Having gathered data from multiple sources concerning the WeRateDogs Twitter account archive. The data is assessed pragmatically and visually.

Visually, Microsoft Excel was used while pandas function like info and head were used to assess the data programmatically.

In the course of assessment, certain inconsistencies were noticed and cleaned to enable us analyze.

Here are some of them:

- For the twitter archive, the retweeted rows contain data for retweets, and this wasn't needed as we aren't concerned with retweets but as to how the dogs are rated and original tweets. These rows had to be dropped.

- The timestamp column which registered the time the tweet was made was not in time format. This was in string format and needed to be changed to the appropriate format which was datetime.

- The inreply columns was dropped as it contained data concerning replies to tweets and was also not important for this analysis.

- For the names of the dog, 'None' as a string was used in rows where the dog name wasn't know instead of using null values. This needed to be corrected to promote the data's tidiness.

- Upper and lower case inconsistencies in the p1,p2 and p3 columns as some ords started with capital letter and others did not. All the first letters were capitalized.

- The names in p1,p2,p3 columns also had underscores separating names instead of spaces. The underscore was replaced with spaces.

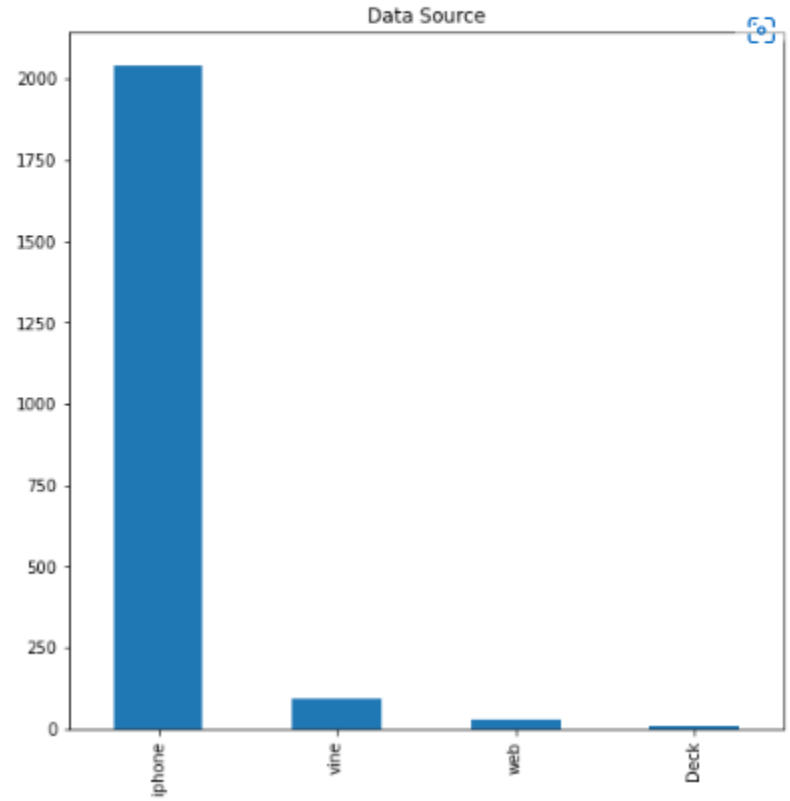
- The source column originally contained urls but this was not necessary as all that is needed is a description of the source of the data. The strings were extracted and replaced the urls. This is also aesthetically pleasing to the eyes

Upon cleaning the data and merging to one pandas dataframe we were able to get some meaningful insights form the now clean data.

We can see that the source for majority of the data was gotten from Twitter for iPhone, followed by Vine, Twitter for Web and Tweetdeck coming up last.

```
: # Illustrates the data source

labels = ('iphone', 'vine', 'web','Deck')
plt.title("Data Source")
df_copy.source.value_counts().plot(kind= 'bar', figsize = (8,8));
plt.xticks(np.arange(4), labels);
```



It is also notable that the dog stage with the highest likes is the popper stage

```
#dog stage with the highest likes
labels = ( 'doggo', 'floofer', 'pupper', 'puppo')
plt.title("Likes By Dog stages")
df_copy.groupby('value')['favorite_count'].count().plot(kind= 'bar',
                                                        color = ['grey', 'pink', 'red', 'indigo'], figsize = (8,8));
plt.xticks(np.arange(4), labels);
```