**DATA ANALYSIS REPORT**

***By: Asimiea Nicholas Walson***

**A black and white dog

Description automatically generated with medium confidence**

After cleaning and merging the three datasets into a master dataset, it was saved to a csv file named **twitter\_archive\_master.csv.**

The following steps were taken to draw insight from the data.

**Read the master dataset into a pandas dataframe named df\_master**

**Graphical user interface, text

Description automatically generated**

**Establish the source of most of the tweets in the dataframe**

**Graphical user interface, text, email

Description automatically generated**

**Establish the most popular dog breed from p1, p2 and p3 columns respectively**

**Graphical user interface, text

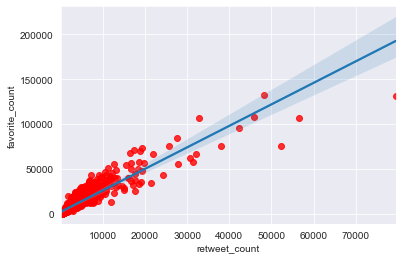
Description automatically generated**

**Insights:**

1. 98% of the tweets were made using an iPhone
2. Most frequent rating numerator is 12 and most frequent rating denominator is 10.
3. Most popular dog breeds are golden retriever followed by the labrador retriever before the chihuahua.

**Visualization:**

Question: Is there a relationship between retweet count and favorite count?



The plot above attempts to visualize the relationship between retweet\_count and favorite\_count.

We can see there is a positive linear relationship between the two series which implies the tendency of increment in either retweet\_count as favorite\_count increases and vice versa. This would not always be, but most likely be the case.