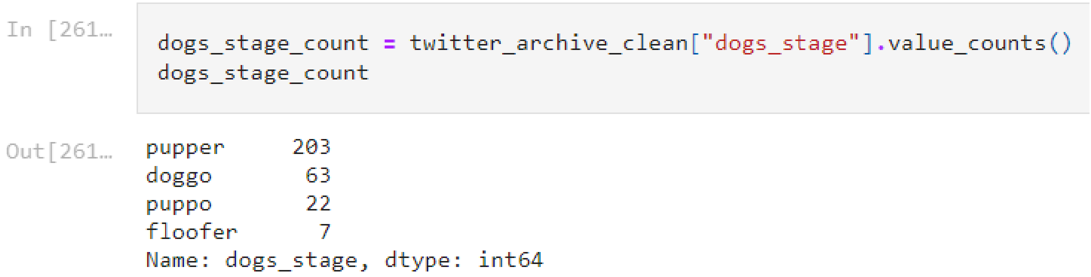
Data Analysis and Visualization Report

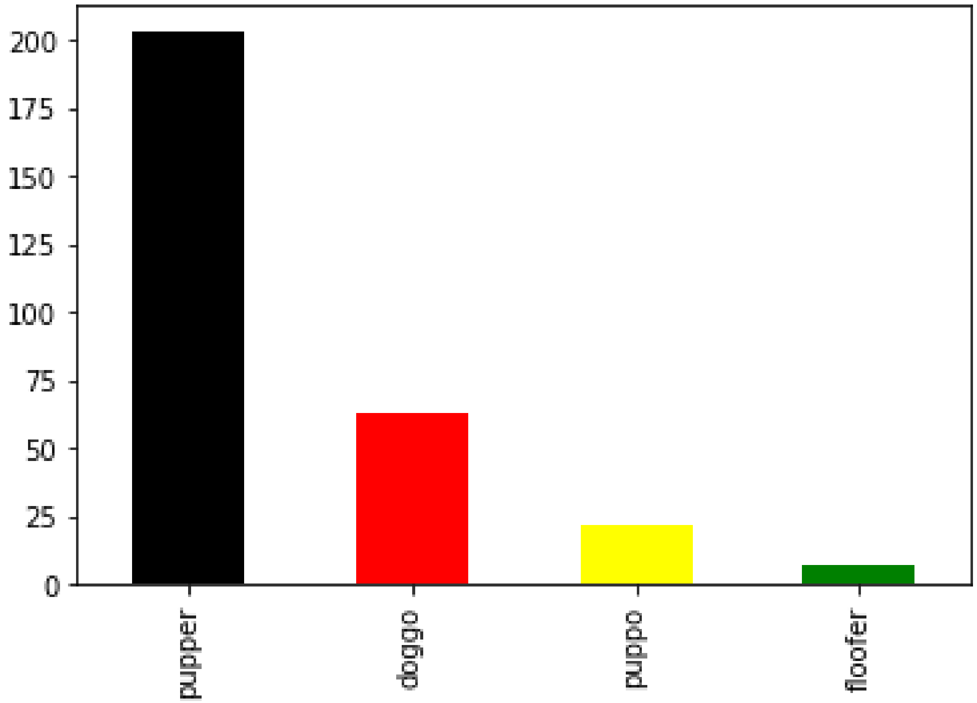
Analysis can be done in various ways. I used the following methods for analysis and plotting Visualizations to answer my questions.

1- What dog’s stage more common in tweets?



```
In [261…
dogs_stage_count = twitter_archive_clean["dogs_stage"].value_counts()
dogs_stage_count
```

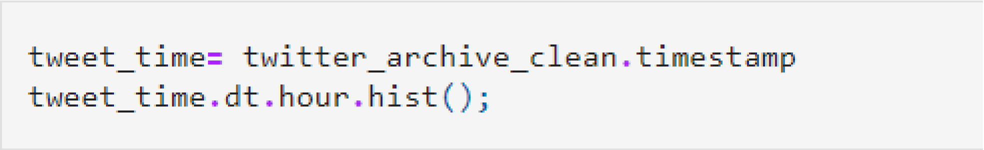
```
Out[261…
pupper     203
doggo       63
puppo       22
floofer      7
Name: dogs_stage, dtype: int64
```



