act_report

October 2, 2022

0.1 Report: act_report

- Create a **250-word-minimum written report** called "act_report.pdf" or "act_report.html" that communicates the insights and displays the visualization(s) produced from your wrangled data. This is to be framed as an external document, like a blog post or magazine article, for example.

We have three steps in the data wrangling process. Namely; Data Gathering, Data Accessing and Data Cleaning. We have completed these 3 steps, now we have to draw insights from the clean data presented and also so some visualization

0.1.1 We made 3 insights and visualization on the clean data, regarding the ratings, the names and the breed of the dogs

0.1.2 Insights:

1. From the analysis we notice that 10,11,and 12 are the highest dog ratings
2. From the analysis we notice that while most dog names are missing('None'), some dogs still share the same names. The most common are Tuker, Oliver, Penny, cooper and charlie
3. From the analysis we notice that while most dog breeds were missing, however the most common dog breeds are golden_retriever and labrador_retriever

0.1.3 Visualization

```
In [2]: #Lets read the wrangled data so we can give insights and visualize
        import pandas as pd
        wrangled_data = pd.read_csv('twitter_archive_master.csv')
        wrangled_data.head()
```

```
Out[2]:              tweet_id            timestamp  \
        0  892420643555336193  2017-08-01 16:23:56
        1  892177421306343426  2017-08-01 00:17:27
        2  891815181378084864  2017-07-31 00:18:03
        3  891689557279858688  2017-07-30 15:58:51
        4  891327558926688256  2017-07-29 16:00:24

                                            source  \
```

```
0  http://twitter.com/download/iphone
1  http://twitter.com/download/iphone
2  http://twitter.com/download/iphone
3  http://twitter.com/download/iphone
4  http://twitter.com/download/iphone

                                                text  \
0  This is Phineas. He's a mystical boy. Only eve...
1  This is Tilly. She's just checking pup on you...
2  This is Archie. He is a rare Norwegian Pouncin...
3  This is Darla. She commenced a snooze mid meal...
4  This is Franklin. He would like you to stop ca...

                                        expanded_urls  rating_numerator  \
0  https://twitter.com/dog_rates/status/892420643...                13
1  https://twitter.com/dog_rates/status/892177421...                13
2  https://twitter.com/dog_rates/status/891815181...                12
3  https://twitter.com/dog_rates/status/891689557...                13
4  https://twitter.com/dog_rates/status/891327558...                12

   rating_denominator      name stage  retweet_count  favorite_count  \
0                  10   Phineas  None         6961.0         33635.0
1                  10     Tilly  None         5266.0         29179.0
2                  10    Archie  None         3463.0         21949.0
3                  10     Darla  None         7183.0         36743.0
4                  10  Franklin  None         7707.0         35101.0

       breed
0       None
1  Chihuahua
2  Chihuahua
3       None
4     basset
```

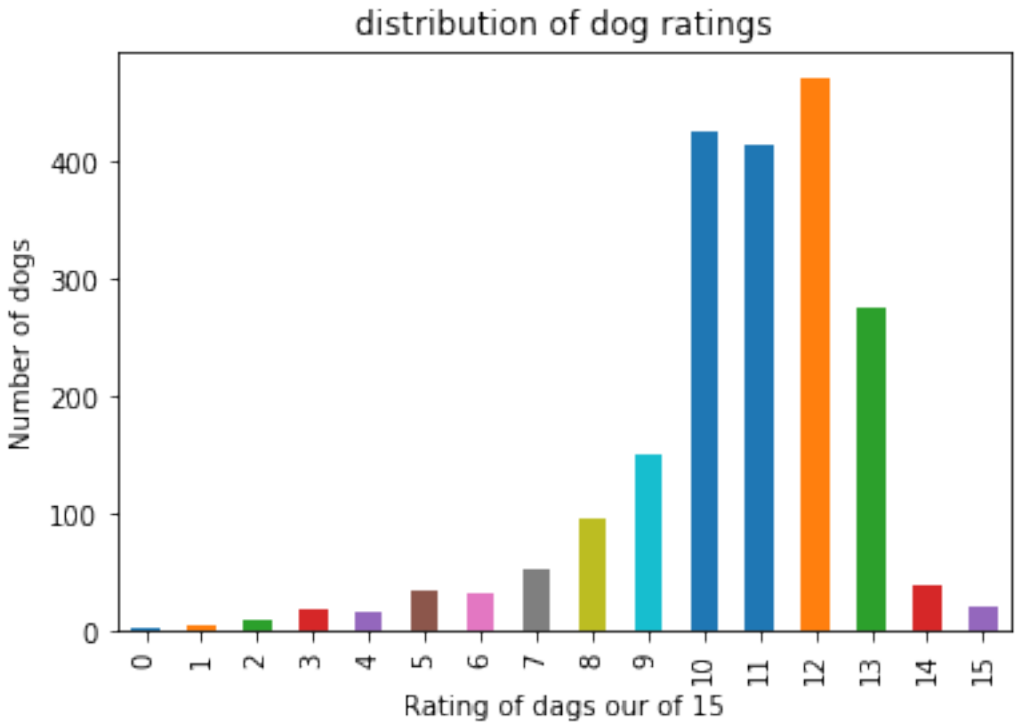
```
In [3]: #Lets see the highest breed of dogs
        import pandas as pd
        import matplotlib.pyplot as plt
        % matplotlib inline
```

```
In [7]: #insight 1, lets analyze the dog distribution
        wrangled_data['rating_numerator'].value_counts().sort_index()
```

```
Out[7]: 0       2
        1       5
        2       9
        3      19
        4      16
        5      34
```

```
6      32
7      52
8      95
9     151
10    425
11    413
12    471
13    275
14     38
15     20
Name: rating_numerator, dtype: int64
```

