

# act\_report

February 28, 2021

## 1 Act Report

I have worked on three source files for Udacity's data wrangling project. Firstly, I collected data from CSV files, JSON text files, and scraping data from Twitter API. Then I merged the data into a single data frame. I assessed the data for underlying issues. I resolved them one by one. I described in detail each step in the Wrangel report. So, here I am going to talk about the only visualizations and overall results of data in detail.

```
In [1]: # all needed libraries
import matplotlib
import pandas as pd
import numpy as np
import requests
import tweepy
import os
import json
import time
import re
import matplotlib.pyplot as plt
import warnings
import seaborn as sns

%matplotlib inline

In [ ]: df_twitter_clean = pd.read_csv('twitter_archive_master.csv') # saved master data file
df_twitter_clean

In [ ]: df_twitter_clean['tweet_id'] = df_twitter_clean['tweet_id'].astype(object)
df_twitter_clean['timestamp'] = pd.to_datetime(df_twitter_clean.timestamp)
df_twitter_clean['source'] = df_twitter_clean['source'].astype('category')
df_twitter_clean['life_stage'] = df_twitter_clean['life_stage'].astype('category')
df_twitter_clean['dog_gender'] = df_twitter_clean['dog_gender'].astype('category')

df_twitter_clean.set_index('timestamp', inplace=True)
```

## 2 Correlation Heat Map

```
In [ ]: #correlation map
f,ax = plt.subplots(figsize=(12, 12))
sns.heatmap(df_twitter_clean[['source','user_followers','favorites','retweets',
                             'rating_numerator']].corr(), annot=True, linewidths=.8, fmt= '.1f',ax=ax)
plt.title('Correlation Heat Map')
```

## 3 Pie chart of Life stage of dogs

```
In [ ]: df_count_stages = list(df_twitter_clean[df_twitter_clean['life_stage'] != 'None']['life_
life_stages = df_twitter_clean[df_twitter_clean['life_stage'] != 'None']['life_stage'].v
explode = (0.2, 0.1, 0.1, 0.1) #we can use the explosion property of slices of a pie cha

figure1, axis1 = plt.subplots()
axis1.pie(df_count_stages, explode = explode, labels = life_stages, shadow = True, start
axis1.axis('equal')
```

From above pie chart,dogs in pupper life stage got most tweets followed by doggo and puppo.

## 4 Scatter plot between favorites and retweets.

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

### 4.1 Analysis

- There is a strong correlation between favorites and retweets.
- Negative correlation among user followers and retweets.
- Dogs in the Pupper stage get most tweets.
- There is a big positive 0.9 correlation shown by the correlation heat map between the retweets and favorites. This means retweets and favorites have a strong positive relationship.

### 4.2 Conclusion

Through all the analysis I made, I can say that the Twitter data of WeRateDogs gives us a lot of interesting perceptions about people who tweet about the different breeds of dogs within their different life stages. Interestingly, the Pupper stage of dog's life got the most of tweets followed by Doggo. Moreover, having more followers on Twitter doesn't increase the retweets of user's tweets.

```
In [ ]:
```