Report

1.Introduction

After the wrangling of our data, now it's time to start analyzing and disecting it. On top of that, we need some visualization.

2.Analyzing and visualizing the data

Some of the question that intreaged me where :

Q1:What type of source does these tweets come from?

This question is very easy to exploreand answer because we already have the source column in the twitter_archive_mastered dataset. This mean that we can directly use the value_counts() methode to find the answer that we are looking for.

From the result we can see that 1845 people uses the iphone to tweet,25 are web clients, and 9 are using tweetdeck.

Q2:What type of dog breed has the biggest favorite count?
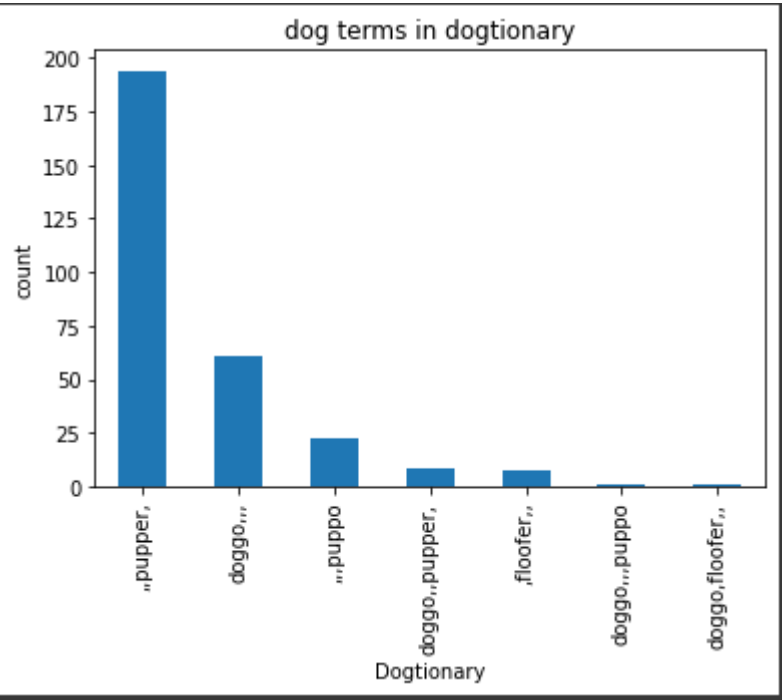
In this part, it is very interessting for us to know which dog breed is the favorite amoung our data because this could help us understand our data even more. using code we grouped the data and then sort it to find this result:

```
p1
golden_retriever      1638035
labrador_retriever    1093097
pembroke               954213
chihuahua              673876
samoyed                491651
                        ...
```

as we can see the golden retriever breed is the most favorite breed of dogs.

Q3: Which dogtionary term is the most used?

Just as the last question, knowing which dogtionary term is used could make understanding the dataset much easier, and visualizing it cwill make it even cleaner thats why we have :

here we see that the pupper term is the most used one out of the bunch which indicate according to the dogtionary that the favorite dogs are small, younger dogs are the favorites