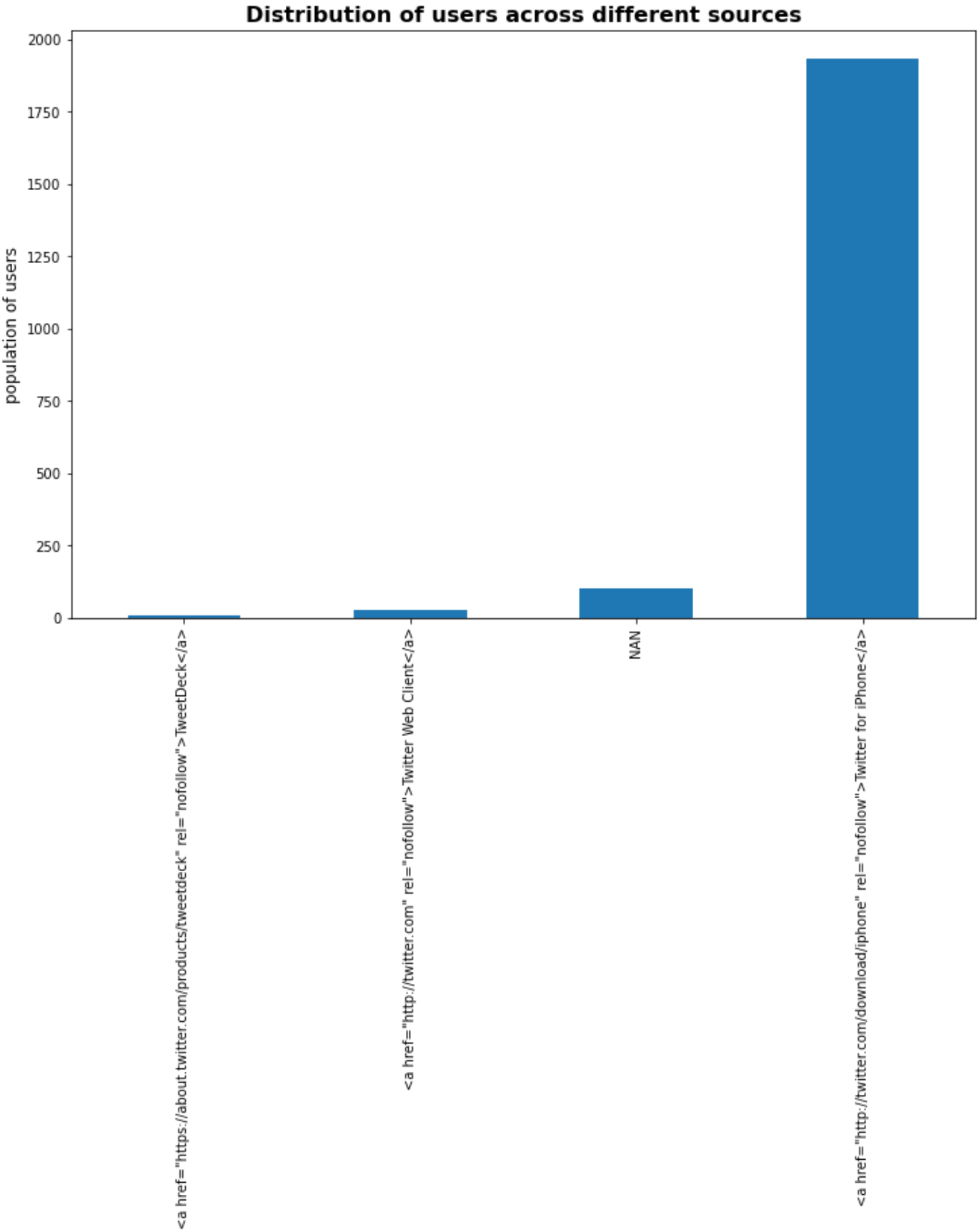
Introduction

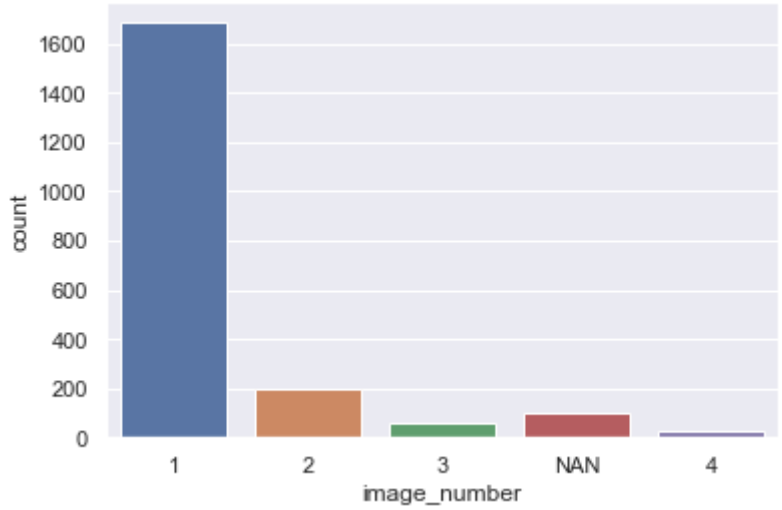
This report reveals the insights and visualiztion gotten from the wrangled data which was stored into a single dataframe called the master_twitter_archive

