UDACITY

أكاديمية مسك
MISK ACADEMY

Project #5

Ghadi Alowaimer

Data Analyst

Wrangle and Analyze Data

“Act Report”

Introduction :

Real-world data rarely comes clean. In this project I will use Python and its libraries, I will gather data from a variety of sources and in a variety of formats, assess its quality and tidiness, then clean it. This is called data wrangling.

The dataset that I will be wrangling (and analyzing and visualizing) is the tweet archive of Twitter user **@dog_rates**, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

Insights :

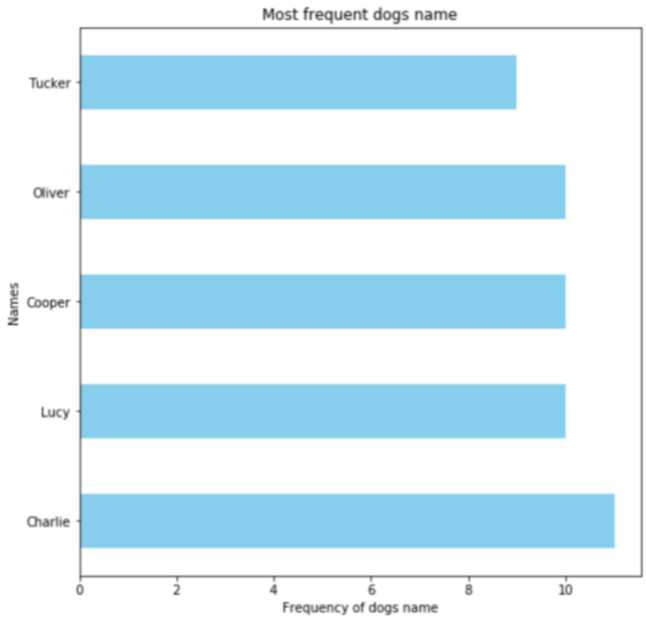
1-Group the image by the image number inorder to see the most frequent image. I found that image number 1is most frequent

2-Counting the boolean values for p1_dog, p2_dog and p3_dog columns to check which prediction algorithm is predicting well. Prediction algorithm number 2 was the best

3-Five number summary for both the rating numerator column and rating_denominator column

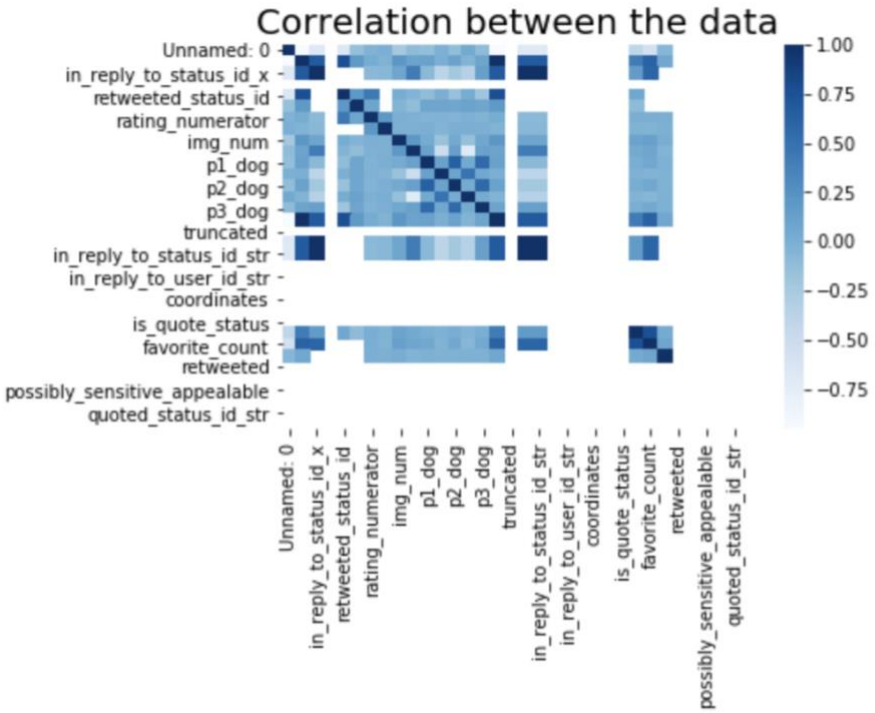
Visualization :

Most frequent dogs name



I observed that Charlie is the most frequent dog name

Correlation between the data



There is strong correlation between some attributes

Trying to extract the first image