Project Act Report

Introduction

The data used for analysis was the combined twitter archive, image and tweets data. Afater merging the dataframe had 1917 rows and 13 columns

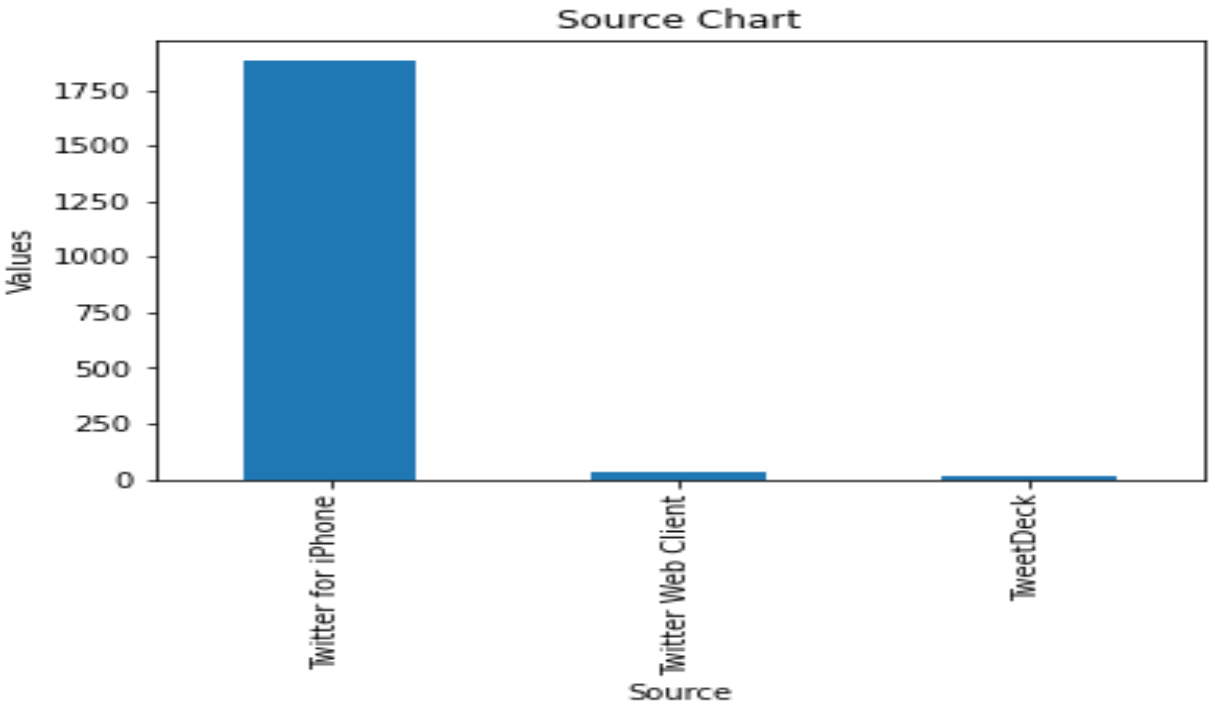
Analysis

Questions

1. What was the most popular source of tweeting from?
2. Which breed of dog was the most popular based on the number of retweet?
3. Which breed of dog had the highest number of favorite count?
4. Which dog is the best dog breed based on the rating, retweet and favorite count?

Answers and Insights

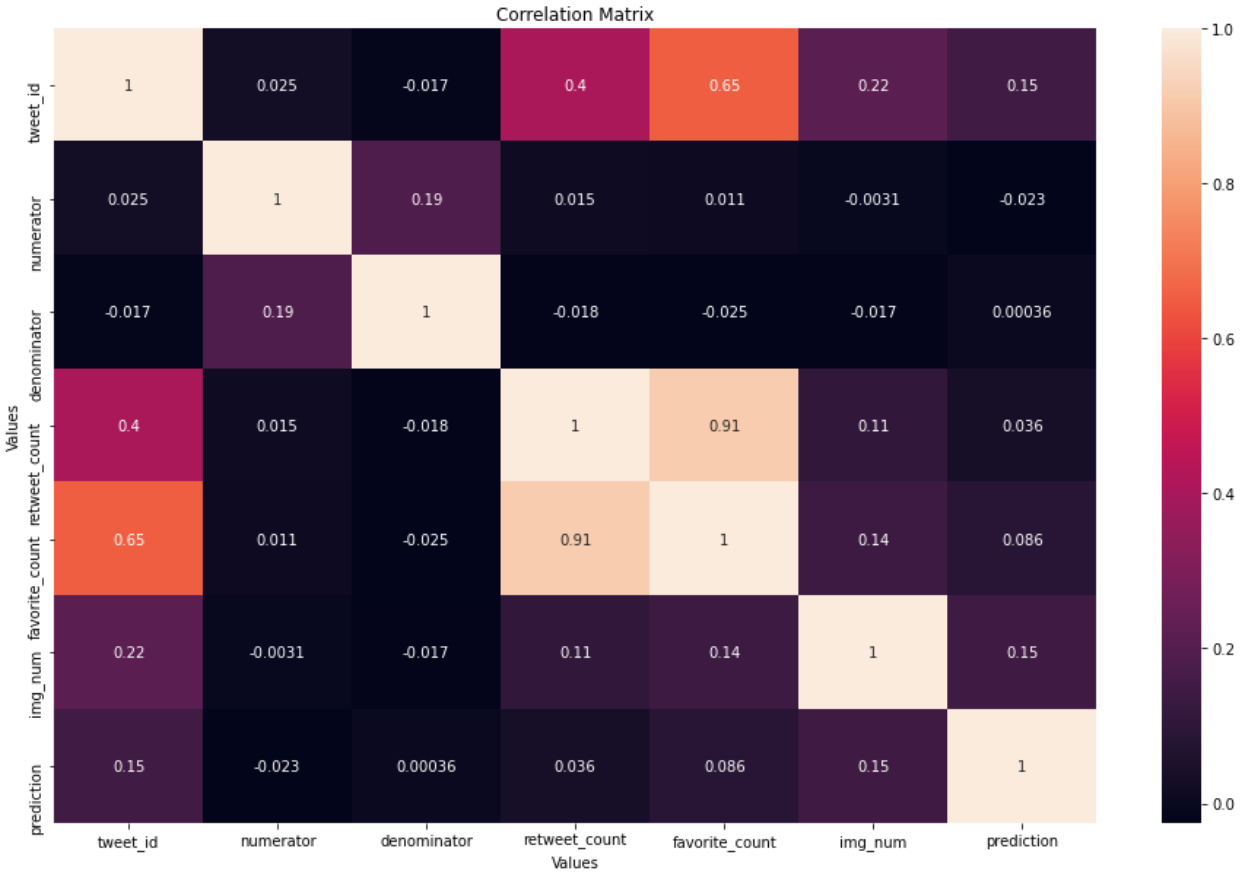
1. The most popular source of tweet ratings was from **twitter for iphone**. It could point to the most popular mobile phone in the population.



