PURPOSE: COMMUNICATING THE FINDINGS AND PRESENTING VISUALIZATIONS FROM OUR DATA

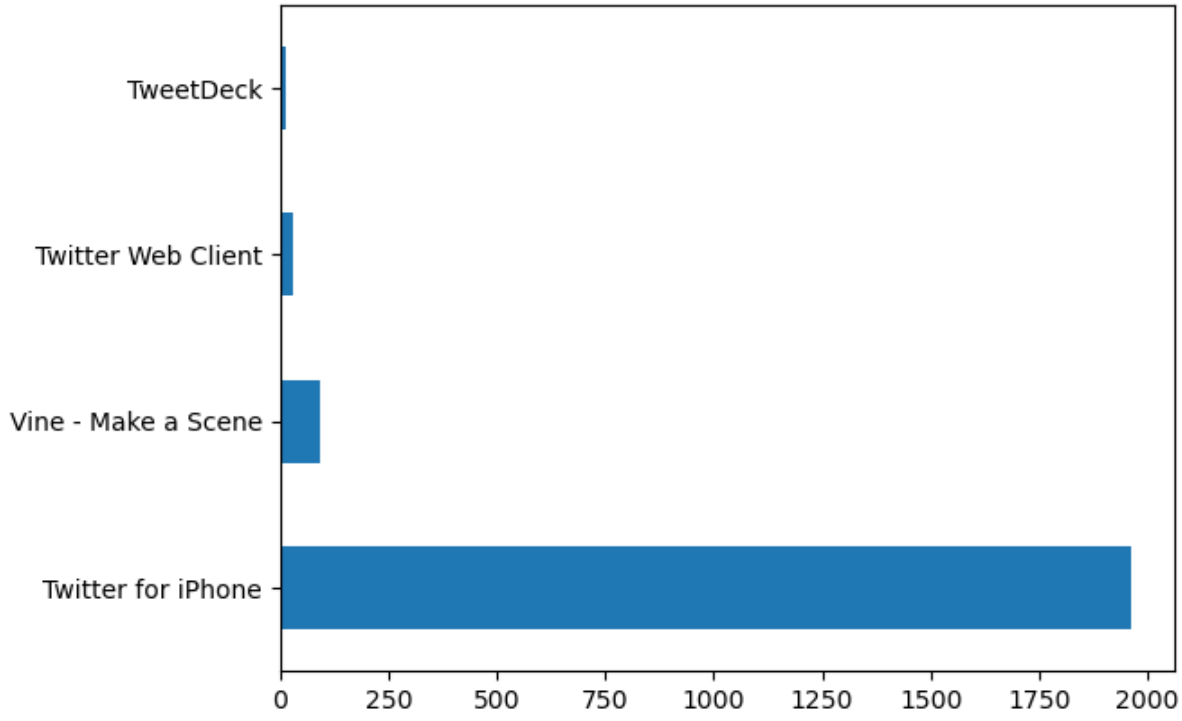
Following the data gathering, assessment and cleaning that was made, the following we came up with three insights, one of which had a plot (visualization generated).

The following were the insights:

1. Which is the source from which users tweeted from most?
2. Which breed of dogs had the highest ratings and what were the ratings?
3. Which language is most used by those on this site in their tweets?
4. How accurate was our model in its predictions?

INSIGHT 1:

To find out which source most tweets came from, a bar chart was plotted using matplotlib



From this visualization, we could conclude that most tweets came from iPhones.

INSIGHT 2:

To discover which breed of dogs had the highest ratings, we used the function *.nlargest* to return the top 10 highest ratings which led us to discover that twitter apparently has a thing for retrievers and terriers!

INSIGHT 3:

In order to resolve this, we used the function value_counts on our language column to get the count of each language, which led us to discovering that English was the dominant language.

INSIGHT 4:

Lastly, our final Insight was on knowing how accurate was our model's prediction. This was assessed by using pandas .describe() function to get a statistical summary of our conf_degree column(This column displays the degree of confidence of each accurate prediction that was made). From this summary we observed that though our standard deviation is lower than the mean, it is more than 0.5 of its value...relatively it could be considered large and can be seen as an indicator that our model is not performing optimally and further investigation could be beneficial.