DAND Wrangle And analyze Data Project Wrangling Report

By Naif Abdulrazaq Alsofyani

Contents

Introduction

After finishing the wrangling process, the analysis begins by creating a master DataFrame that contains all the datasets. In our master DataFrame, we have multiple columns that I will be going through each one of them, and then going through questions that were answered by analyzing the data.

Dataset Columns

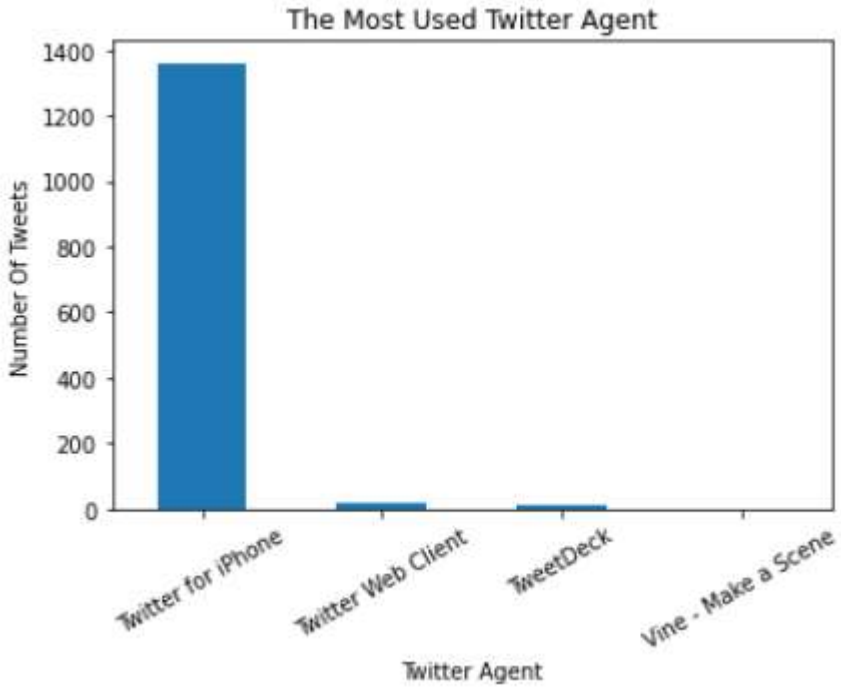
- `tweet_id` (ID associated with each tweet)
- `timestamp` (Tweet's publishing date)
- `source` (Twitter agent, i.e., Twitter for iPhone … etc.)
- `text` (Tweet's content i.e., the tweet itself)
- `expanded_urls` (Tweet's URL)
- `name` (The dog's name)
- `rating_ratio` (The dog's rating, was obtained by dividing numerator over deno minator)
- `type` (The dog's type)
- `favorite_count` (Favorites number for a tweet)
- `retweet_count` (Retweet number for a tweet)
- `url` (Tweet's shortened URL)
- `jpg_url` (Images URL)
- `img_num` (Number of images for a tweet)
- `p1` (The algorithm's #1 prediction for the image in the tweet)
- `p1_conf` (How confident the algorithm is in its #1 prediction)
- `p1_dog` (Whether or not the #1 prediction is a breed of dog)
- `p2` (The algorithm's second most likely prediction)
- `p2_conf` (How confident the algorithm is in its #2 prediction)
- `p2_dog` (whether or not the #2 prediction is a breed of dog)
- `p3` (The algorithm's third most likely prediction)
- `p3_conf` (How confident the algorithm is in its #3 prediction)
- `p3_dog` (whether or not the #3 prediction is a breed of dog)

Analysis and Visualization

Q1) What is the average rating ratio of dogs?

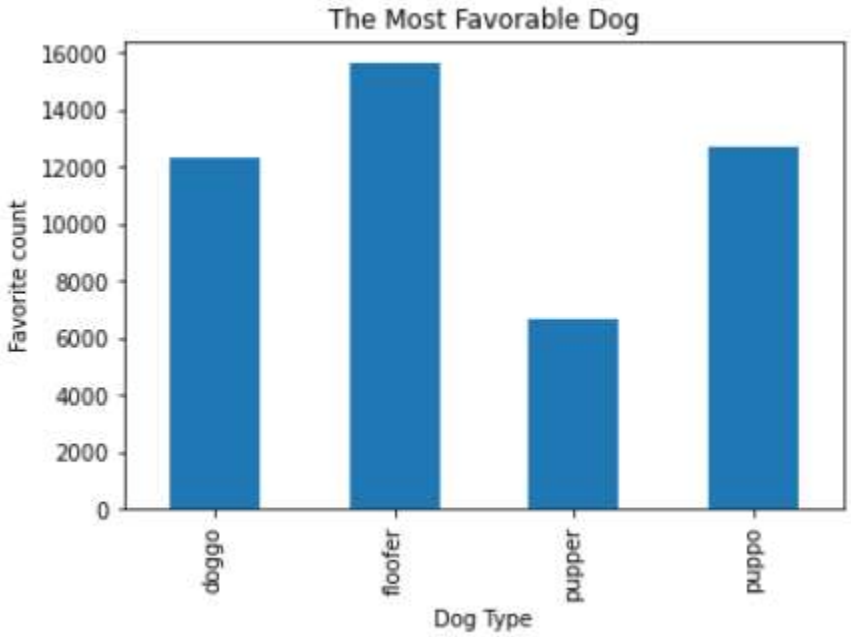
```
Dogs average rating ratio: 1.20
```

Q2) What is the most used twitter agent?



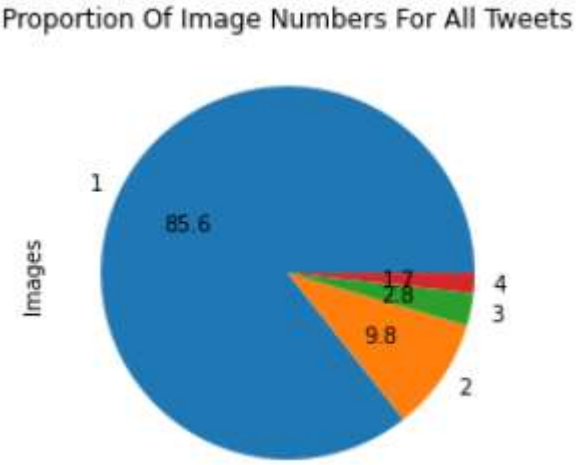
The bar chart showed that most users are using the official Twitter app on iPhone

Q3) What is the most favorable dog based on the tweet favorites number?



The bar chart showed that "floofer" dog type is the most loved dog based on all tweets' favorite count

Q4) What are the proportion of image numbers for all tweets?



The majority of tweets (85.6%) have only 1 picture