

# Project 2 - Data Wrangling

Alx-t programme- UDACITY





## **Insight 1 - Most common dog name given to dogs by their owners**

Every dog owner calls their dog by a name just like any other owned pet. From the WeRateDogs dataset, we gained a few insights, we observed there are some common names given to dogs. The name Charlie tops the table of the most used or popular dog name. Cooper, Oliver, Lucy, Tucker, Penny, Sadie, Winston, Lola, Daisy, Toby follows respectively.

There are a number of dog names but these few, seems to be the most popular amongst the lot.

## Insight 1 - Most common dog name given to dogs by their owners

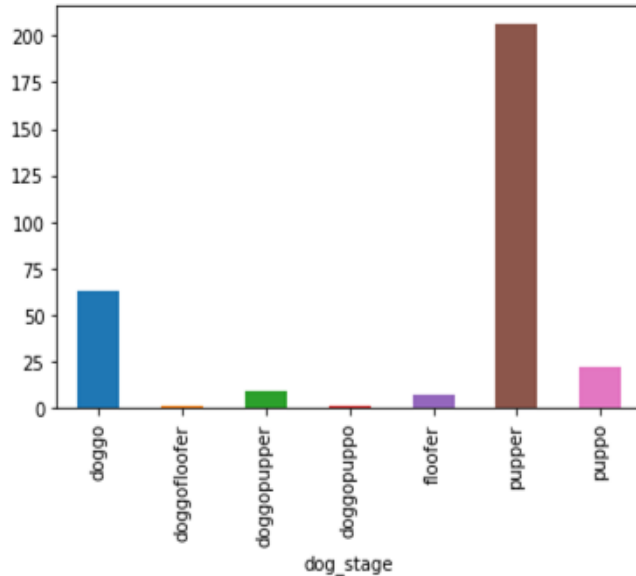
```
Out[9]: Charlie      11
        Cooper       10
        Oliver       10
        Lucy         10
        Tucker        9
        Penny         9
        Sadie         8
        Winston       8
        Lola          7
        Daisy         7
        Toby          7
        Bo            6
        Koda          6
        Jax           6
        Bella         6
        Stanley       6
        Rusty         5
        Leo           5
        Chester       5
        Oscar         5
        Bailey        5
        Scout         5
        Dave          5
        Milo          5
        Louis         5
        Buddy         5
        one           4
        Bentley       4
        Sophie        4
```



## **Insight 2 - In the 'dog\_stage' column, the 'pupper' emerged with the highest favourite\_count.**

The second insight, the pupper is seen in the bar chart as the favorite dog stage on twitter. We arrived at this insight using the bar chart to visually communicate and verify out insight . Daggo comes after pupper as the second highest favorite count, which too is visible from the chart.

**Insight 2**-In the 'dog\_stage' column, the 'pupper' emerged with the highest favourite\_count.



## Insight 2 - the picture of pupper with the highest favorite count on twitter.



Out[188]:





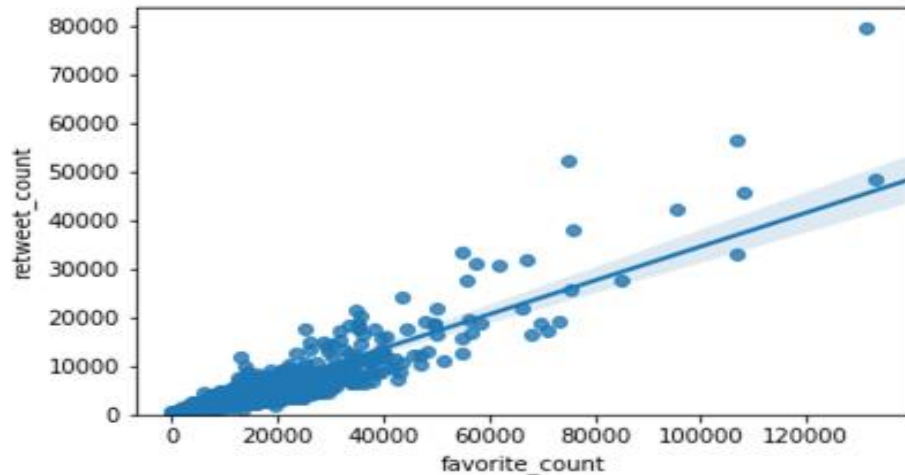
## **Insight 3** - positive linear relationship between favorite\_count and retweet\_count.

