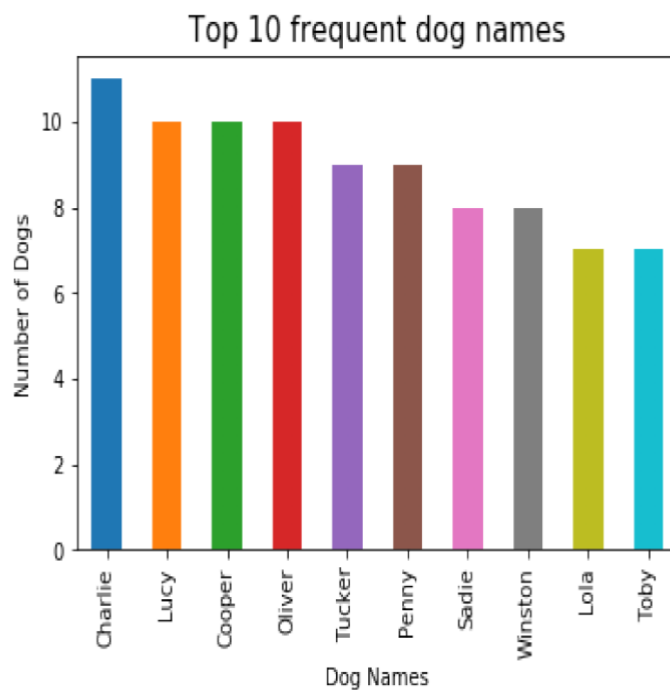


act_report

- **Findings of the analysis**

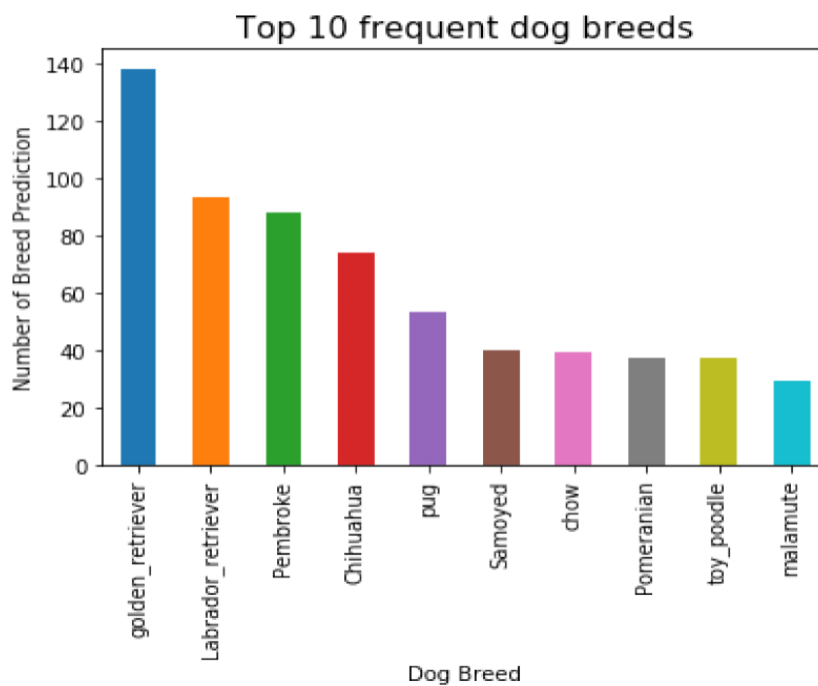
1. Most common name in this dataset is Charlie.
2. There is a positive relationship between favorite count and retweet count.
3. Most of the dogs are of names: Charlie, OLiver,Cooper, Penny, Tucker, Lucy, Sadie, Winston, Daisy,Lola
4. The pred_breed column is created based on the the confidence level of minimum 20% and 'p1_dog', 'p2_dog' and 'p3_dog' statements
5. Based on dog types: doggo, floofer, pupper, puppo, 'doggo, puppo', 'doggo, pupper', 'doggo, floofer', only one categorical column is created named as 'stage'
6. tweet_id is set as object type as it is not going to use for calculation.
7. A Master dataframe is created using df_clean, image_df_clean, and tweet_json_clean dataframes
8. Dog Names Issue got rectified
9. Inconsistency in pred_breed got removed
10. All retweets get deleted to get unique tweets
11. The columns such as in_reply_to_status_id, in_reply_to_user_id, retweeted_status_id, retweeted_status_user_id, and retweeted_status_timestamp is removed which is not needed
12. Timestamp format got corrected to datetime format
13. Extra HTML tags from source column get refracted
14. Dog ratings get standardized for denom of 10.

Most of the dogs are of names: Charlie, OLiver,Cooper, Penny, Tucker, Lucy, Sadie, Winston, Daisy,Lola



- Charlie 11
- Lucy 10
- Oliver 10
- Cooper 10
- Tucker 9
- Penny 9
- Sadie 8
- Winston 8
- Daisy 7
- Toby 7

Most of the dogs have golden retriever, labrador retriever as breed which all are rated.



- golden_retriever 137
- Labrador_retriever 93
- Pembroke 88
- Chihuahua 74
- pug 53
- Samoyed 40
- chow 39
- Pomeranian 37
- toy_poodle 37
- malamute 29