

CREATING AN ACT_REPORT DOCUMENT FOR THE INSIGHT AND VISUALIZATIONS PRODUCED FROM THE WRANGLLED DATA.

Datasets Used for the project;

1. Twitter_archive_enhanced.csv
2. Image_prediction.tsv
3. Tweet-json.txt.

The first dataset was provided by Udacity, it contained basic information about the tweets I worked on. I did not have to download programmatically since it was already provided so I downloaded it manually.

For the second dataset; a tsv file, it was already provided on the Udacity server and all I did was to programmatically download the file using the code screenshotted below;

```
# URL
url = 'https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions/image-predictions.tsv'
response = requests.get(url)

# Code to open a tsv file and save the response content
with open('image-predictions.tsv', mode='wb') as file:
    file.write(response.content)

# Read the TSV file
image_predict = pd.read_csv('image-predictions.tsv', sep='\t')
```

The third dataset needed me to create a twitter developer account and scrap using a Twitter Api created using python's Tweepy Library, unfortunately, I could not create due to some logging issues so I used the dataset Udacity procured for us.

The data set used for the insights and Visualization exercise was created by merging the three data sets after accessing, cleaning and scouring for quality and tidiness issues. The name of the data set is Twitter_archive_master.csv.

```
#Let us save to a folder named twitter_archive_master.csv
data.to_csv("twitter_archive_master.csv", index=False)
```

```
#Let us check if our code worked
data = pd.read_csv("twitter_archive_master.csv")
data.head(20)
```

After ensuring the data was saved appropriately, I made some insights in the data and visualizations.

INSIGHTS

1. The minimum rating numerator is zero, meaning our doges were always rated and the maximum rating numerator is 1776
2. The minimum rating denominator is Two (2), while the maximum is 170.
3. The name Charlie has the most appearances as dog names.

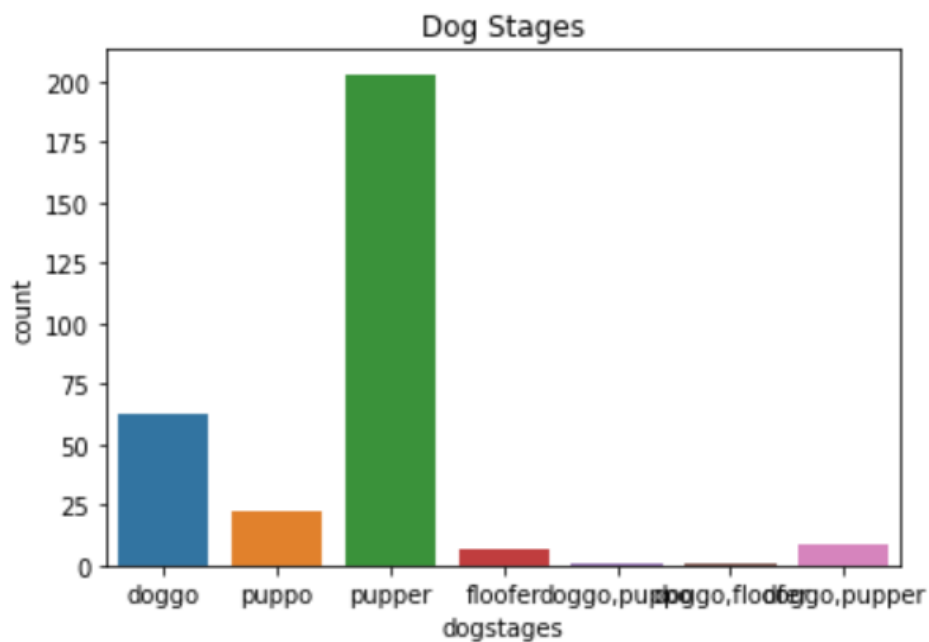
```
In [110]: data.name.value_counts()
```

```
Out[110]: None          644  
Charlie         11  
Cooper          10  
Lucy            10  
Oliver          10  
...  
Leela           1  
Glenn           1  
Shelby          1  
Sephie          1  
Christopher     1  
Name: name, Length: 914, dtype: int64
```

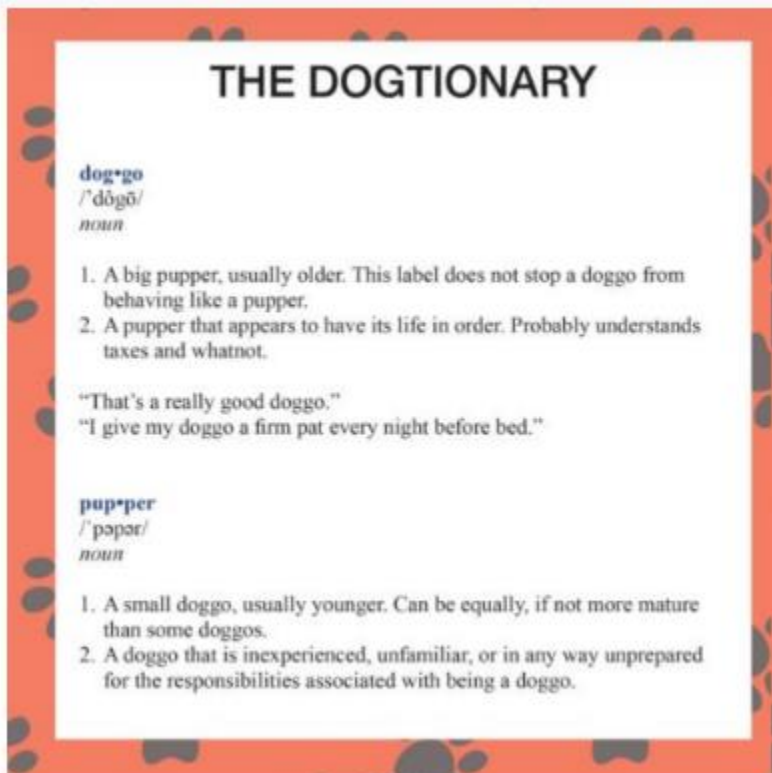
4. Img_number 1 appeared more frequently than most in the data.

VISUALIZATIONS

Since our project is more focused on data wrangling, I made one visualization to ascertain the most popular dog stages, which was Pupper;



From the chart above, pupper is the most popular dog stage in our data, and from our dogtionalary provided in the project, it is the stage where dogs are more cuter.



Thank you.