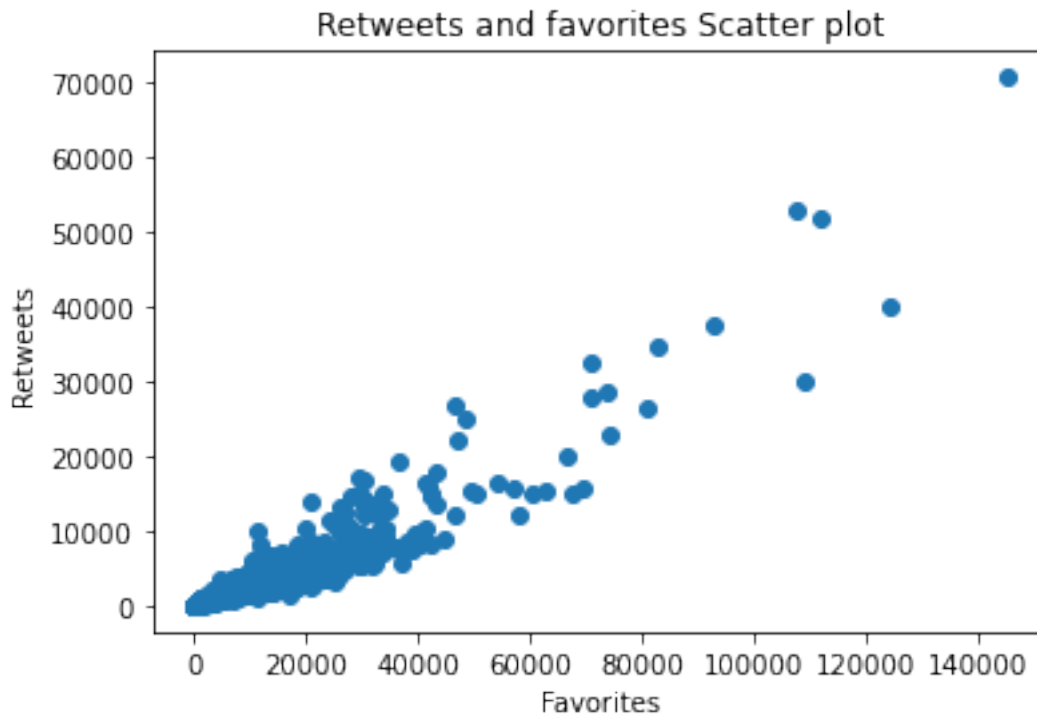


Scatterplot

This section is to plot a scatterplot showing the correlation between favorites and retweets

```
plt.scatter(df_enhanced_clean['favorites'],
df_enhanced_clean['retweets'])
plt.xlabel('Favorites')
plt.ylabel('Retweets')
plt.title('Retweets and favorites Scatter plot')
```

```
Text(0.5, 1.0, 'Retweets and favorites Scatter plot')
```



Conclusion

I can say that the analysis of WeRateDogs' tweets reveals a lot of fascinating information about how people's attitudes toward various dog breeds and canine life phases change over time. It also reveals that an individual's Twitter following is not directly correlated with the number of times their tweets are retweeted.