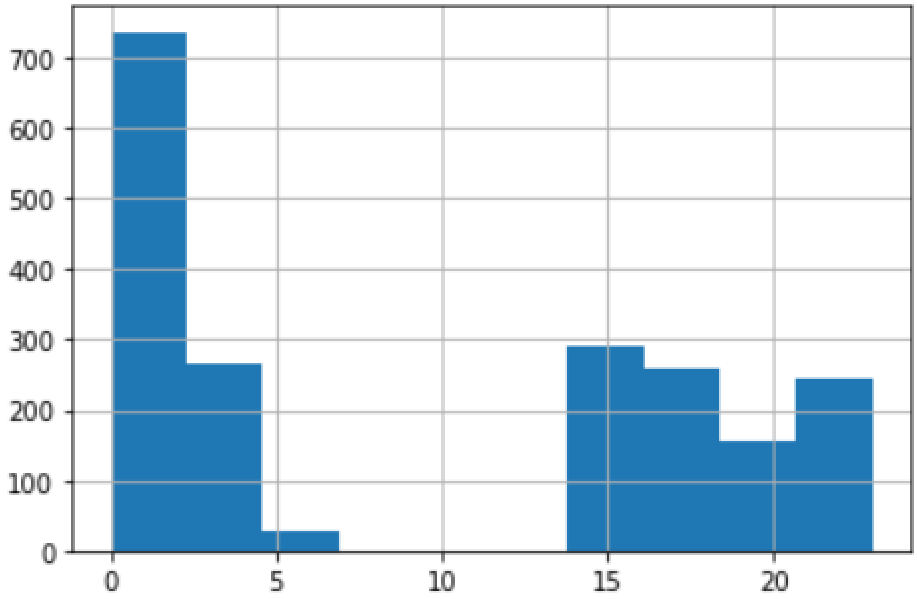
From above analysis and visualization, we have the following insights:

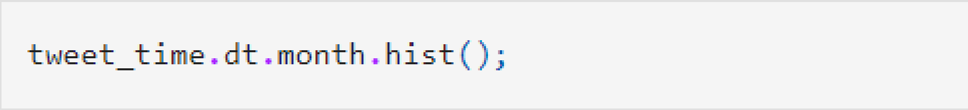
1- "pupper" is the most common.
2- "floofer" is the least common.

2- What is the most active time on the page?

