Act Report, by Rami Salman

This project is the 7th project of data analyst nanodegree program from udacity , it's about wrangling data steps (gathering , assessing , then cleaning) . The data is archived data from WeRateDogs twitter account , and the data is not ready , that is gathered in three different ways as the first step of the project , this data includes :

1- Archive data

This data file is given in the classroom and downloaded manually. This file is the archive of WeRateDogs contains data about the account tweets with 2356 row (tweet).

2- image predictions data

The given file is in a link to download programmatically using python . This file contains predictions of images(dogs) using neural network that can classify breeds of dogs.

3- json data from twitter api

This is the most challenging part of data gathering , this is the first time for me to get data from ready api. It takes about two hours of contacting with twitter to create my developer account then I got the credentials and start collecting data , it takes about 35 minutes of collecting data of tweet id's . Then I saved the collected data in txt file before convert them to csv file to use it , the selected columns/features are the following : 'id','retweet_count','favorite_count','created_at'.

After that , I assessed some problems in the data (Data Quality & Tidiness) then I cleaned them .

After I finished the wrangling data steps , I need to create visualizations to extract insights about the data , so I used the tableau software to create the visualization , you can find it at the following link: https://public.tableau.com/profile/rami.salman#!/vizhome/weRateDogs/WeRateDogsStory?publish=yes .

After the visualization, I found many insights in the data, see the following:

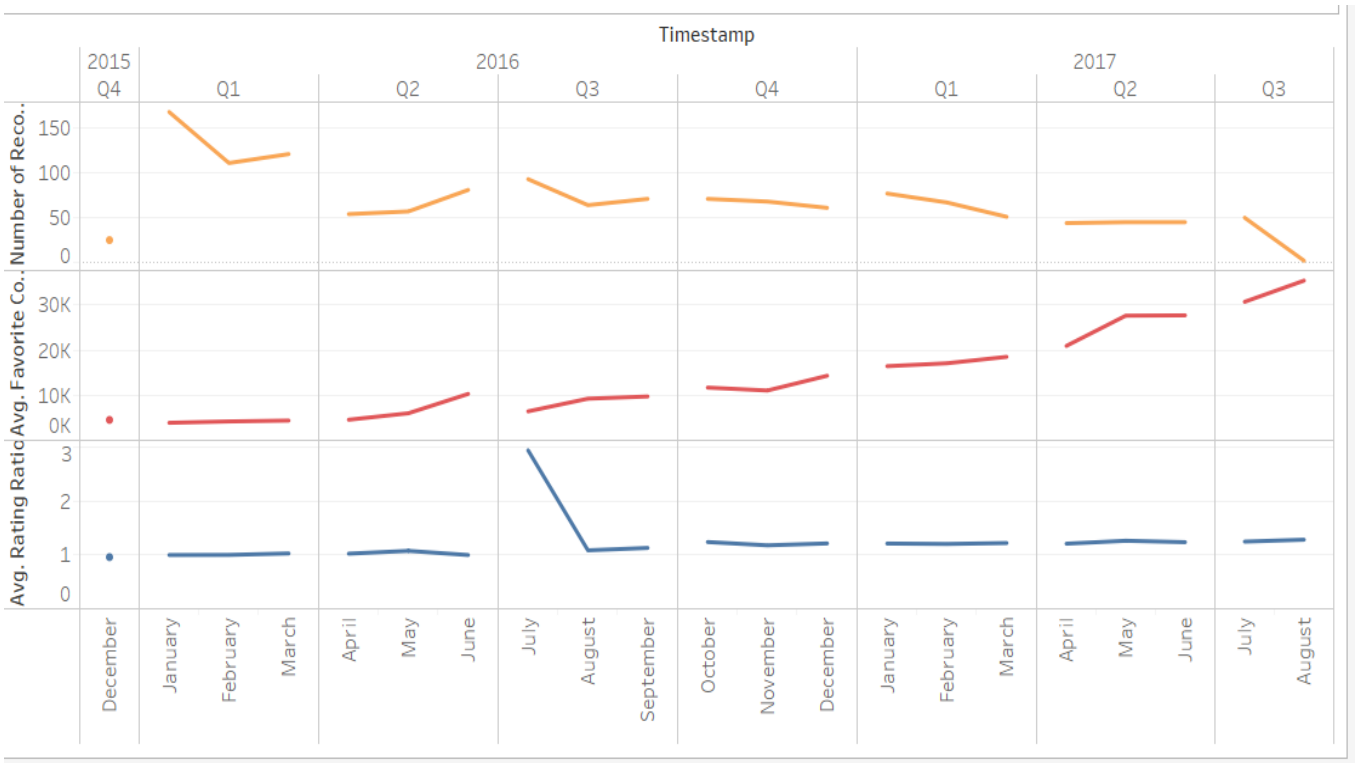

Figure1: number of records , avg. favorite count and avg. ratio among years.

- **From the previous plot , we can extract the following insights:**

1- Number of records decreases by forwarding days.

2- Average of favorite number increases by forwarding days.

3- Average of rating ratio increases in very slow rate by forwarding days.