Analysis and Visualization

the first thing i looked into is the distribution of users among the various sources present in the data and from my graph it reveals that majoritynof the users are from iphone source



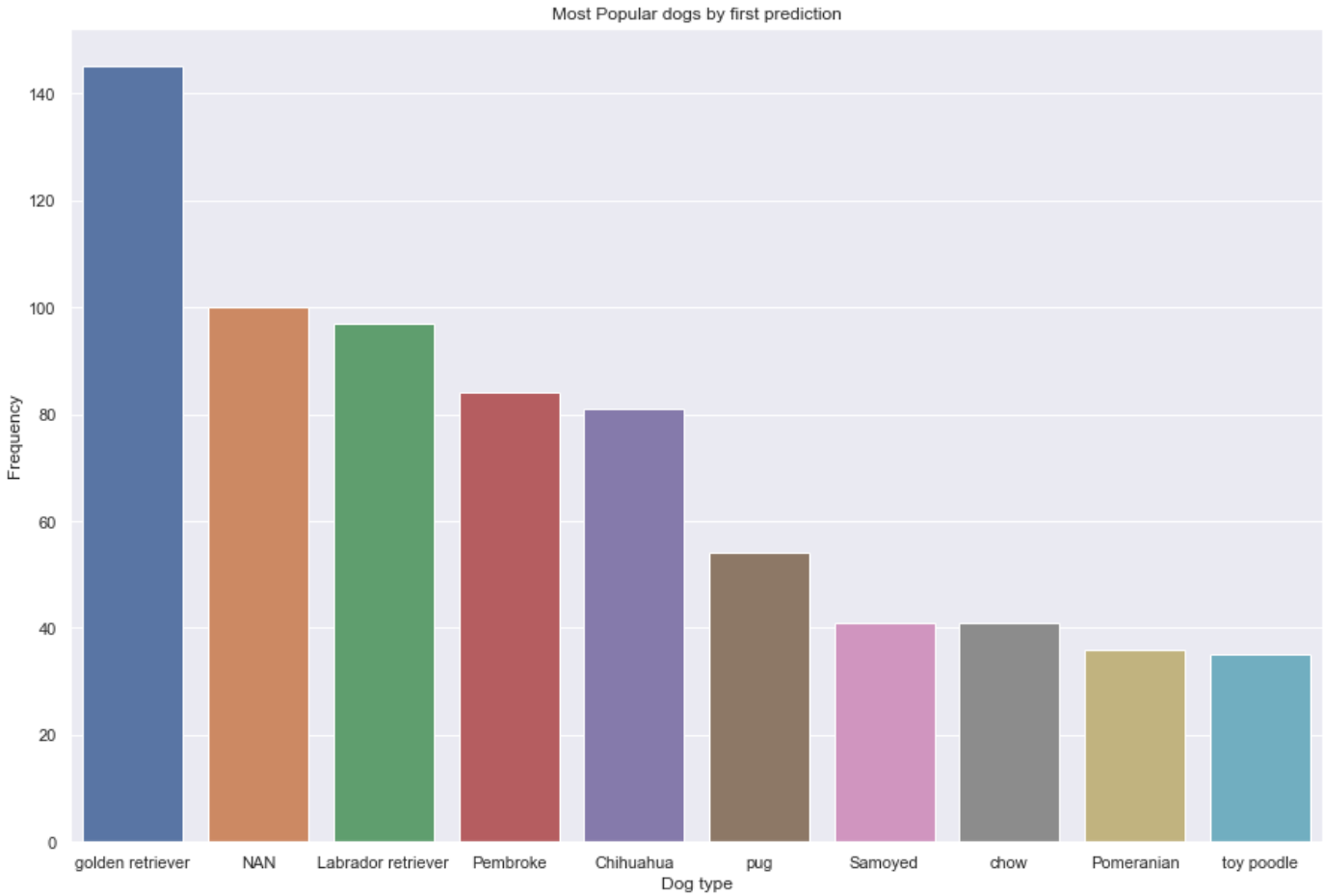
looking at the image number,the most occured image is 1 while there is 4 in some instance,though i think the case of 4 is an outlier but i cant solve at the moment



in the breed_type comprises of pupper,floofer,pupper and puppo,pupper has the highest count of more than 80% of the total breed count

```
pupper            201
NAN               100
doggo              65
puppo              23
doggo, pupper       9
floofer             7
doggo, puppo        1
doggo, floofer      1
Name: breed_type, dtype: int64
```

doing a seaborn count plot of the first prediction shows the dog breed goldern retriever as the most popular dog ¶



challenges: the dropna() method wasnt working i tried all i could but wasnt able to remove it thats the reason why NAN was still showing in my result

In [ ]: