DATA ANALAYSIS REPORT ON WeRateDogs TWEET

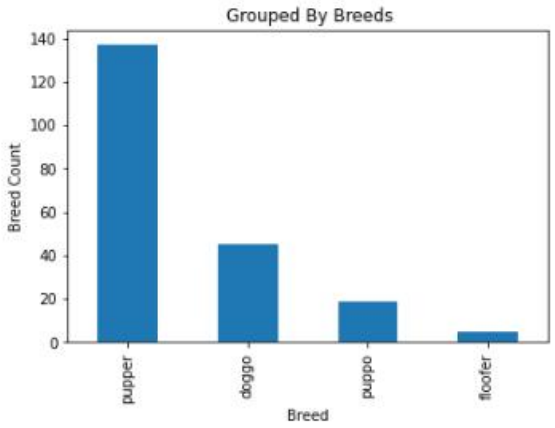
Data analysis entails the insights discovery journey and appropriate reach to a better and informed decision. There are two methodologies to analyzing data commonly known as qualitative data analysis and quantitative data analysis. For this task, the nature of data attracted the type of analysis to performed. Quantitative analysis entails analysis of values based variable, or otherwise categorical data. Quantitative data analysis entails data analysis on numerical data. For most of the , the analysis made was qualitative as with the nature of data, occasional was the quantitative analysis done.

Project Questions

1. What are the most occurring dogs

```
In [140]: merged.Breed.value_counts().plot(kind='bar')
          plt.title(" Grouped By Breeds")
          plt.xlabel('Breed')
          plt.ylabel('Breed Count')
```

```
Out[140]: Text(0, 0.5, 'Breed Count')
```



Analysis results showed that the most commonly desired breed to be pupper, it was followed by doggo, then puppo and lastly was the breed of floofer.

2 which breed had the most rating?

```
In [107]: tempdf=merged.reset_index()
          grouped_breed=tempdf.groupby(tempdf['Breed'])['rating_numerator', 'rating_denominator'].sum()
          # grouped_breed = grouped_breed.sort_values(by=[],ascending=False)
          grouped_breed.plot.bar()
          plt.title("Total By Breed Group")
          plt.xlabel('Rating Numerator || Rating Denominator')
          plt.ylabel("Total animals In the Breed Category")
```

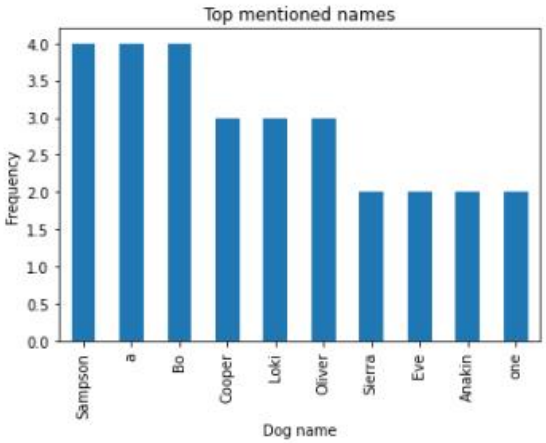
```
Out[107]: Text(0, 0.5, 'Total animals In the Breed Category')
```

Rating had the dimension of numerator and denominator, both appeared to have a correlation. Breed of pupper had the most of both dimensions followed by doggo, then ppuppo and lastly was the breed of floofer.

3 By name which dogs were the most mentioned

```
# visualizing the most mentioned dog
merged.name.value_counts().nlargest(10).plot.bar(rot=90)
plt.title("Top mentioned names")
plt.xlabel("Dog name")
plt.ylabel("Frequency")
```

```
Out[120]: Text(0, 0.5, 'Frequency')
```

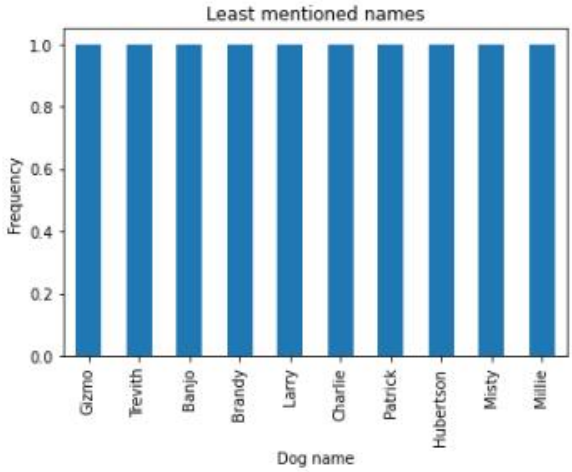


As observed from analysis, sampson was mentioned the most, a followed then bo and so forth.

4 What names were the least mentioned?

```
In [122]: # visualizing the least mentioned dog
merged.name.value_counts().nsmallest(10).plot.bar(rot=90)
plt.title("Least mentioned names")
plt.xlabel("Dog name")
plt.ylabel("Frequency")
```

```
Out[122]: Text(0, 0.5, 'Frequency')
```



As for this question there was a tie for the least names, the first one was gizmo, followed by trevith as observed in the image above.
As with the analysis performed, its clear to me that data analysis and insights discovery is as important as the data quality itself, I was surprised how slashed the data was after data cleaning.
Data collection mechanism to better collect and stored data needs to be engineered that data is cleaned before before stored, this will provide easy and short time for analysis.