

Data Analysis and Visualization is the last part of the project, as we start extracting meaningful and useful insights from our cleaned data and back them up with visualization whenever possible.

And the insights I thought to be useful were:

1- Which tweet has the most retweet count, and which one was the most favorite:

And as you might suspect the most favorite tweet was also the most retweeted one. Where tweet id number '744234799360020481' had a favorite count of '154225' and retweet count of '76485'.

2- What is the most common dog breed in the algorithm's first prediction:

The Golden Retriever came in the first place with a total count of '138', while the Labrador Retriever came second with a total count of '93'.

3- What is the 2nd most common dog breed:

The Labrador Retriever came in the second place with a total count of '96', while the Golden Retriever came second with a total count of '81'.

4- Investigating the dog breeds which the algorithm's first prediction was over 90% confident:

Here I tried to see which dog breed the algorithm was most confident about, and it showed that the Golden Retriever had the highest count. However, while the Labrador Retriever was the second most common dog breed in the algorithm's first prediction, it came in the 6th place in the first prediction confidence. Where Pembroke, Bug, Samoyed, and Pomeranian landed in the 2nd to 5th place respectively.

As for the visualization part, I saw it fitting to see the relation between the rating and the retweet count. While it seems fair to think that the highest rating would correspond the highest retweet count, the graph showed a little bit difference. As the highest retweet count was for the ratings approximately equal to 1.3.