Act report

I found this picture of a dog named atticus and it got 4888 favorite, 2332 retweet and it is the highest rated with 1776 out of 10 this is her picture

Data collection
Gathering data from the required sources. Uploading the archive file and forecasting images and data from Twitter APi.

Data assessing
Evaluate data visually and programmatically and document problems in the data.

Quality assessing

- **Archive, Image Predictions and tweet table:**

1. Change the data type to a string

2. Convert id type to object

- **Archive table:**

3. Drope columns of (in_reply_to_status_id, in_reply_to_user_id, retweeted_status_id, retweeted_status_user_id, retweeted_status_timestamp, expanded_urls, timestamp)

4. There are two tweets (315, 1016) that do not contain pictures of dogs

5. Two of the photos got zero reviews (315, 1016)

6. Extracting the types of sources used in writing tweets

7. Convert rating_denominator to float

8. Remove the dog names lowercase values

- **Image Predictions table:**

1. The first letters are lowercase in p1, p2, p3

2. remove the img_num column

- **Tweet table:**

1. Drope columns of (id_str, truncated, entities, extended_entities, in_reply_to_status_id, in_reply_to_status_id_str, in_reply_to_user_id, in_reply_to_user_id_str, in_reply_to_screen_name, user, coordinates, contributors, is_quote_status, possibly_sensitive, possibly_sensitive_appealable, retweeted_status, quoted_status_id, quoted_status_id_str, quoted_status, favorited, retweeted, friends_count, listed_count, geo, time_zone, place, created_at, place, profile_background_tile, photo, profile_text_tile, withheld_in_countries, source, hashtags)

2. Convert favorite_count retweet_count float to integer

Tidiness assessing

- **All tables**

The data is divided into 3 tables

- **Archive table:**

The type of dog (doggo, floofer, pupper, puppo) can be collected in one column

- **Image Predictions table:**

Collect the 3 columns of (p1, p2, p3) in one column dog types

Data cleaning
Clean and fix documented issues to get data ready for analysis.
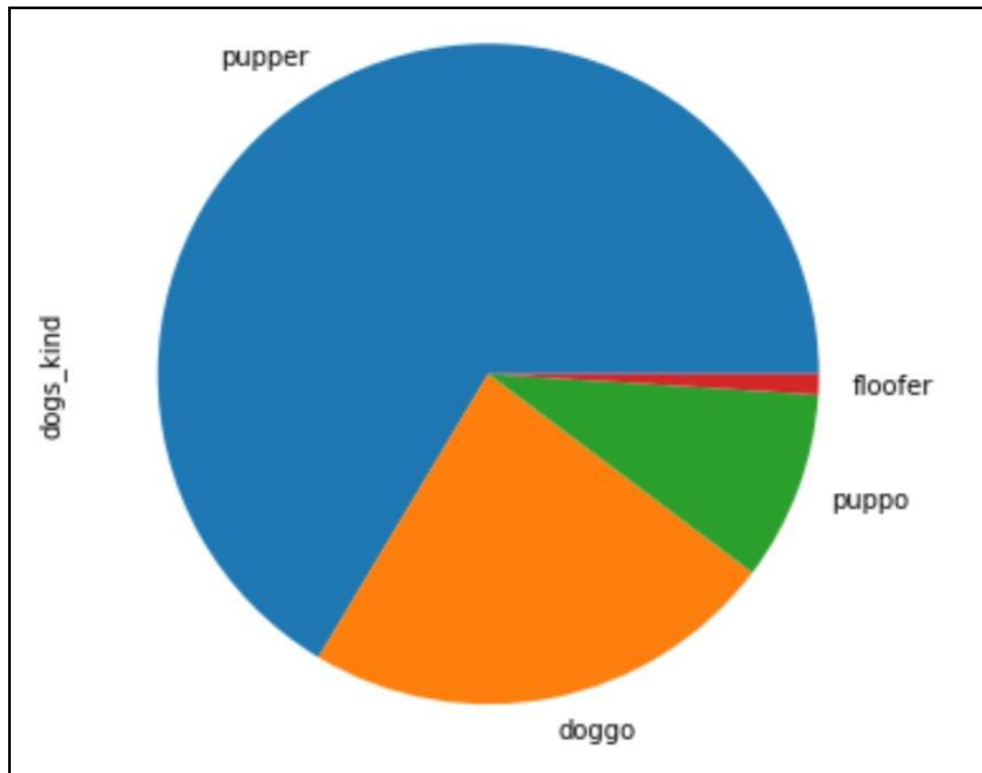
Data storage
Store all data after the cleaning process in a new file.