There seem to be a positive linear relationship between favorite\_count and retweet\_count. Meaning the number of likes on a post directly affects retweet counts.

The higher or greater the favorite count, the higher the retweet count.

The linear regression plot was used in visualising this insight

## Insight 3 - positive linear relationship between favorite\_count and retweet\_count.







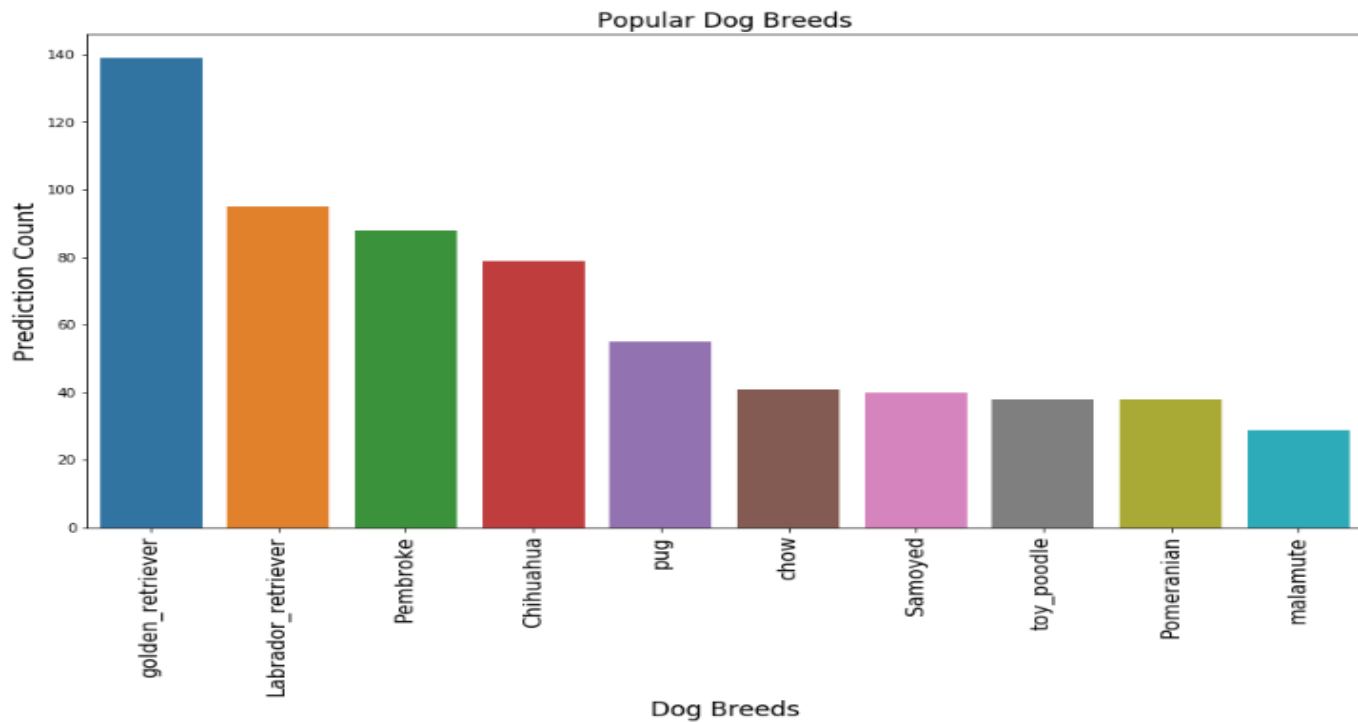
# VISUALISATION

The visualisation in the previous slide displays a barplot showing the most popular dog in our dataset. Our top dog breed appears to be the “golden retriever”, followed by labrador retriever, then the pembroke.

The results of the bar plot are hierarchical making it visually comprehensive.

The end.

# VISUALIZATION



it appears the 'Golden retriever' is the most popular dog in this dataset.