Project: Act Report

Table of Contents

Introduction

In this project, we are concerned about making meaningful insights from tweets on the @WeRateDogs Twitter page. We extracted three datasets from this page using the Twitter Developer API and came up with a final dataset *twitter_archive_master_data.csv* after wrangling was done.

We carried out Exploratory Data Analysis (EDA) on this master data (check Analyzing and Visualizing Data section in *wrangle_act.ipynb* notebook)

Exploratory Data Analysis

This report focuses on communicating all the insights and displaying the visualization(s) we got from analyzing the master data.

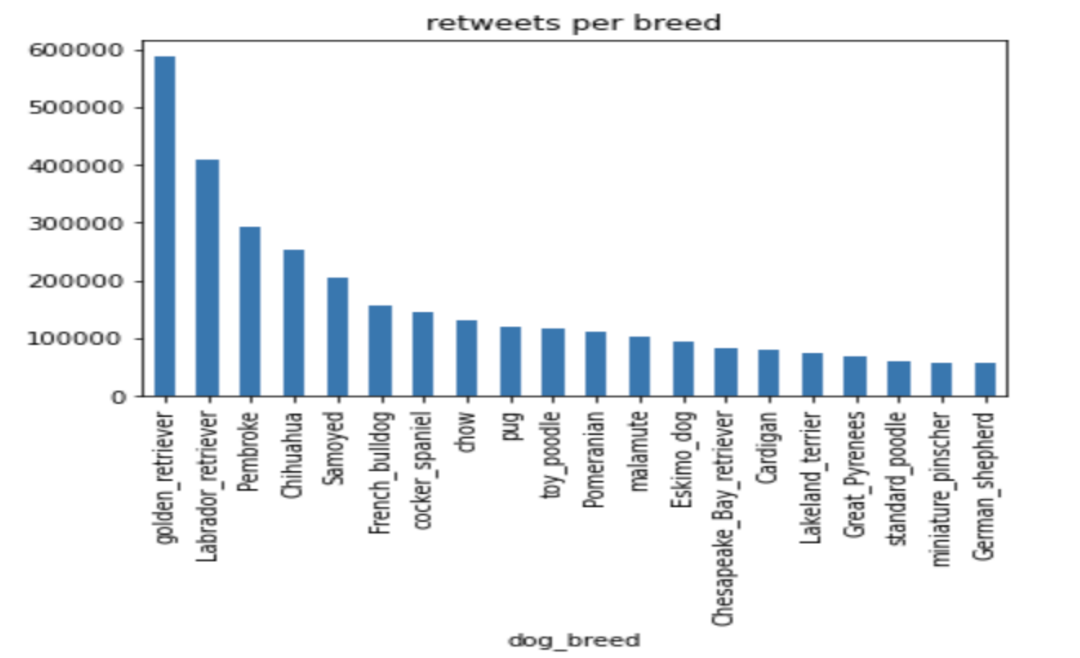
Let's take a look at this master dataset below:

