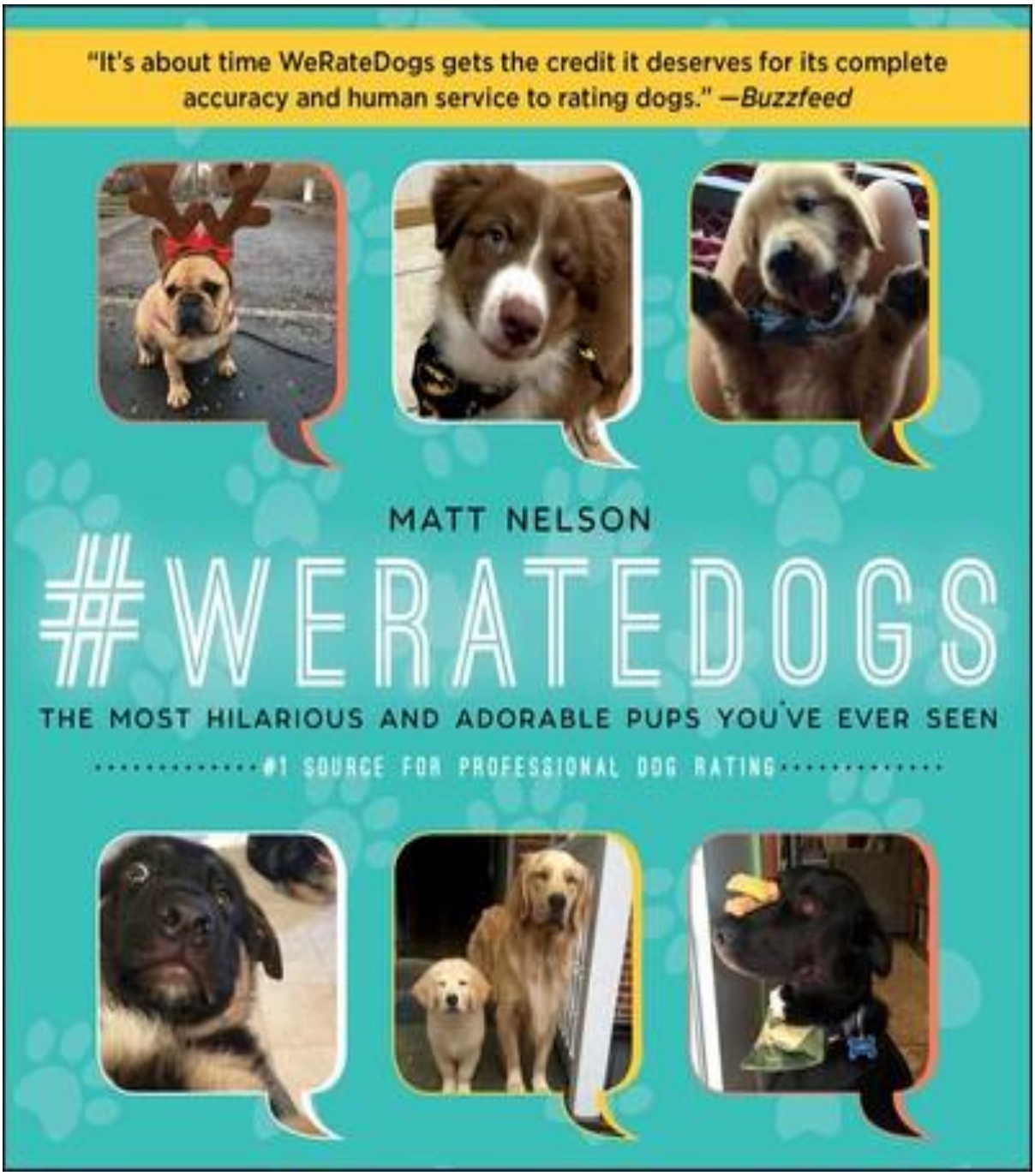
INSIGHTS AND VISUALITZATIONS FROM DATA WRANGING – WeRateDogs



PhotoCredit:
https://www.simonandschuster.com/books/WeRateDogs/Matt-Nelson/9781510767263

INTRODUCTION

The dataset that is wrangled is the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. This Twitter account rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10 and numerators always greater than 10. The twitter account has over 4 million followers and has received international media coverage. This report documents my wrangling efforts as part of the ALX/Udacity data analysis program.

DESCRIPTIVE STATISTICS OF THE DATASET

The minimum rating for a tweet in the dataset was zero. This means that there were tweets that were not rated in the dataset.

The least number of images per tweet is 1 with the maximum being 4.

