

Act Report

By Mohammed Al Mujadib

Introduction:

After I cleaned Dataset, I started with Dataset analysis. I created file for the Whole datasets that I cleaned, merged them, named it master dataframe. The Result that I have in my dataset 27 columns and 2071 rows. columns name After merged:

- 1-tweet_id
- 2-in_replay_to_status_id
- 3-in_replay_to_user_id
- 4-timestamp
- 5-source
- 6-text
- 7-Retweeted_status_id
- 8-Retweeted_status_user_id
- 9-Retweeted_status_timestamp
- 10-expanded_urls
- 11-name
- 12-Rating
- 13-type
- 14-Tweet_id
- 15-Likes
- 16-RT
- 17-jpg_url
- 18-img_num
- 19-p1
- 20-p1_conf
- 21-p1_dog
- 22-p2
- 23-p2_conf
- 24-p2_dog
- 25-p3
- 26-p3_conf
- 27-p3_dog

Data Analysis:

**I explored data and I found that I have Three questions two of them
Two of them with visualization:**

Question1: What is the most popular dog in our dataset?

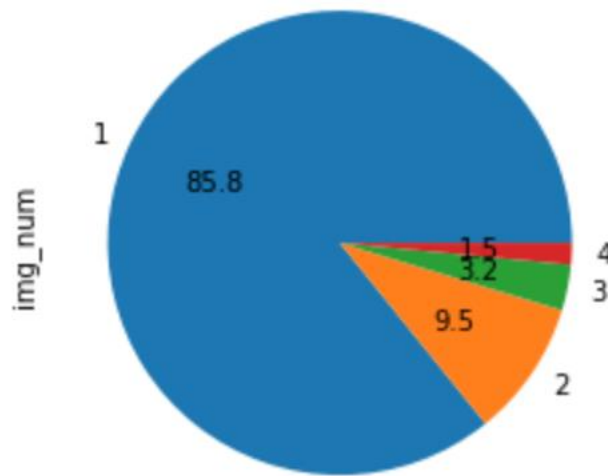
Awnser1: the most popular type is pupper with counts of 229 by far from the others. Then, doggo is second with counts of 75. After that, puppo is third with counts of 29. Last, floofer with counts of 3 only.

```
pupper      229
doggo        75
puppo        29
floofer       3
Name: type, dtype: int64
```

Question2: What is the most image number in each tweet?

Awnser2: I did visualization for it, I found that one image is the most Number by far from the rest with 85.8 %. Then, two images with 9.5 %. After that, three images with 3.2 %. Last, four images with 1.5 %.

What are the most number of images in each tweet



Question 3: What is the most source of tweets?

Answer 3: I found that the most source of tweets is twitter for iPhone by very far from the others with almost 2000 tweets. After that, twitter web clients come second with less 250 tweets. Then, tweetDeck comes at third. The last, Vine – make a scene.

