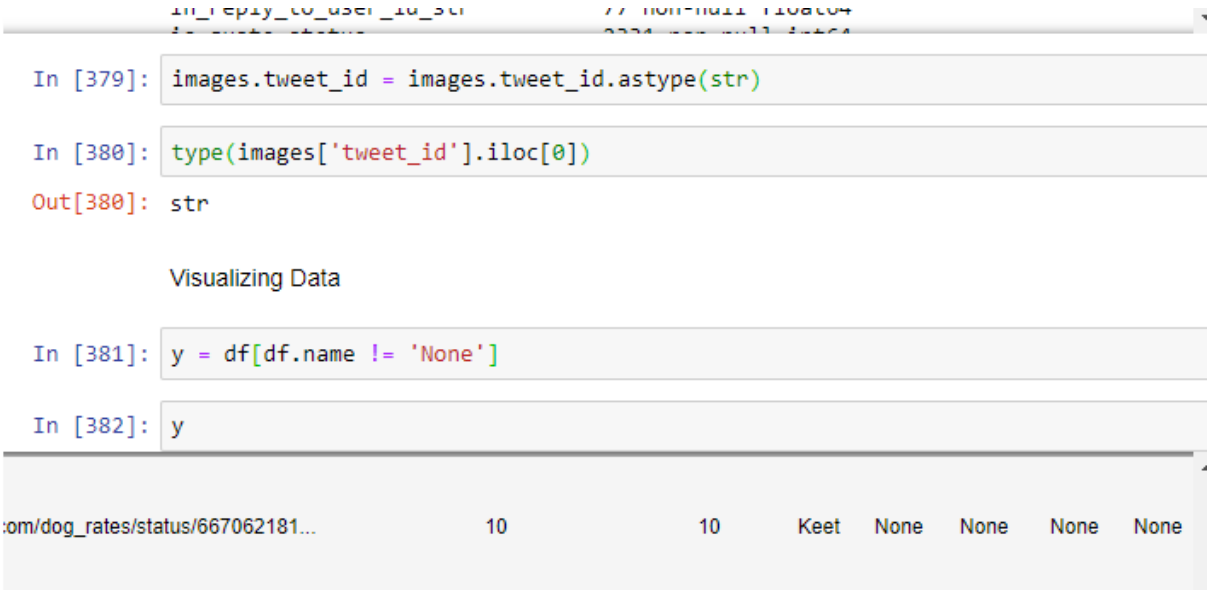
This report builds upon the previous report “wrangle report” that describes that the task being done is the Udacity wrangle and analyze data project which uses python programming language to apply Data analysis methods “We rate dogs’ ” data.

After data was cleaned. I did not know at the beginning what visual should I produce as there actually many columns that could be used. saved them to the workspace and made a visualization to show the most common dog names.

I thought after some time of making a graph to show the repeatability of the names and where is the trend going in dogs’ names. When I made the visual at first, I encountered a couple of problems. First some entries don’t have all the data entered, so some had the dog’s name as ‘None’.

So, I first made a new dataframe to remove rows having None in dogs’ names column as shown in line In [381].



```
In [379]: images.tweet_id = images.tweet_id.astype(str)

In [380]: type(images['tweet_id'].iloc[0])
Out[380]: str

          Visualizing Data

In [381]: y = df[df.name != 'None']

In [382]: y
```

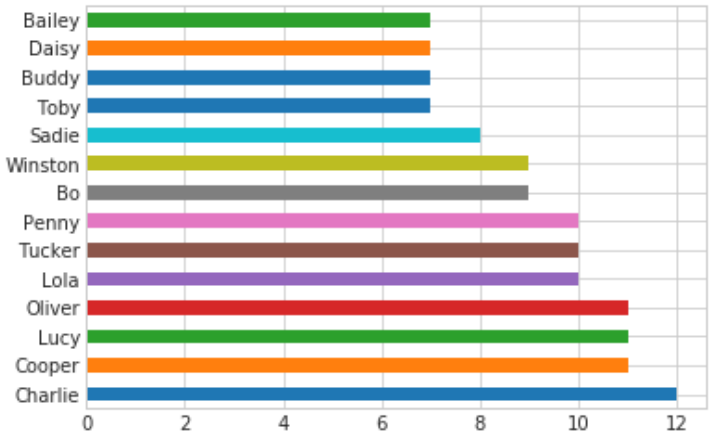
:om/dog_rates/status/667062181...	10	10	Keet	None	None	None	None

Second, the graph produced after this procurement had loads of rows as some names were repeated only once, you can imagine having like 70 names written vertically on the y-axis. I then made a lower limit of seven to cut the slack of the names and produce a visual that is clear, informative and appealing.

