ANALYSIS AND INSIGHTS

Udacity DAND: Wrangle and Analyze Data Project

By: Ebtihal Alkhalifah

INTRODUCTION

The wrangle and analyze project is part of Udacity Data Analysis Nanodegree program. It is about wrangling data from multiple sources related to tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. The goal is to wrangle WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations, In this report I gathered the datasets, Assessed the quality and tidiness issues, depending on these issues I cleaned the data and started to analyze and visualize after that. So the analysis was depending on some questions came to my mind, Which dog had the highest favorite count among all? Finding the statistical result of retweet count and favorite count (min, max, mean,…)? What is the most common dog stage? And find the relation between the retweet count and favorite count?.

HIGHEST FAVORITE COUNT

Most of the dogs were highly rated but to satisfy my curiosity I

wanted to see the one with highest likes which was the one below, after seeing the pic I would really like to check the percentage of each gender of those who liked the photo ;), but this won't be easy.

Figure 1. Dog with highest favorite count.

MOST COMMON DOG STAGE?

The plot bar below shows that the most common dog stage is Pupper.

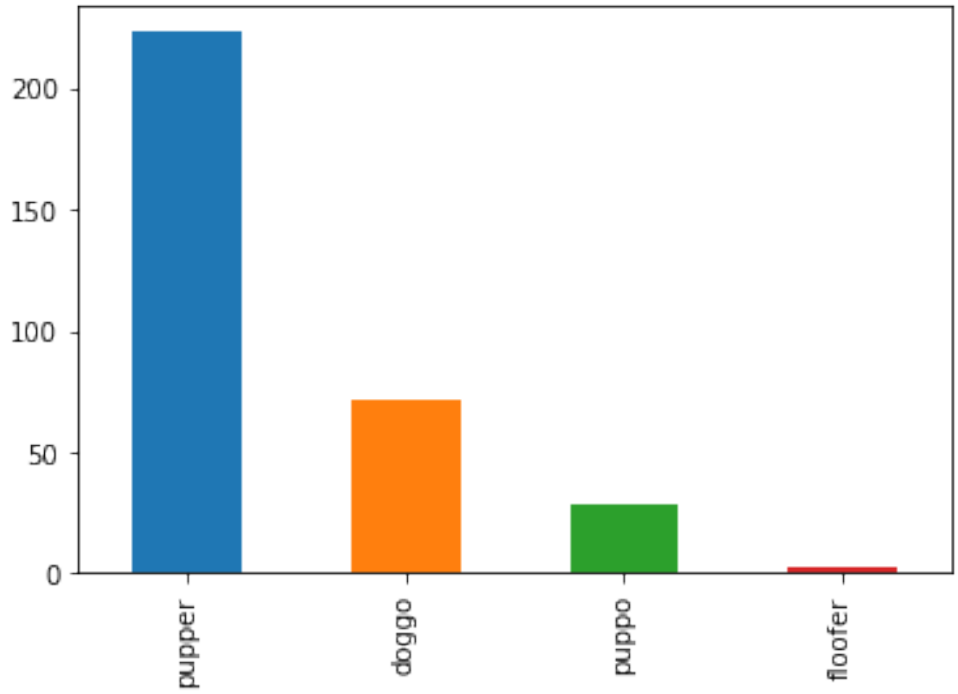

Figure 2. Most common dog stage.

THE RELATIONSHIP BETWEEN RETWEETS AND FAVORITES?

The scatter plot below shows that the retweets count VS favorites count are positively correlated, where when the retweets decrease the favorite decreases as well.

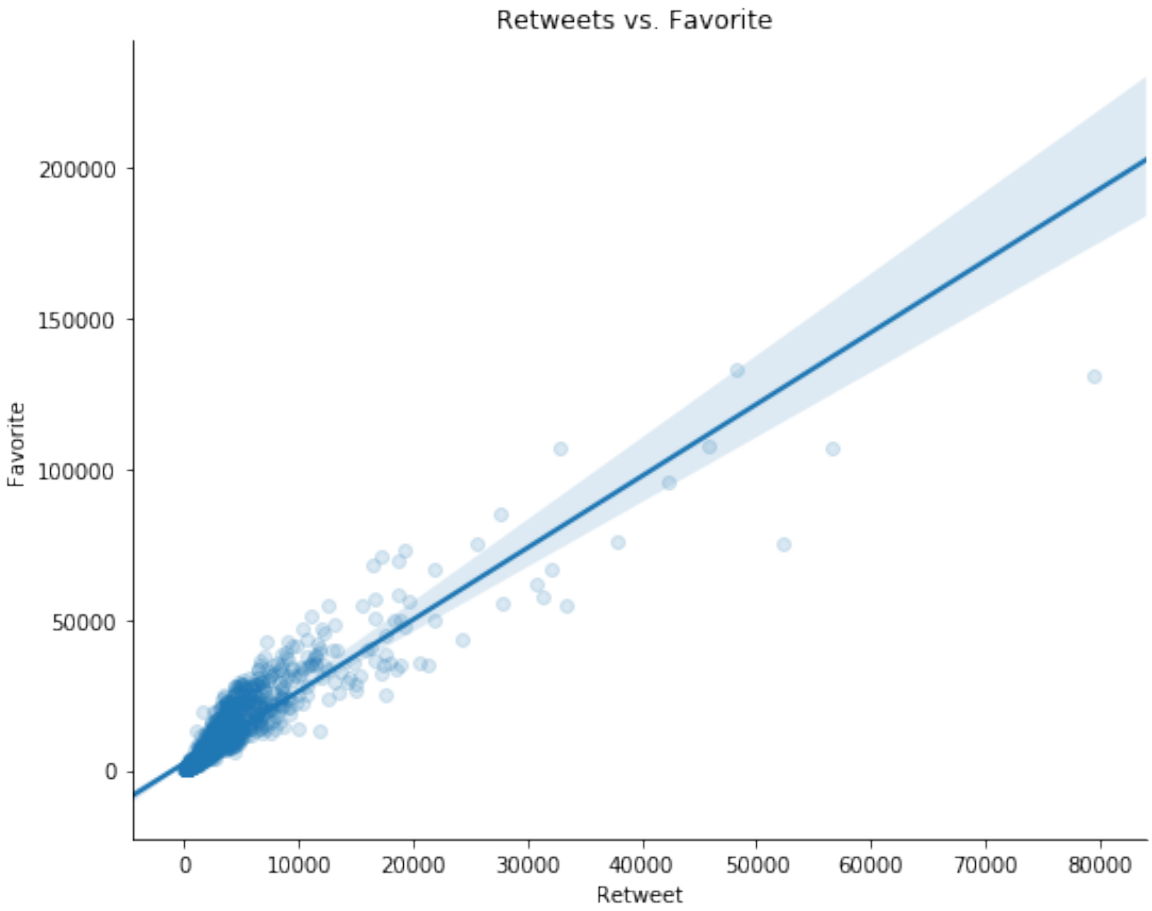

Figure 3. Retweet count vs favorite.