Report: act_report

Insights

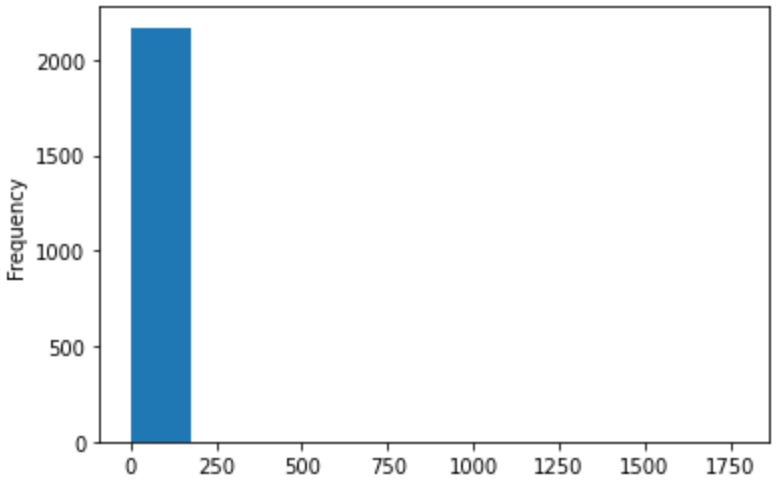
The following are the insight gotten from the wrangled data:

In [2]:
```
# import libraries
import pandas as pd
%matplotlib inline
import matplotlib.pyplot as plt
```

In [3]:
```
# impor files
df_combined = pd.read_csv('twitter_archive_master.csv')
```

Insight 1. Dog rating numerators are between 10 to 15.

In [4]:
```
df_combined['rating_numerator'].plot(kind='hist');
```



In [5]:
```
round(df_combined.rating_numerator.value_counts()/len(df_combined)*100)
```

Out[5]:

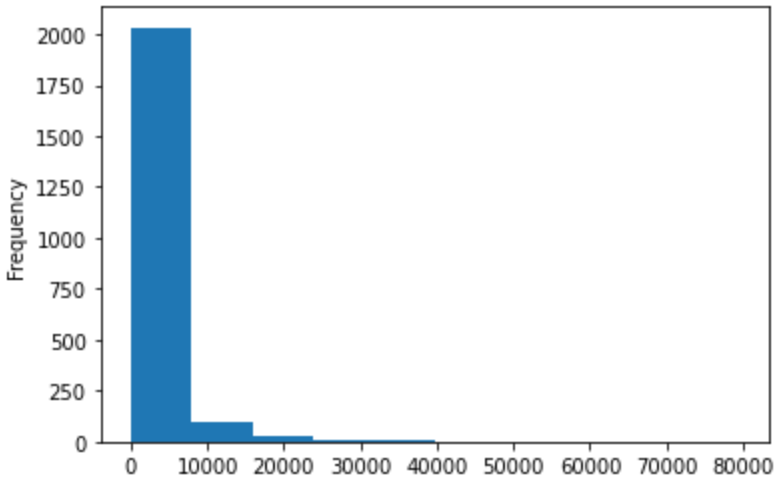
12	23.0
10	20.0
11	20.0
13	14.0
9	7.0
8	5.0
7	2.0
14	2.0
5	2.0
6	1.0
3	1.0
4	1.0
2	0.0
1	0.0
420	0.0
0	0.0
17	0.0
45	0.0
60	0.0
143	0.0
44	0.0
99	0.0

```
121      0.0
20       0.0
26       0.0
144      0.0
80       0.0
165      0.0
50       0.0
204      0.0
1776     0.0
27       0.0
75       0.0
24       0.0
84       0.0
960      0.0
15       0.0
182      0.0
666      0.0
88       0.0
Name: rating_numerator, dtype: float64
```

The above plot further shows that 38 percent of rating_numerator is 10, 25 percent are 12, 20 percent are 11, 15 percent are 13 and the remaining 2 percent are 14.

Insight 2. Most people provide just 1 image for prediction..