	tweet_id	timestamp	source	text	expanded_urls	rating_numerator	rating
0	892420643555336193	2017-08-01 16:23:56+00:00	href="http://twitter.com/download/iphone" r...	This is Phineas. He's a mystical boy. Only eve...	https://twitter.com/dog_rates/status/892420643...	13	
1	892177421306343426	2017-08-01 00:17:27+00:00	href="http://twitter.com/download/iphone" r...	This is Tilly. She's just checking pup on you....	https://twitter.com/dog_rates/status/892177421...	13	
2	891815181378084864	2017-07-31 00:18:03+00:00	href="http://twitter.com/download/iphone" r...	This is Archie. He is a rare Norwegian Pouncin...	https://twitter.com/dog_rates/status/891815181...	12	
3	891689557279858688	2017-07-30 15:58:51+00:00	href="http://twitter.com/download/iphone" r...	This is Darla. She commenced a snooze mid meal...	https://twitter.com/dog_rates/status/891689557...	13	
4	891327558926688256	2017-07-29 16:00:24+00:00	href="http://twitter.com/download/iphone" r...	This is Franklin. He would like you to stop ca...	https://twitter.com/dog_rates/status/891327558...	12	

```
RangeIndex: 1994 entries, 0 to 1993
Data columns (total 23 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   tweet_id              1994 non-null   int64  
 1   timestamp             1994 non-null   object 
 2   source                1994 non-null   object 
 3   text                  1994 non-null   object 
 4   expanded_urls         1994 non-null   object 
 5   rating_numerator      1994 non-null   int64  
 6   rating_denominator    1994 non-null   int64  
 7   name                  1990 non-null   object 
 8   retweet               1994 non-null   int64  
 9   favorite              1994 non-null   int64  
 10  jpg_url               1994 non-null   object 
 11  img_num               1994 non-null   int64  
 12  p1                    1994 non-null   object 
 13  p1_conf               1994 non-null   float64
 14  p1_dog                1994 non-null   bool   
 15  p2                    1994 non-null   object 
 16  p2_conf               1994 non-null   float64
 17  p2_dog                1994 non-null   bool   
 18  p3                    1994 non-null   object 
 19  p3_conf               1994 non-null   float64
 20  p3_dog                1994 non-null   bool   
 21  dog_stage             326 non-null    object 
 22  dog_breed             1994 non-null   object 
dtypes: bool(3), float64(3), int64(6), object(11)
```

