WRANGLE AND ANALYZE DATA

ACT REPORT

Weijia Xu

WRANGLE AND ANALYZE DATA

Purpose:

In this project I took tweet data from the popular WeRateDogs twitter account,the purpose of this documents is to communicate the insights and visualizations produced as a result of my wrangling efforts against the WeRateDogs archive.

INSIGHTS:
INSIGHT 1 :
Tweet with Hight rating winner749981277374128128

INSIGHTS:
INSIGHT 2 :
Tweet with lowest rating 668537837512433665

Insight 3

VISUALIZATION:

Getting count from all tweets predictions source there are more interesting things as following.

There are (3) sources of the tweets in our dataset as follows:

1926 (Twitter for iPhone)

27 (Twitter Web Client)

11 (TweetDeck)

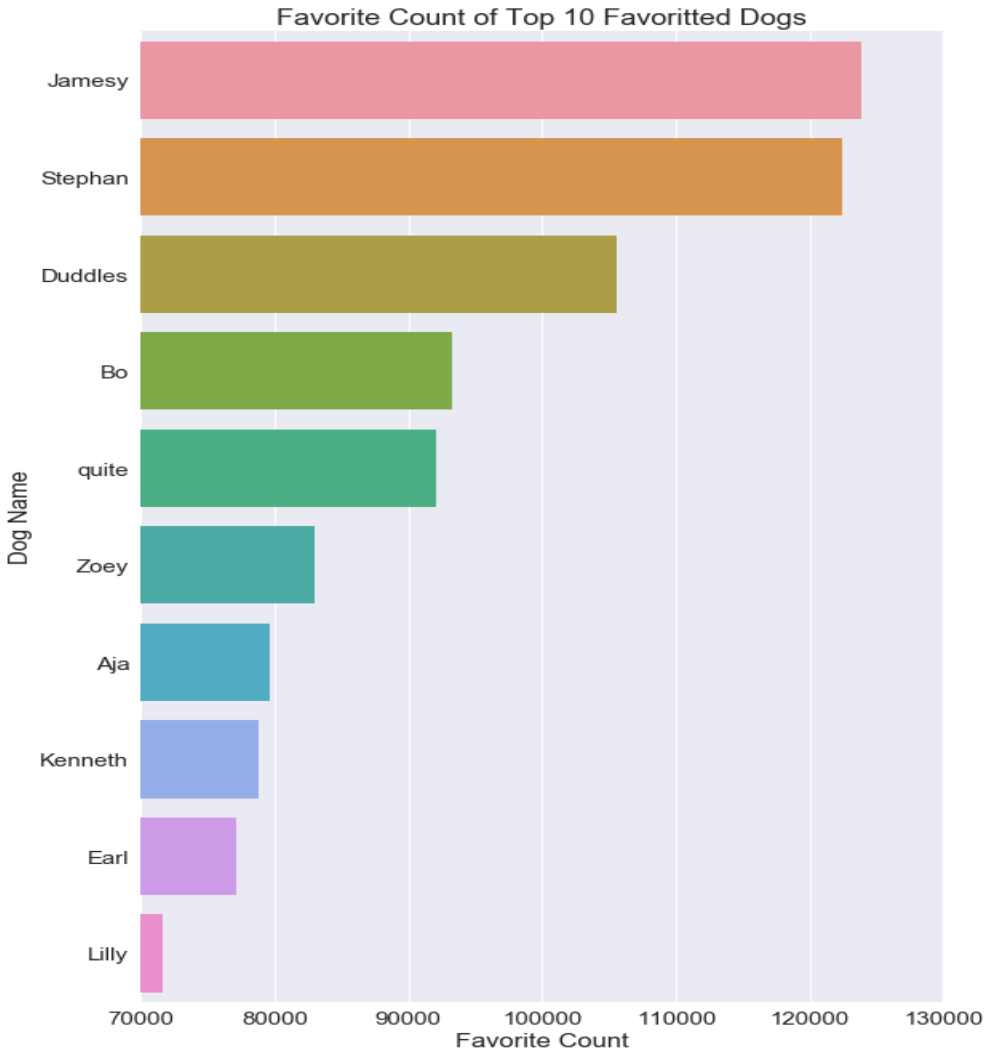
Insights:

1. My dataset consists of 20 variables with 1964 oberservations
2. The average rating for all the dogs was about 116 percent
3. The favorite count was about 8879 which I am interesting about their relationship.

I have created 2 visualizations, the first one I created the variable of top 10 and create a data set that contains the top 10 favoritted dogs then plot the top favorited dogs.

The second one, I would be interesting to see if appears to be any relationship between favorite count and retweet count that can appears to be a positive linear relationship between them.

VISUALIZATION 1 – top 10 favorited dogs by favorite count
This visualization shows the top favorited dog by favorite count for names dogs. The top favorited dog named Jamesy.



VISUALIZATION 2– Favorite count by retweet count
Here we can see what looks like a positive relationship between favorite count and retweet count. The more a tweet gets retweet,the more likely it seems that it would be favorited.