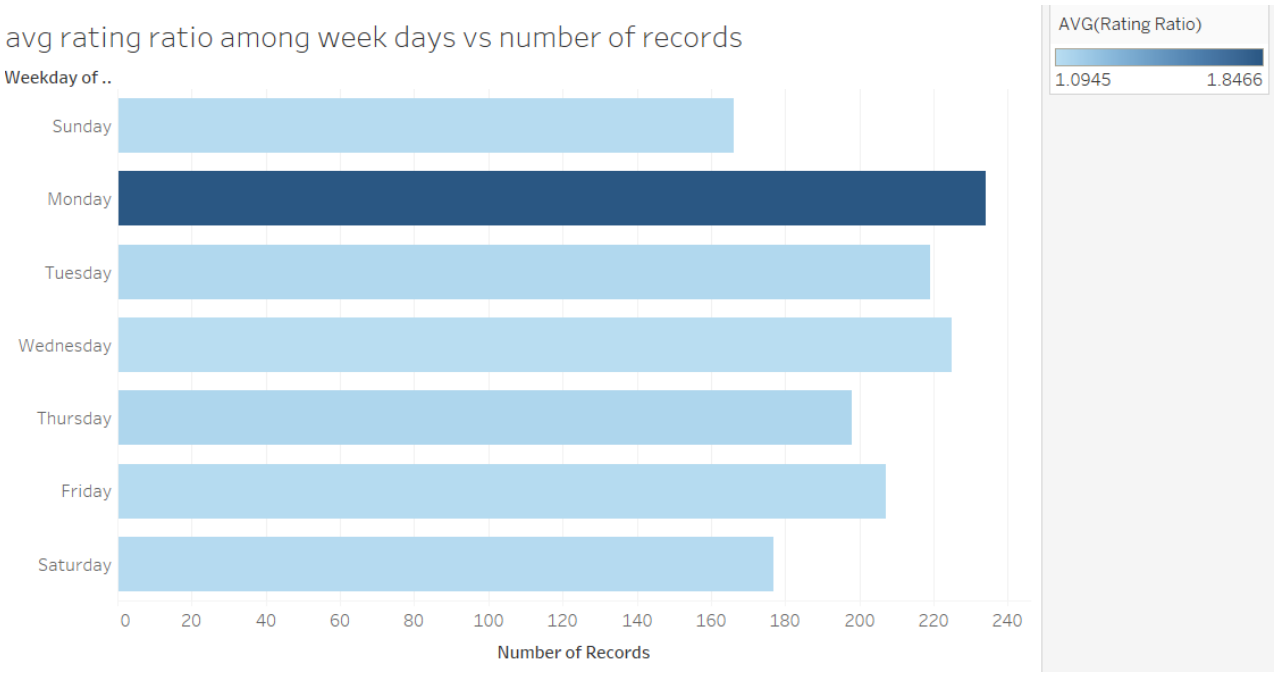

figure2: avg. rating ratio among week days vs number of records

- **From the previous plot , we can extract the following insights:**

4- The highest number of tweets is on Monday followed by Tuesday , on the other hand , the lowest number of tweets is on Saturday.

5- The highest average rating ratio is on Monday.

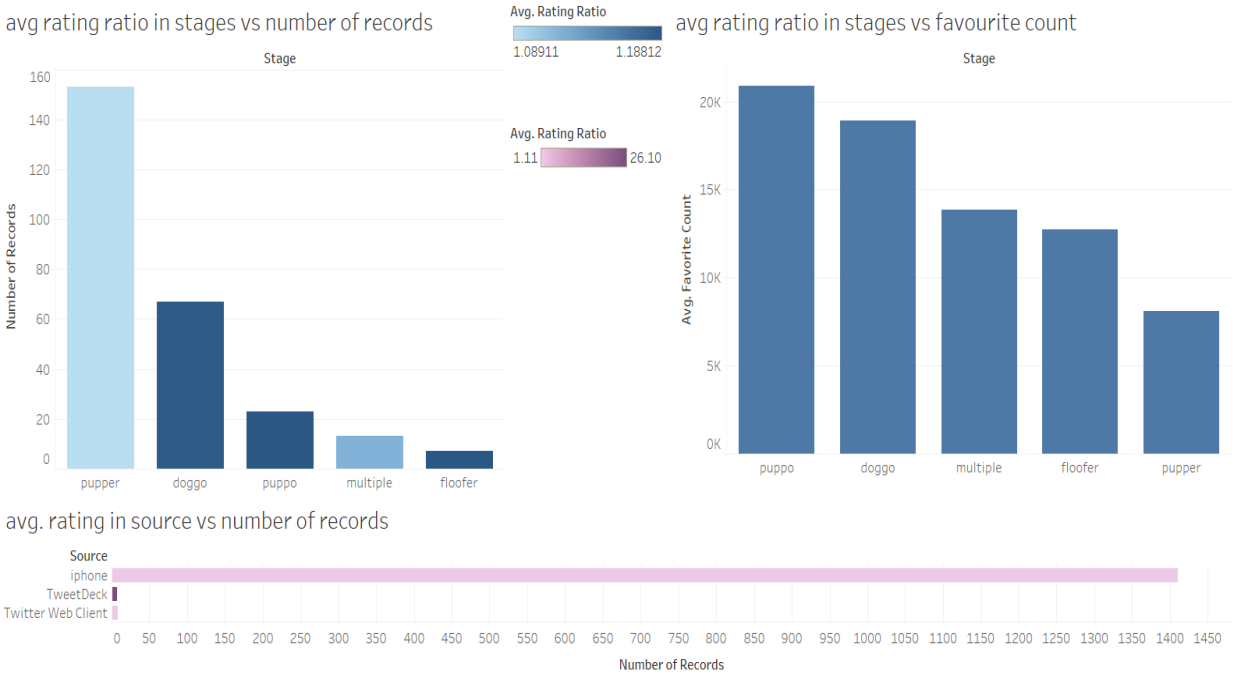

figure3 : dashboard of tweets with dog stage and source .

- **From the previous plot , we can extract the following insights:**

6- The highest avg. favorite count is for puppo dogs .
7- The highest number of records is for pupper dogs .
8- The highest average rating ratio is for floofer dogs with 1.88 .
9- Most of tweets are from iphone .
10- The tweets uploaded from tweetDeck have the highest rating ratio of 26.1 .