Shahad Nasser Asseri
Project 5:
Wrangle and Analyze Data

(Act Report)

July 23, 2020
UDACITY
Data Analysis Nanodegree

Introduction :

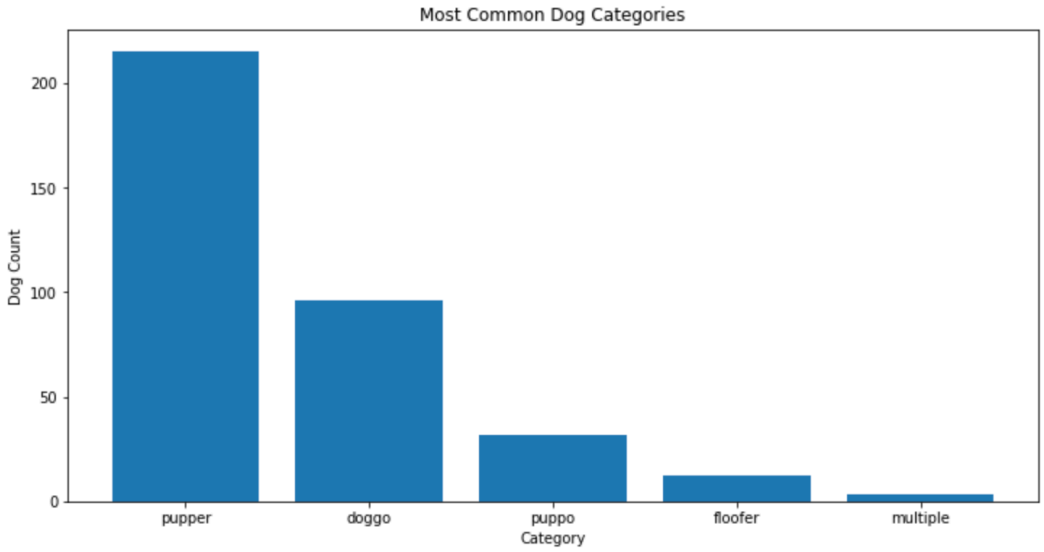
In this project I will use Python and its libraries such as (pandas, numpy, matplotlib, requests, json, os and seaborn).first I will gather data from a different of sources and in a variety of formats such as (Json,text file,API) to complete this project, then assess its quality and tidiness, and last step is clean data all these steps collated under data wrangling,then store it and visualise it, this report contain two visualises one most common dog categories and Correlation between the data.

The dataset that I will be wrangling (Sorting and analyzing and visualizing) is the tweet archive of Twitter user @dog_rates, also known as 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, but now in 2020 has over 8 million followers.

Most common dog categories:

visualization.

```
plt.rcParams["figure.figsize"] = [12, 9]
dog_type = ['pupper', 'doggo', 'puppo', 'floofer', 'multiple']
dog_counts = [215, 96, 32, 12, 3]
fig,ax = plt.subplots(figsize = (12,6))
ax.bar(dog_type, dog_counts, width = 0.8)
ax.set_ylabel('Dog Count')
ax.set_xlabel('Category')
plt.title("Most Common Dog Categories")
plt.show()
```



Insights Found After Analysis:

- The most popular dog type is a "pupper".
- The most popular dog based on image predictions is a Golden Retriever with 150.
- The median and mean ratings out of ten are 11 and 12.
- The median and mean retweet count are 1408 and 2976.
- The median and mean favorite count are 3864 and 8556.

Correlation between the data :