In [6]:
```
df_combined['img_num'].plot(kind='hist');
```



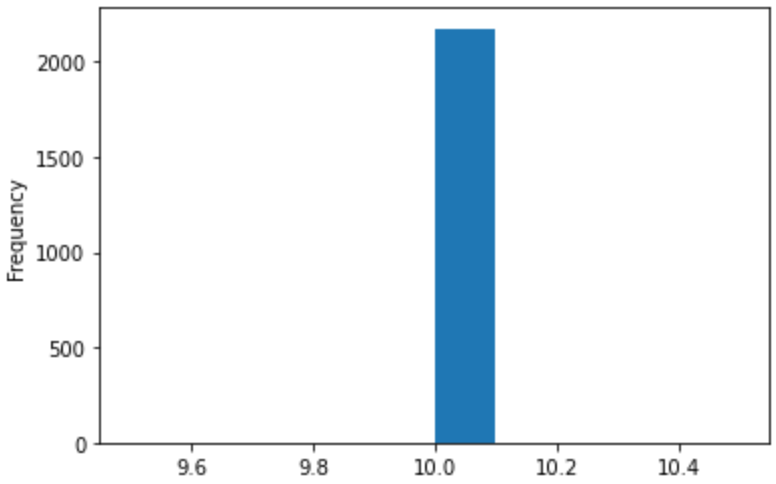
In [7]:
```
round(df_combined.img_num.value_counts()/len(df_combined)*100)
```

Out[7]:
```
577      0.0
83       0.0
516      0.0
3652     0.0
71       0.0
        ...
4492     0.0
6478     0.0
2035     0.0
5488     0.0
147      0.0
Name: img_num, Length: 1713, dtype: float64
```

The analysis shows that 73 percent provided 1 image for prediction, 8 percent provided 2 images, 3 percent provided 2 images, while only 1 percent provided 4 images for prediction.

Insight 3. Dog ratings denorminator is 10.

In [8]:
```
df_combined['rating_denominator'].plot(kind='hist');
```



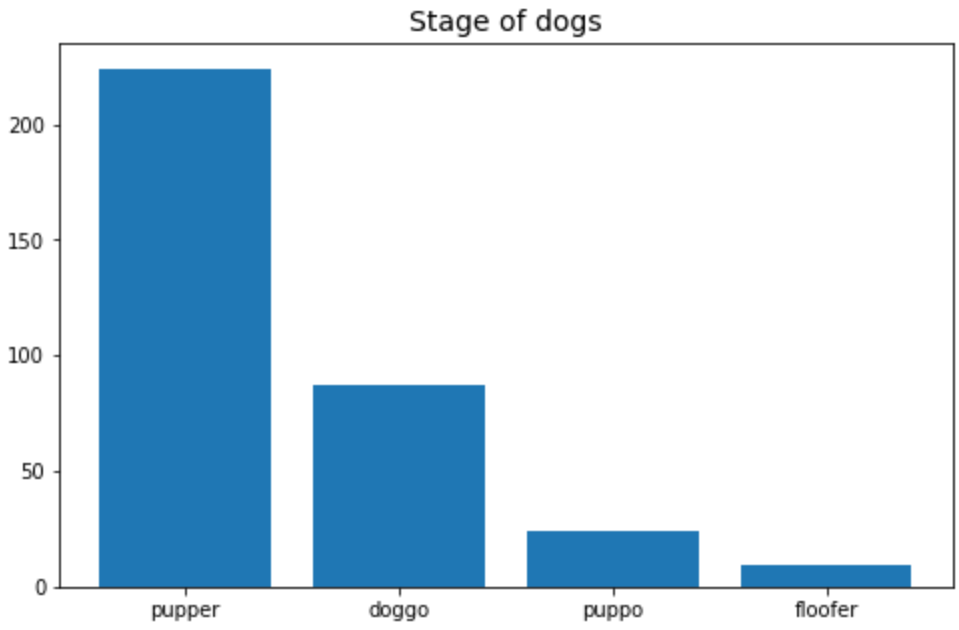
In [9]:
```
round(df_combined.rating_denominator.value_counts()/len(df_combined)*100)
```

Out[9]:
```
10    100.0
Name: rating_denominator, dtype: float64
```

This shows that all the rating denominators are '10' after cleanning

In [10]:
```
plt.figure(figsize=(8,5))
plt.title("Stage of dogs", fontsize=14)
plt.bar(x=df_combined['dog_stage'].value_counts().index,
        height=df_combined.dog_stage.value_counts().values)
```

Out[10]:
```
<BarContainer object of 4 artists>
```



The chart above shows that pupper dog_stage is the most common at weratedogs twitter archive.

In [ ]: