The WeRateDogs Project

Data Wrangling Project, Udacity Data Analyst Nanodegree

Author: Jaroslav Tran
Date:06/03/2020

The aim of this project was to gather, assess, clean and analyse multiple datasets from @dog_rates Twitter account also known as WeRateDogs. WeRateDogs is a Twitter account with over 8.6 Million followers worldwide.

The reason behind its popularity is the fact that the account rate incredible dog picture with an entertaining content and quirky rating of the dogs. The rating is a fraction where the denominator is 10 (per dog for multiple dogs on the picture) and denominator often greater than 10.

Three datasets were used during this project. The first was a twitter archive generously provided by WeRateDogs for exclusive use by Udacity student community for learning purposes in a csv format. It contained a few thousand tweets with a fairly basic tweet data such as timestamp, text, etc.
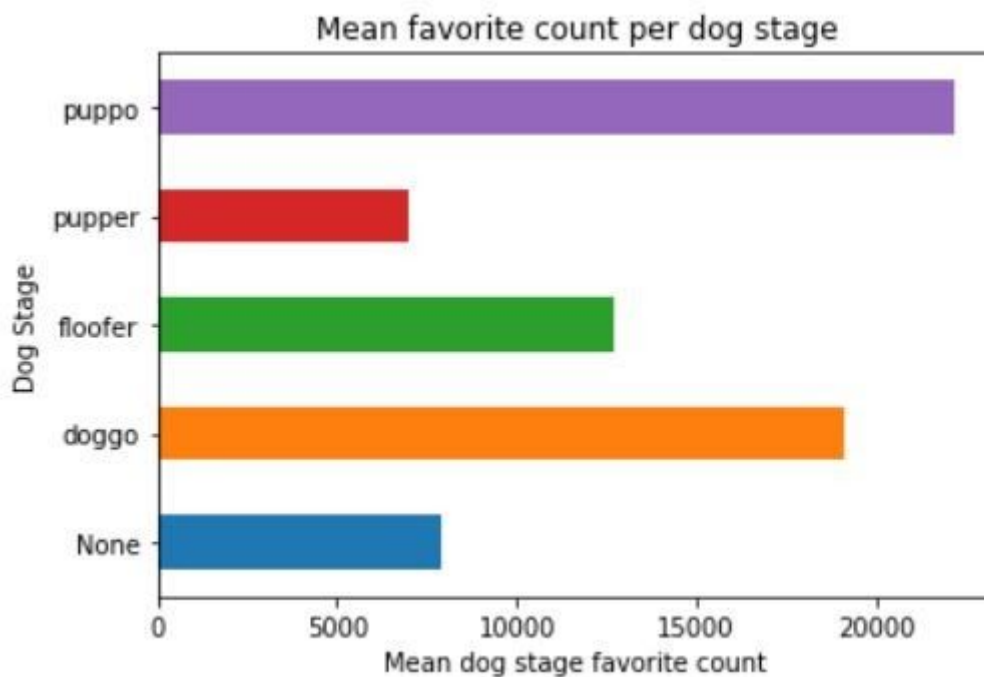
The next dataset was provided by the Udacity, in which it run convolutional neural network algorithm on tweet images to identify a dog breed predictions along with the confidence intervals.

The last dataset was gathered by using Tweepy library to query each tweet´s ID using Twitter API and get more granular data about those tweets.

After wrangling the data and overcoming the most serious data quality and data tidiness issues, I dived into statistical analysis, answering a few questions:
1) What stage of dogs has the most favorites(likes)? In other words, which dog stage seems to be most liked?
2) What are the top 10 dogs with the highest number of followers?
3) Using the first prediction, what are the top 10 dog breeds with the highest rating?

Q1: Which dog stage seems to be most liked?

Mean favorite count per dog stage

Comments: The most liked dog stage was unsurprisingly puppies. Everybody loves puppies and their average favorite count is off the charts.

Q2: What are the top 10 dogs with the highest number of followers?

name	Score	followers count	jpg url
Kial	10/10	8681947.0	https://pbs.twimg.com/media/CUOYBbbWIAAXQGU.jpg
Spark	8/10	8681945.0	https://pbs.twimg.com/media/CUcfnWIWsAAzlwE.jpg
Maks	10/10	8681945.0	https://pbs.twimg.com/media/CUdvambWoAA007z.jpg
Jackie	10/10	8681945.0	https://pbs.twimg.co

			m/media/CVUovvH WwAAD-nu.jpg
Unknown	9/10	8681942.0	https://pbs.twimg.com/media/CWZgPPU WUAAUOvu.jpg
Unknown	9/10	8681942.0	https://pbs.twimg.com/media/CV6ZOPq WsAA20Uj.jpg
Kawhi	10/10	8681942.0	https://pbs.twimg.com/media/CXBdJxLU sAAWql2.jpg
Eve	8/10	8681942.0	https://pbs.twimg.com/media/CVgGc9h WlAIe1bn.jpg
Pip	11/10	8681941.0	https://pbs.twimg.com/media/CVGjflNW oAEwgrQ.jpg
Cassie	10/10	8681937.0	https://pbs.twimg.com/media/Ca-XjfiUsA AUa8f.jpg

Conclusion: Surprisingly, the top 10 dogs with the most follower count are not that far from each other and all have over 8.681 million followers.

Q3: Using the first prediction, what are the top 10 accounts/tweets with the highest score (Score = rating_num / rating_denom)

Name	Jpg url	Score
Atticus	https://pbs.twimg.com/media/CmgBZ7kWcAAlzFD.jpg	177.6
Unknown (Multiple dogs)	https://pbs.twimg.com/media/CU9P717W4AAOlKx.jpg	42
Logan	https://pbs.twimg.com/media/CurzvFTXgAA2_AP.jpg	7.5
Sophie	https://pbs.twimg.com/media/Cswbc2yWcAAVsCJ.jpg	2.7

Unknown (Multiple dogs)	https://pbs.twimg.com/media/CXGaVxOWAAADjhF.jpg	2.6
Emmy	https://pbs.twimg.com/media/DC98vABUIAA97pz.jpg	1.4
Unknown	https://pbs.twimg.com/ext_tw_video_thumb/887517108413886465/pu/img/WanJKwssZj4VJvL9.jpg	1.4
Unknown	https://pbs.twimg.com/media/DEYrlZwWsAA2Wo5.jpg	1.4
Unknown	https://pbs.twimg.com/media/C2JXyARUAAE4gbL.jpg	1.4
Gary	https://pbs.twimg.com/media/C0s-XtzWgAAp1W-.jpg	1.4

Conclusion: Atticus clearly dominates this group. Although one might be concerned by the bias of the evaluator, given that the Atticus has an all american look.