Visualization
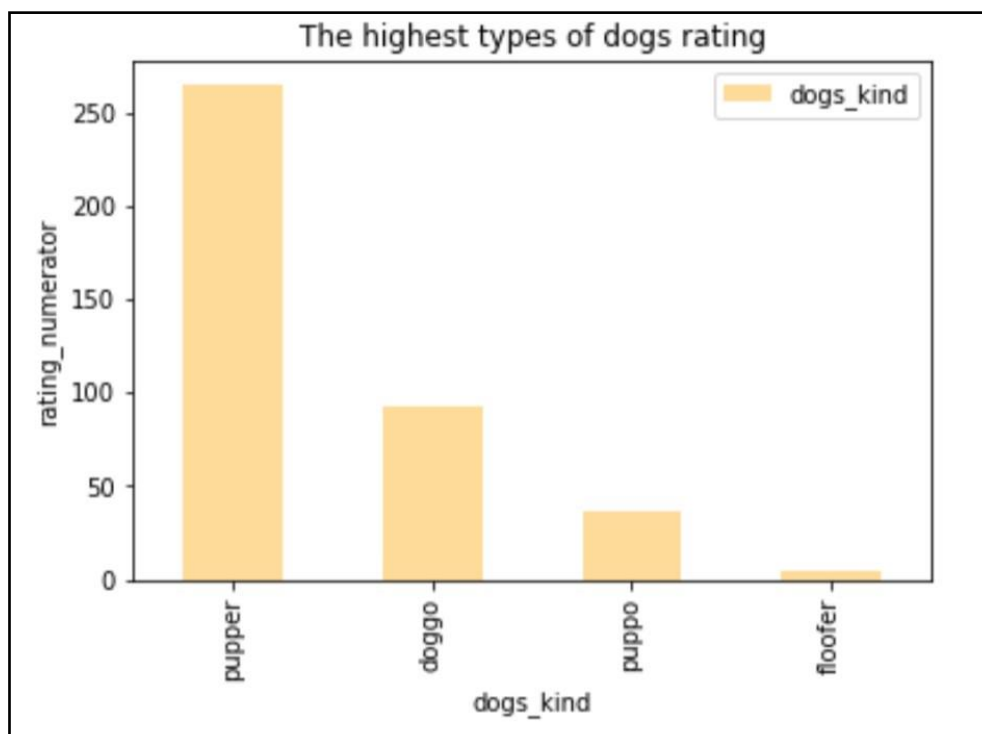Analyze data with 3 insights and visualization

