Data Wrangling & Analyze (Act Report)
Prepared By: Faris AlShammari

Table of Contents

Introduction

After the data wrangling process then I did analysis for the last data frame which I merged all that dataframes which I cleaned it into this dataframe called "twitter archive master".

Columns description:

1. [tweet _ID] it is the unique number for each tweet.
2. [in_reply_to_status_id] if it's a reply tweet so it has a unique id.
3. [in_reply_to_user_id] if it's a reply tweet so it shows the unique id for the user.
4. [timestamp] it for the date and time for the tweet.
5. [source] it for what the tweet comes from an application or website.
6. [text] it is the tweet's content.
7. [retweeted_status_id] the tweet has been retweeted based on tweet ID.
8. [retweeted_status_user_id] ID for who retweeted the origin tweet.
9. [retweeted_status_timestamp] the date of when a tweet has been retweeted.
10. [expanded_urls] the link URL for a tweet.
11. [name] The dog's name.
12. [category] the dog's breed.
13. [Dog_Rate] the dog's rating.
14. [jpg_url] It is the image URL of the dog.
15. [img_num] the number of images in a tweet.
16. [p1], [p2] and [p3] it is the number of the prediction and its algorithms.

17. [p1_conf], [p2_conf] and [p3_conf] it is the percentage of how the prediction is correct.
18. [p1_dog], [p2_dog] and [p3_dog] is whether or not the predictions are breed of dog.
19. [Liked] it is a number of how many people liked a tweet.
20. [RT] it is a number of how many people retweeted a tweet.

Analysis

Q1: The Population of dog breed

Here I search for who is the most popular dog based on its breed and I found out that the "pupper" is the most popular then "doggo" followed by "puppo" and "floofer" is the less popular.

What are the most popular dog breed
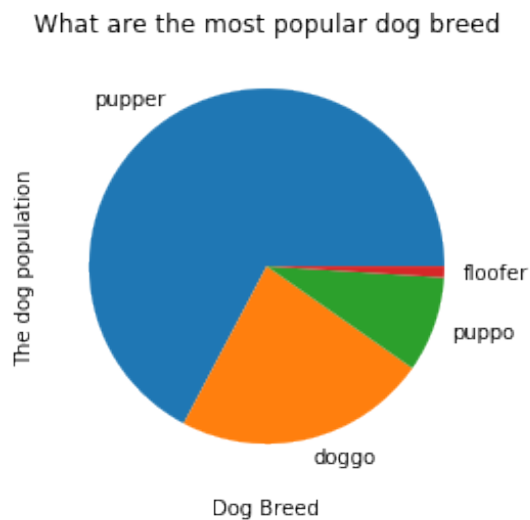
Q2: The source of tweet

Here I search for what is most active with people then I found that many people are tweeted from their iPhone devices. Then from vine followed by the website and less active is "tweet deck" application.

The number of source tweets
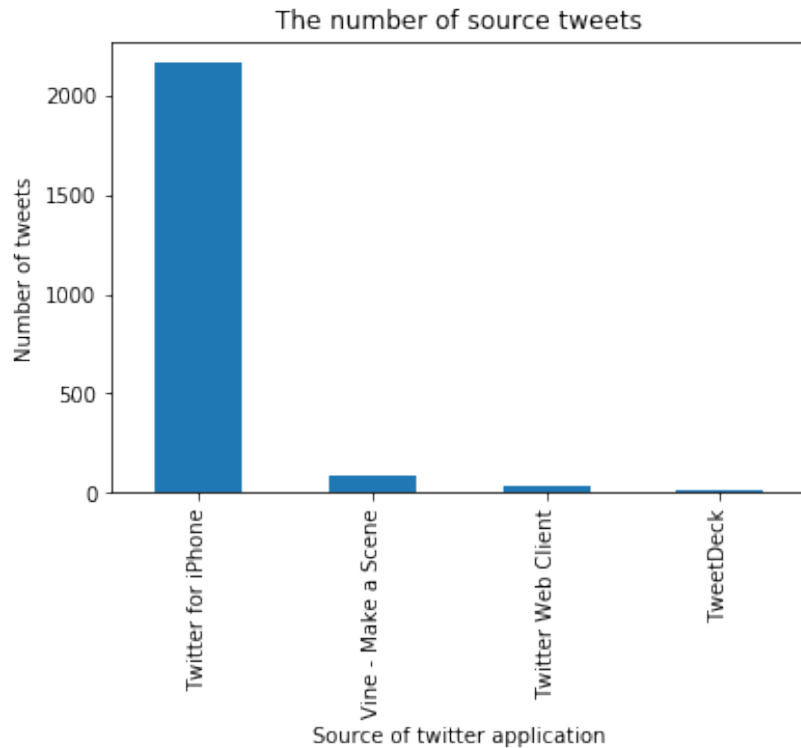
Q3: Best prediction

Here I want to know which the prediction confidant are has the highest mean, then I noticed that prediction 1 is the highest (59.466) then the prediction 2 (13.43) and the lowest is the prediction3 (6.002).

Q4: Best Dog rating

And for searching who is the best dog based on his rate then I found out that "Atticus" has the highest rate (177.6).

Conclusion

That is my Data wrangling and analysis for the dog's dataframes file and thank you for reading.