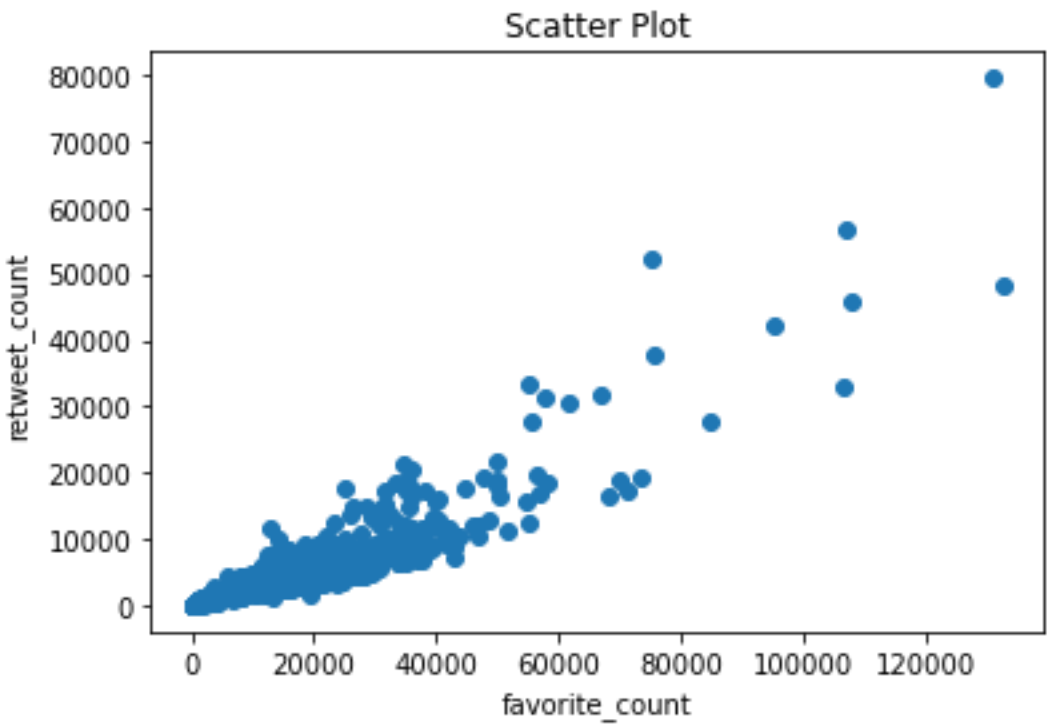
2. Which breed of dog was the most popular based on the number of retweet? The **Labrador_retriever** breed won this prize

3. Which breed of dog had the highest number of favorite count? The **Lakeland_terrier** won this particular prize

I plotted a heatmap to check for any correlation between the variables, and it was noticed that there is a **positive correlation** between the retweet_count and favorite_count. This is expected as people would naturally retweet what they like or resonate with.



I plotted a scatter plot to show the positive correlation between the retweets and favorite count, and it is shown that as the **retweet count increases, the favorite count also increases**.



4. Finally, I wanted to know the most popular dog breed based on the rating numerator, retweet and favorite counts. At the point I assumed it must really be a special dog breed and yes, I was right! The **Lakeland_terrier** brought it home!! Drum rolls. I would [ick the dog as my favorite too because the dog is cute.

Now, I have an idea of the dog to buy for myself.