1. Top types of dogs?
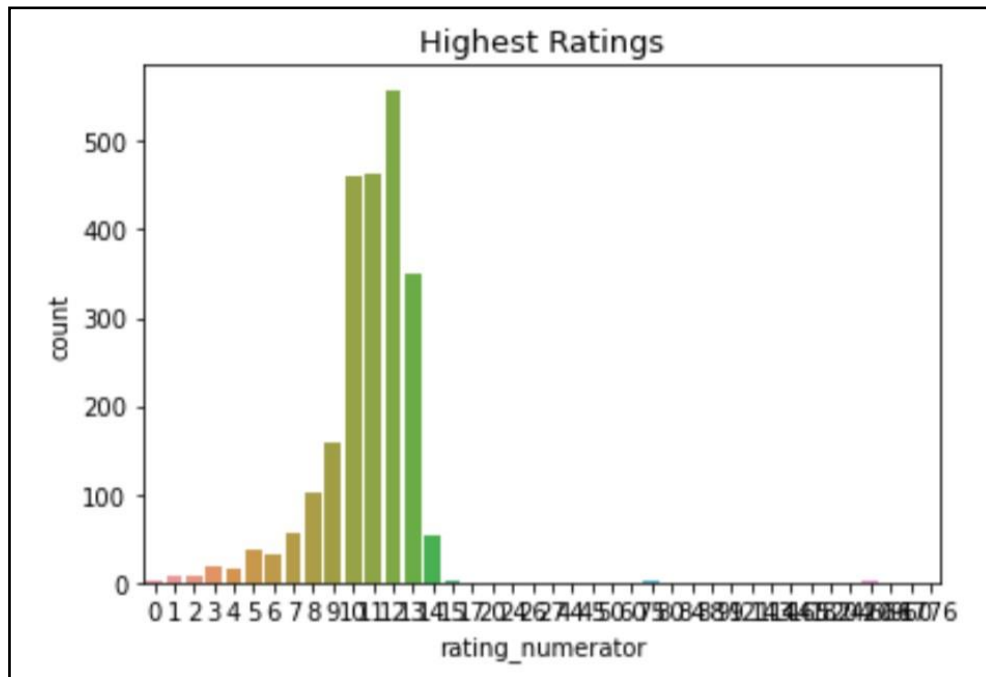
The bobber is the highest type of dog

2. What kind of dogs have highest rating numerator?

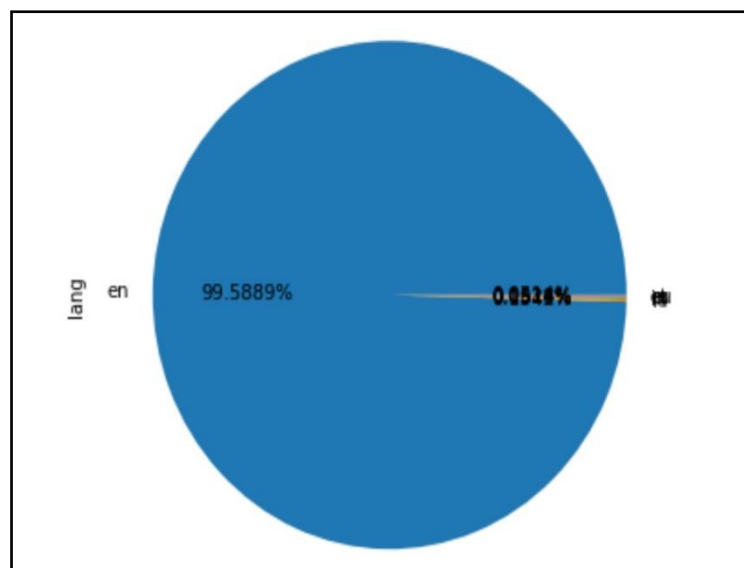The bobber is also the most popular type of dog

3. Highest ratings?

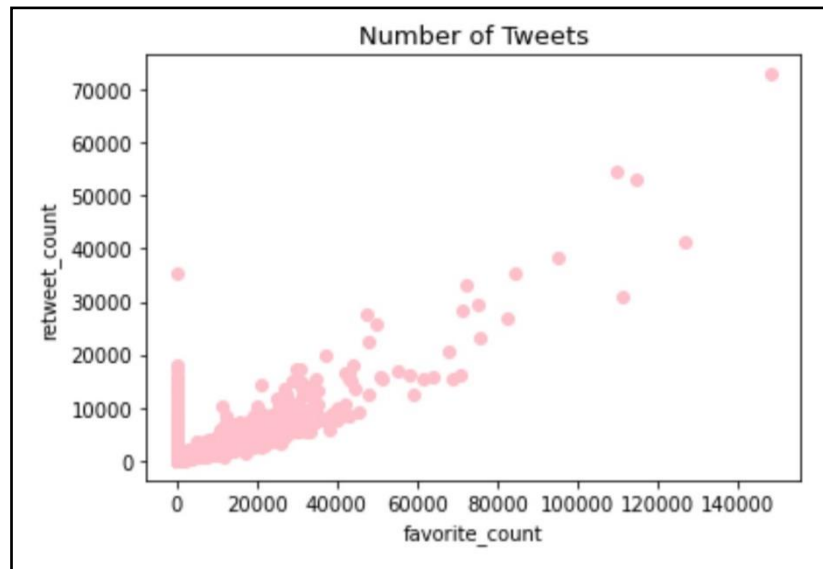The most pictures got a rating equal to 12 out of 10

4. What languages are used in tweets?

English was the only language used to write tweets

5. Compare the number of likes with the number of retweets

It turns out that the rate of likes is greater than the rate of retweets

6. Types of sources used in Twitter?

The highest percentage was from the iPhone