In [4]:
```
#Lets checkout the dog rating distribution visualization (from insight 1)
plot_chart = wrangled_data['rating_numerator'].value_counts().sort_index().plot(kind='ba
plot_chart.set_xlabel("Rating of dags our of 15");
plot_chart.set_ylabel("Number of dogs");
```

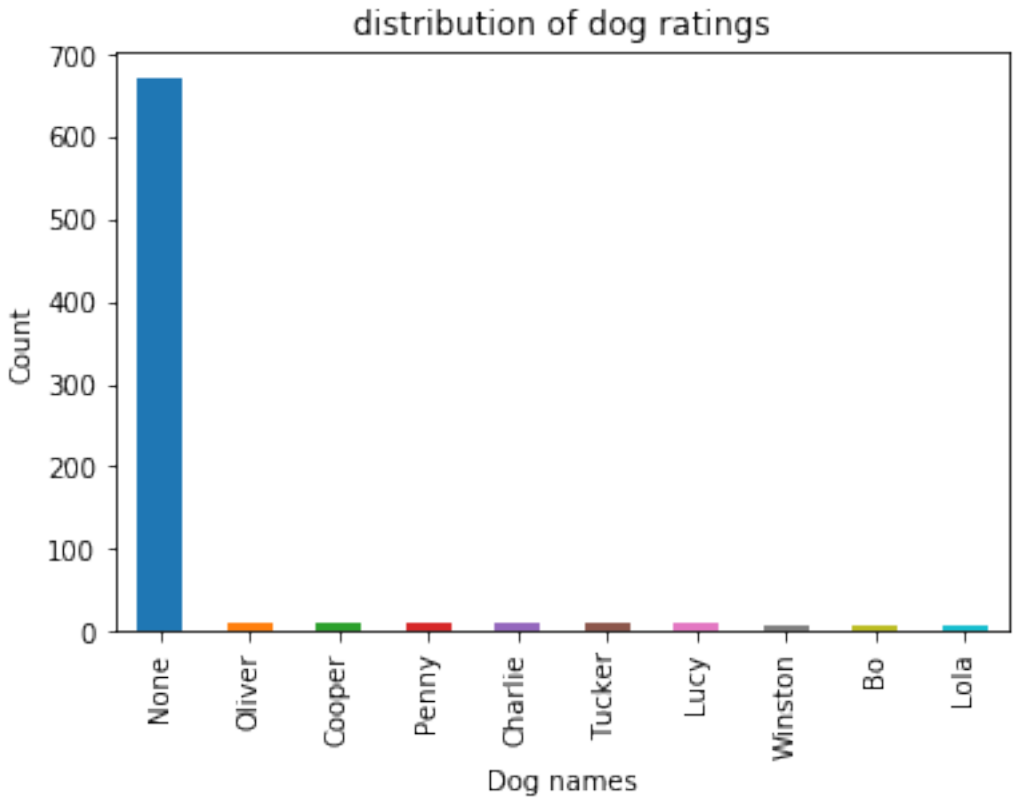


In [10]:
```
#insight 2, lets see if the dogs shared simliar names
wrangled_data['name'].value_counts()[:10]
```

Out[10]:
```
None       671
Oliver      10
Cooper      10
Penny       10
```

```
Charlie     10
Tucker      10
Lucy         9
Winston      8
Bo           8
Lola         8
Name: name, dtype: int64
```

In [5]:
```
#Lets checkout the dog names distribution from insight 2
plot_chart = wrangled_data['name'].value_counts()[:10].plot(kind='bar', title = 'distrib
plot_chart.set_xlabel("Dog names");
plot_chart.set_ylabel("Count");
```



In [9]:
```
#insight 3, lets see the most common dog breeds
wrangled_data['breed'].value_counts()
```