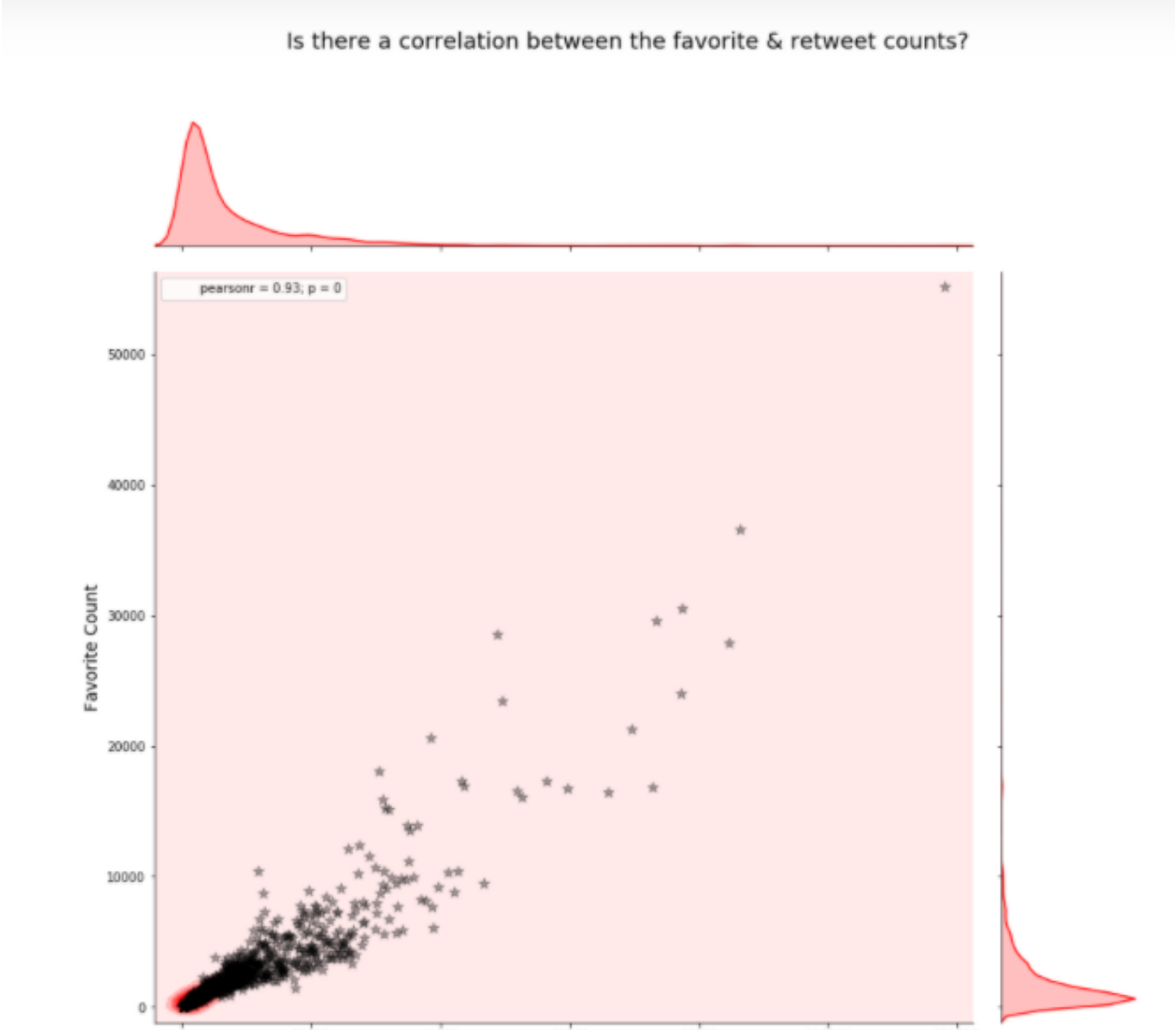
Making this code part and other parts definitely was not a piece of cake, I had to go through stack exchange - one of the most websites coders go to - pages, visit websites and check documentation made for Python programming language.

```
In [383]: #Displaying most common names of Dogs
          y = y.groupby('name').filter(lambda y: len(y) >= 7 )
          y['name'].value_counts().plot(kind = 'barh')
          plt.show

Out[383]: <function matplotlib.pyplot.show(*args, **kw)>
```



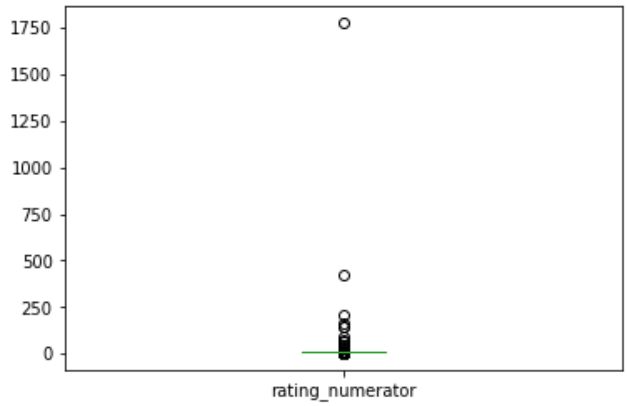
The plot showing most common dog names. The most common one is Charlie.



This is the second visualization I made to depict the relation between favorite and retweet count.

```
In [44]: #Code
         #Third visualization
         df.rating_numerator.plot.box()

Out[44]: <matplotlib.axes._subplots.AxesSubplot at 0x7f7857f806d8>
```




```
In [45]: # Saving files in the workspace
         folder_name = 'Final_Documents'
```

In the third visualization, I made a box plot of the rating numerator.

The final dataframes were saved into folders in the online workspace.

Afterwards, I checked the project rubric and to fulfil a task requirement, the three dataframes were combined into a single one after unifying the ID column's name and its datatype along the three dataframes.