Research Question 1: *Which breed of dogs have the highest retweets?*

Now let's examine which of these breeds gets more retweets

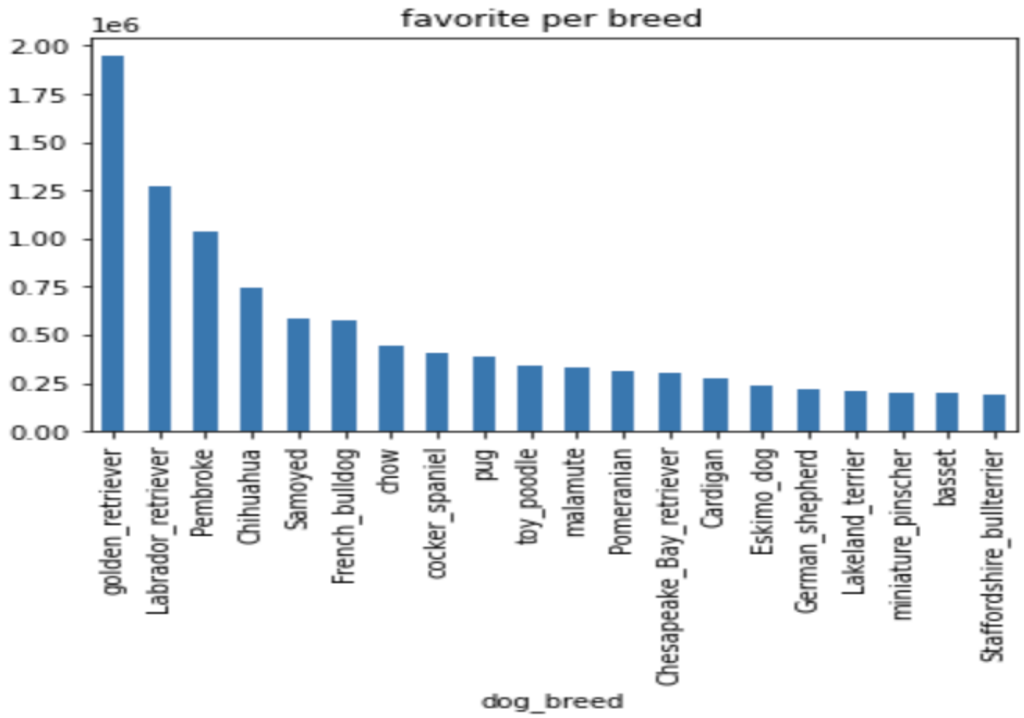


We see that post with golden_retriever has the highest retweets

---

Research Question 2: *Which breed of dogs is people's favorite?*

We would like to know which dog breed get's the most likes

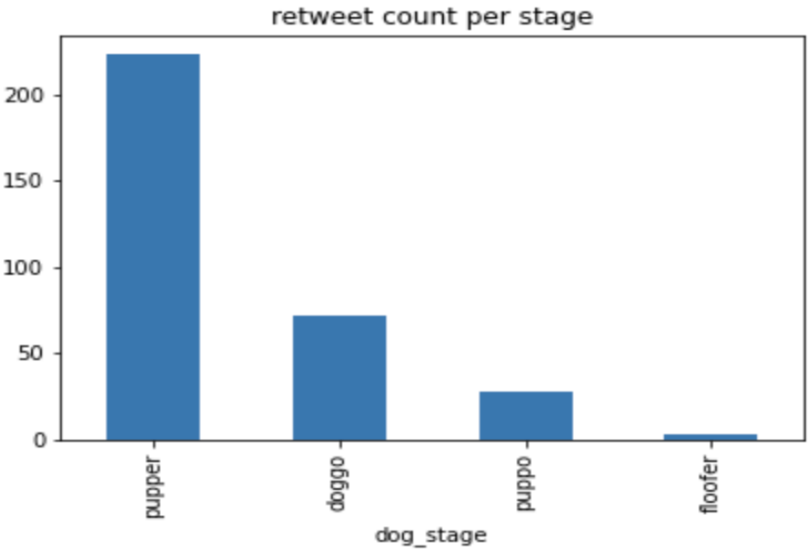


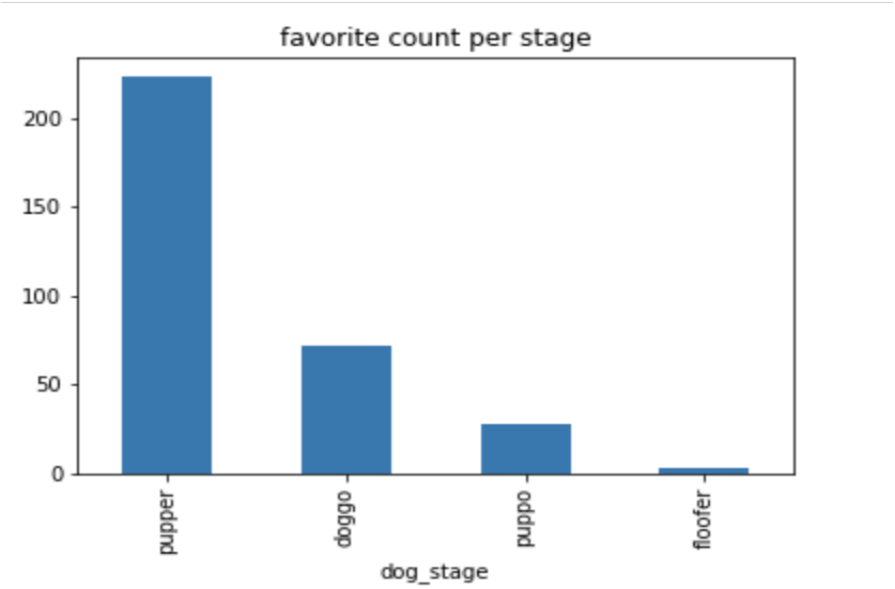
We also see that the golden_retriever is people's favorite breed

---

Research Question 3: *what stage of the dog do people find most adorable?*

We also engineered a column called dog_stage during the wrangling and would like to know which growth stage people find most adorable





Judging by the retweets and likes per stage we can say that people find puppers most adorable

---

Conclusion

This concludes our exploratory analysis of the master dataset.