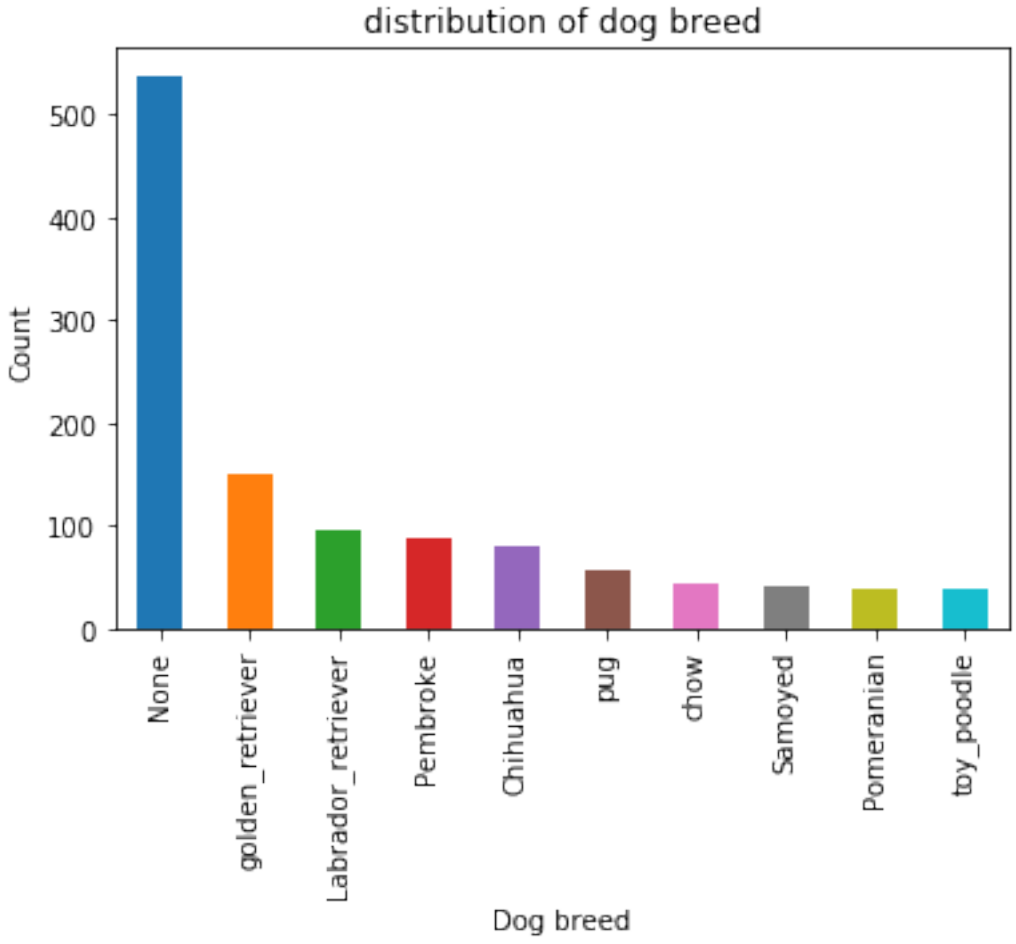
Out[9]:
```
None                    538
golden_retriever        150
Labrador_retriever       96
Pembroke                 88
Chihuahua                81
pug                      57
```

chow	44
Samoyed	42
Pomeranian	38
toy_poodle	38
cocker_spaniel	30
malamute	30
French_bulldog	25
Chesapeake_Bay_retriever	23
miniature_pinscher	23
Siberian_husky	20
German_shepherd	20
Cardigan	19
beagle	18
Eskimo_dog	18
Staffordshire_bullterrier	18
Shetland_sheepdog	18
Maltese_dog	18
Shih-Tzu	17
Lakeland_terrier	17
Rottweiler	17
kuvasz	16
Italian_greyhound	16
West_Highland_white_terrier	14
Great_Pyrenees	14
...	
Norwich_terrier	4
Rhodesian_ridgeback	4
Tibetan_terrier	4
miniature_schnauzer	4
Saluki	4
Ibizan_hound	3
giant_schnauzer	3
Greater_Swiss_Mountain_dog	3
komondor	3
cairn	3
Leonberg	3
Scottish_deerhound	3
briard	3
Brabancon_griffon	3
Welsh_springer_spaniel	3
Irish_water_spaniel	3
curly-coated_retriever	3
Australian_terrier	2
Sussex_spaniel	2
toy_terrier	2
Appenzeller	2
black-and-tan_coonhound	2
wire-haired_fox_terrier	2

```
clumber                          1
standard_schnauzer               1
Japanese_spaniel                 1
groenendael                      1
silky_terrier                    1
EntleBucher                      1
Scotch_terrier                   1
Name: breed, Length: 112, dtype: int64
```

In [6]:
```
#Lets checkout the dog breed distribution from insight 3

plot_chart = wrangled_data['breed'].value_counts()[:10].plot(kind='bar', title = 'distri
plot_chart.set_xlabel("Dog breed");
plot_chart.set_ylabel("Count");
```



In [ ]: