

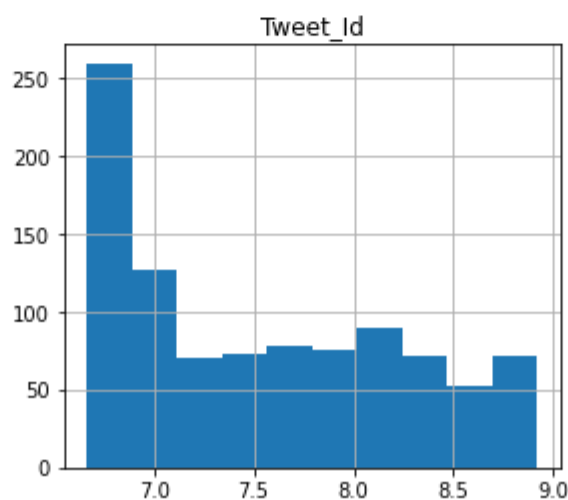
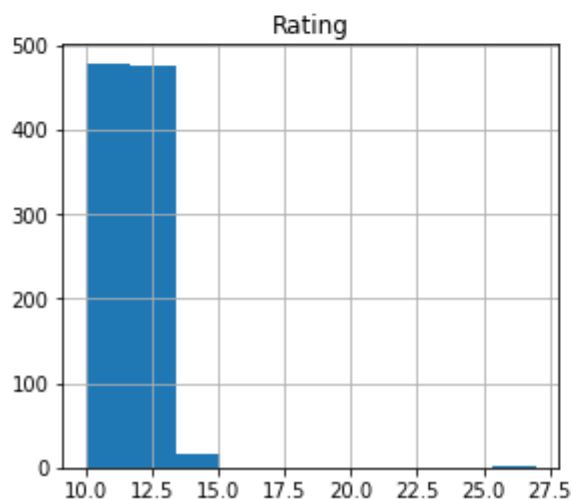
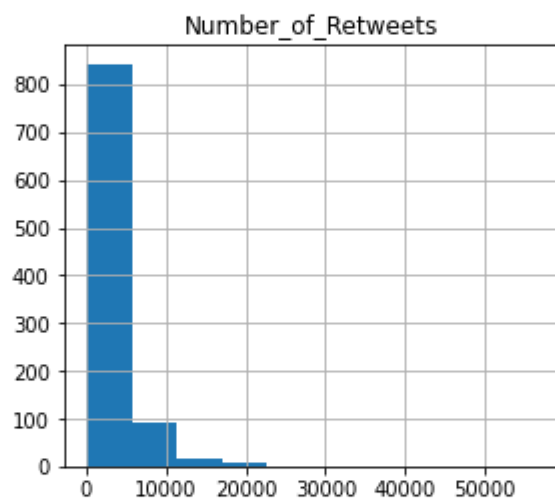
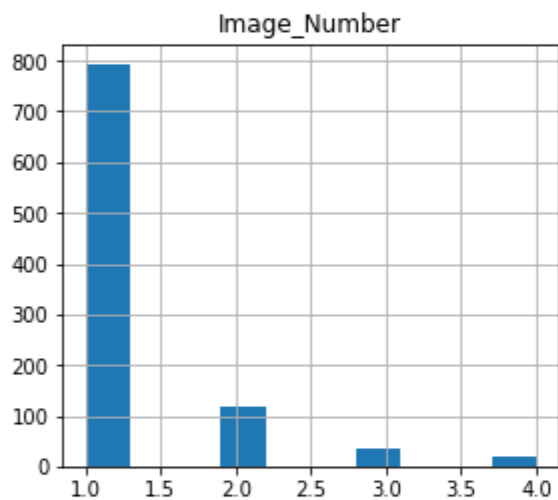
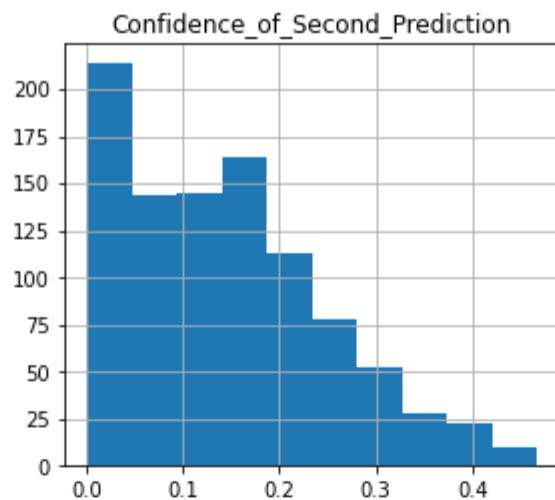
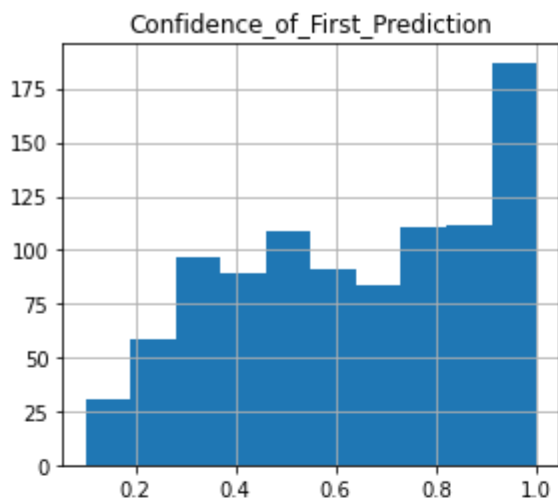
```
import pandas as pd  
df = pd.read_csv("twitter_archive_master.csv")
```

▼ Analyzing, and Visualizing Data

```
graph=df.select_dtypes(["int64","float64"])  
graph.hist(figsize=(16,14))
```



```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7fe7357b5fd0>,
      <matplotlib.axes._subplots.AxesSubplot object at 0x7fe7357952e8>,
      <matplotlib.axes._subplots.AxesSubplot object at 0x7fe735748550>],
      [<matplotlib.axes._subplots.AxesSubplot object at 0x7fe7356f87b8>,
      <matplotlib.axes._subplots.AxesSubplot object at 0x7fe735729a20>,
      <matplotlib.axes._subplots.AxesSubplot object at 0x7fe7356dcc88>],
      [<matplotlib.axes._subplots.AxesSubplot object at 0x7fe735693ef0>,
      <matplotlib.axes._subplots.AxesSubplot object at 0x7fe735651160>,
      <matplotlib.axes._subplots.AxesSubplot object at 0x7fe7356511d0>]],
      dtype=object)
```



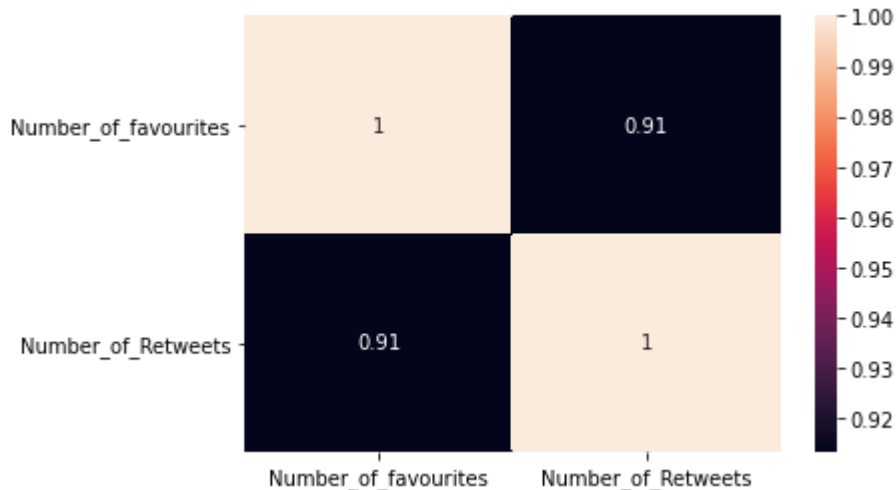
Inference:

- The confidence of First Prediction is the highest and has maximum number of values having a c
- Very few values in confidence of First Prediction have values which are on the lesser side, that is
- Values are almost equally distributed between 0.3 and 0.9 confidence in the confidence of First Pr
- Maximum number have a confidence of 0, that is 0% in the confidence of the second prediction
- In the confidence of the second prediction,, as the confidence keeps on increasing, the number c decreasing
- Least number of tweets have a maximum confidence of 0.45 that is 45% in the confidence of se
- The confidence of Third Prediction is the lowest as compared to the confidence of the first and i
- The confidence of the Third prediction has maximum confidence of 0.22 i.e. 22% with most num
- Very few values having a confidence of 0.22 in the confidence of the third prediction
- Most number of tweets have retweets between 0-1000
- Very less number of tweets have tweets more than 1000
- Maximum retweets received are 5000 retweets
- A very high number of tweets are favourited upto 1500 times
- Number of tweets favourited above 1500 goes on decreasing
- Maximum number of times a tweet is favoutited is 12000
- Most of the tweets have a rating of 10-12.5
- Very few tweets have a rating above 12.5
- The maximum rating received by a tweet on a scale of 10 is 27.5
- For most number of tweets, the image number that corresponded to the most confident predicti
- After image number 1, image number that corresponded to the most confident prediction is ima
- After image number 2, the image number that corresponded to the most confident prediction is
- The number of tweets for which image number 4 corresponds to the most confident prediction

```
import seaborn as sns
corr_new=df[['Number_of_favourites', 'Number_of_Retweets']].corr()
sns.heatmap(corr_new,annot=True)
```



```
/usr/local/lib/python3.6/dist-packages/statsmodels/tools/_testing.py:19: FutureWarning:
import pandas.util.testing as tm
<matplotlib.axes._subplots.AxesSubplot at 0x7fe72a51d240>
```



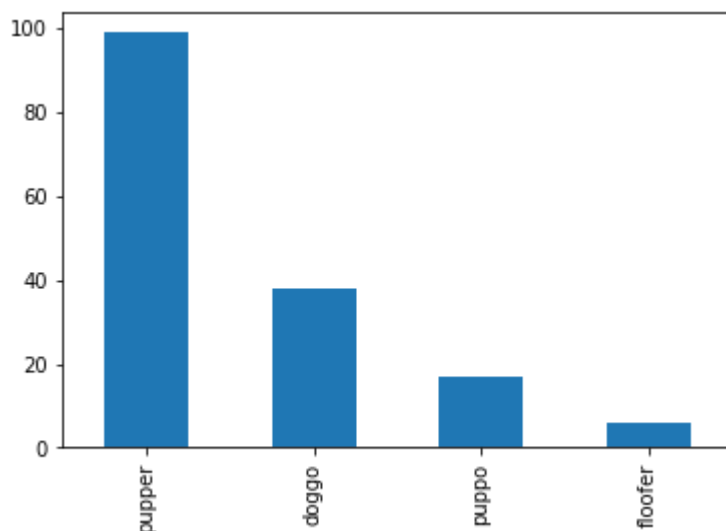
Inference:

- The number of retweets and the number of favourites have a high correlation
- When number of retweets is high, the number of favourites is high
- When number of retweets is less, number of favourites is less as well

```
new_plot = df[df.Life_Cycle.str.len()<8]
```

```
new_plot.Life_Cycle.value_counts().plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fe726a21748>
```



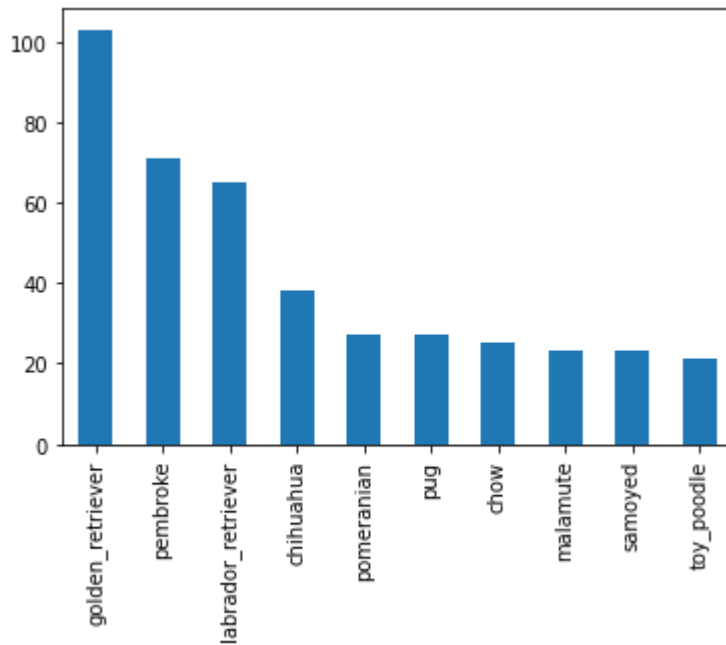
Inference:

- It shows the lifecycle of the dogs

- Most of the dogs have a lifecycle stage of being a of pupper
- Few ddogs are in the lifecycle stage of being a of doggo
- Very few dogs are in the lifecycle stage of being a puppo
- Lease number of dogs are in the lifecycle stage of being a floffer

```
df['First_Prediction'].value_counts().sort_values(ascending=False).head(10).plot(kind='bar')
```

↳ <matplotlib.axes._subplots.AxesSubplot at 0x7fe726995c88>



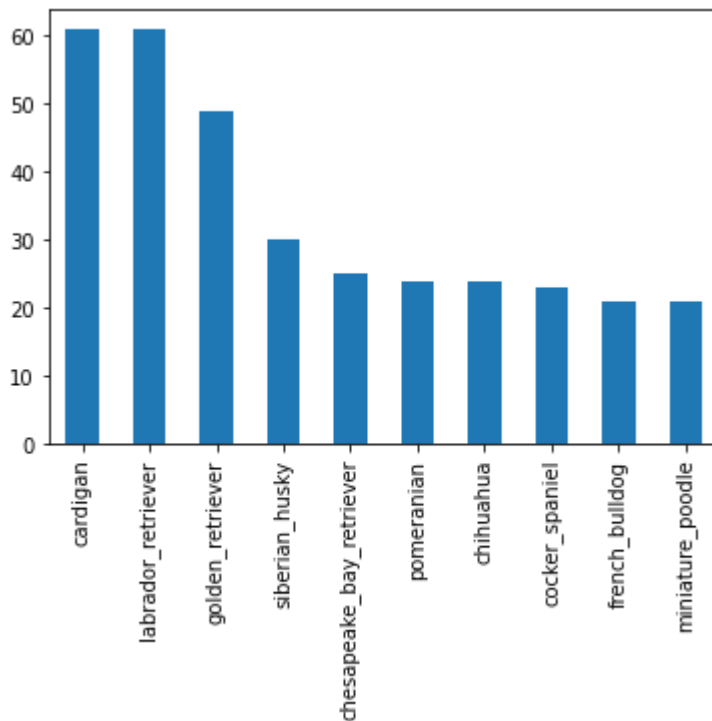
Inference:

- It shows the top 10 breeds that are most likely to be predicted in the First Prediction

```
df['Second_Prediction'].value_counts().sort_values(ascending=False).head(10).plot(kind='bar')
```

↳

<matplotlib.axes._subplots.AxesSubplot at 0x7fe7268f0550>

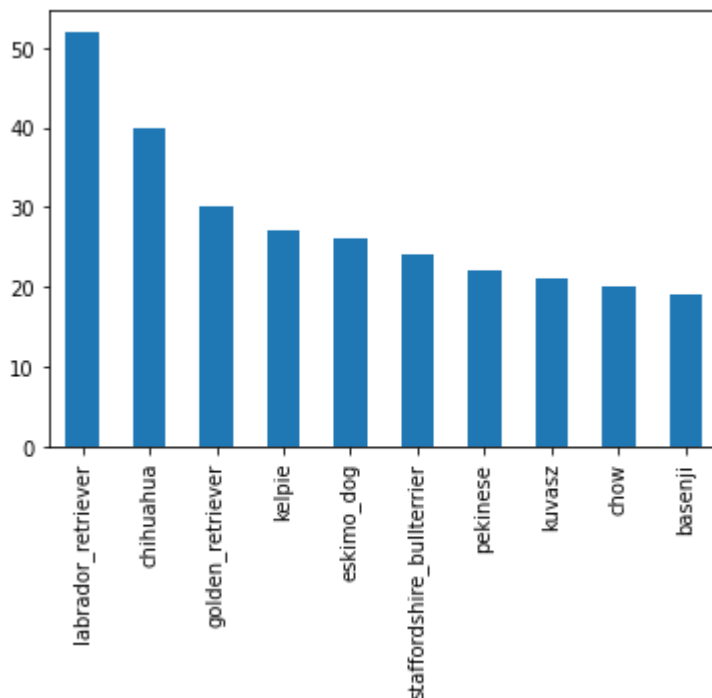


Inference:

- It shows the top 10 breeds that are most likely to be predicted in the Second Prediction

```
df['Third_Prediction'].value_counts().sort_values(ascending=False).head(10).plot(kind='bar')
```

↳ <matplotlib.axes._subplots.AxesSubplot at 0x7fe7268d5748>



Inference:

- It shows the top 10 breeds that are most likely to be predicted in the Third Prediction