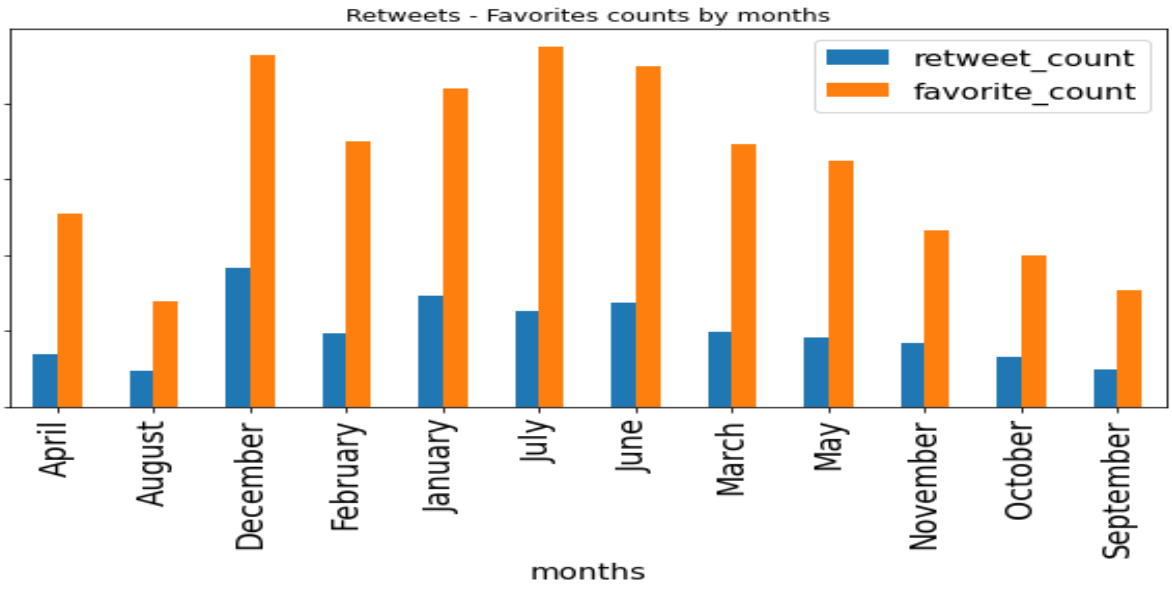
CORRELATION BETWEEN THE VARIABLES

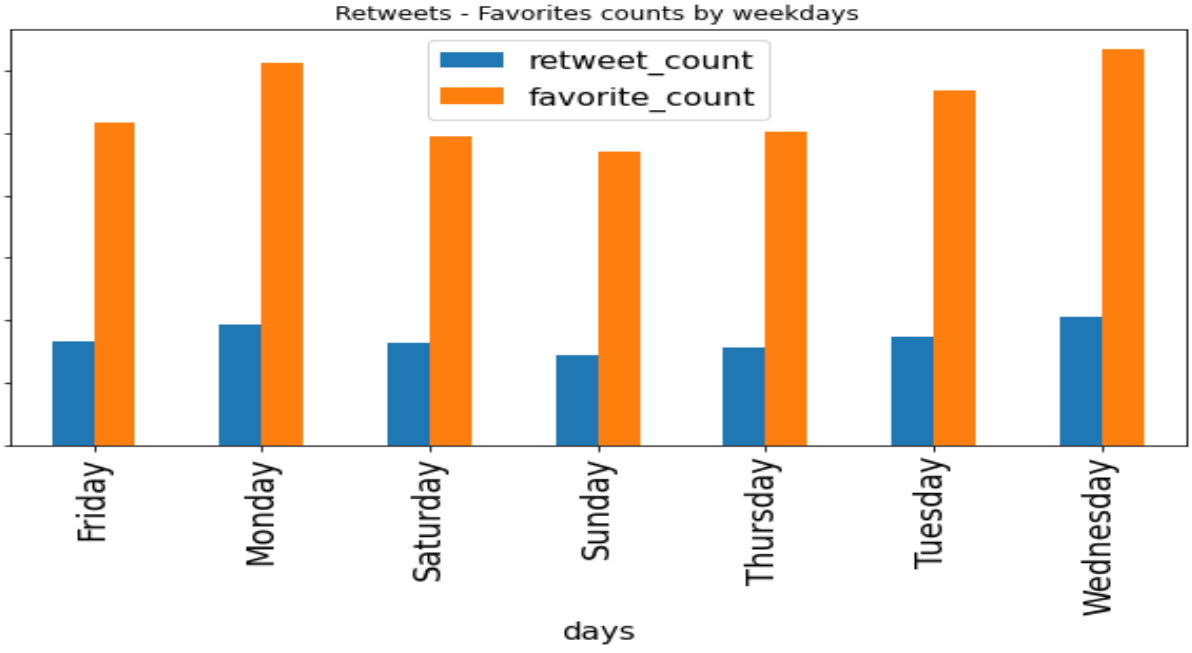
1. The correlation coefficient between retweet_count and favorite_count is 0.911732. This indicates a strong positive correlation between these metrics.

2. There is a negative correlation of -0.707994 between p1_conf and p3_conf. The value of p3_conf is affected by the value of p1_conf. The more confident the first prediction is in the neutral network, the less accurate the subsequent prediction of the network.

FAVOURITE AND RETWEET ANALYSIS BY MONTH AND DAYS

1. The month of December had the highest number of retweets.
2. July had the highest number of favorites.
3. December was the month with the highest activity.
4. Wednesdays have the most activity during the week.


Retweets - Favorites counts by months
retweet_count
favorite_count
April
August
December
February
January
July
June
March
May
November
October
September
months


Retweets - Favorites counts by weekdays
retweet_count
favorite_count
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
days

ANALYSES OF DOG STAGES

#'doggo' and 'puppo' have received the highest number of retweets and favorites.