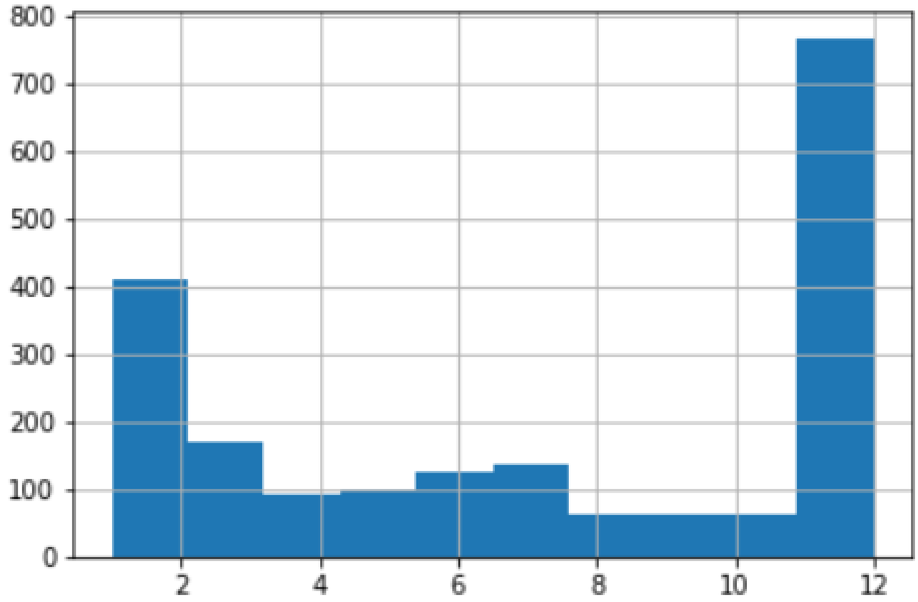


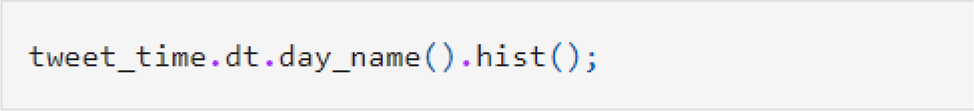
```
tweet_time= twitter_archive_clean.timestamp
tweet_time.dt.hour.hist();
```



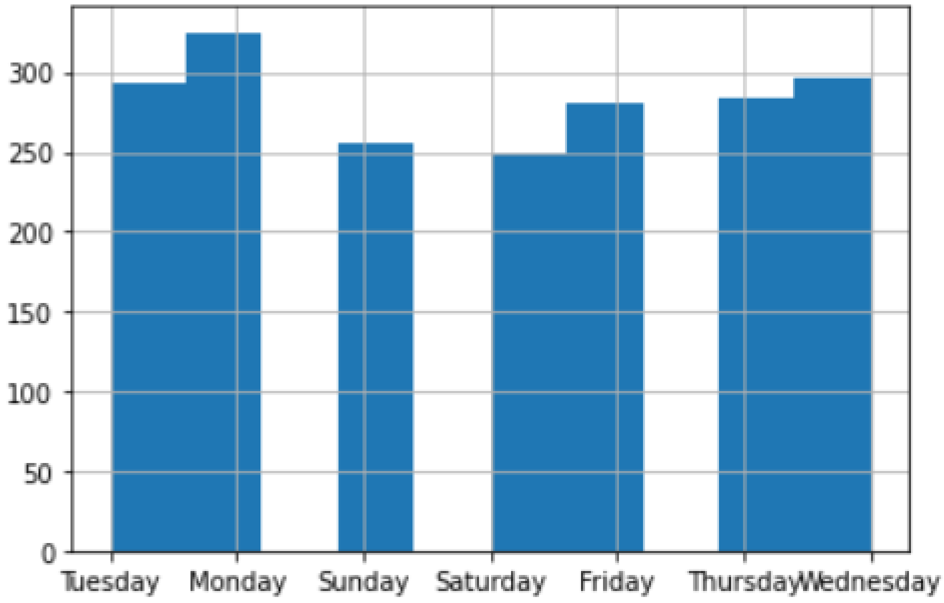


```
tweet_time.dt.month.hist();
```





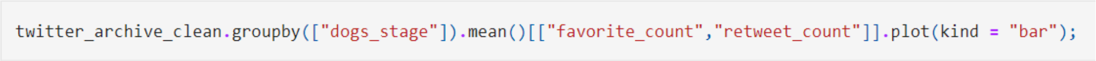
```
tweet_time.dt.day_name().hist();
```



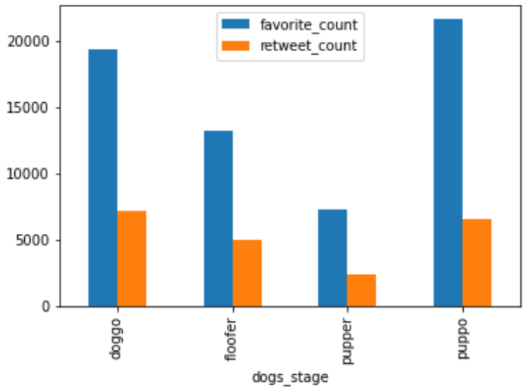
From above analysis and visualization, we have the following insights:

1) About [12 am : 2am] is the most time that page is active.
2) About [7 am : 2pm] has the lowest activity.
3) Monday has the highest activity.
4) Saturday has the lowest activity.
5) The last two month of the year has the highest activity on the page.

3- What is the most dog’s stage that gets highest favorite counts and retweet counts?



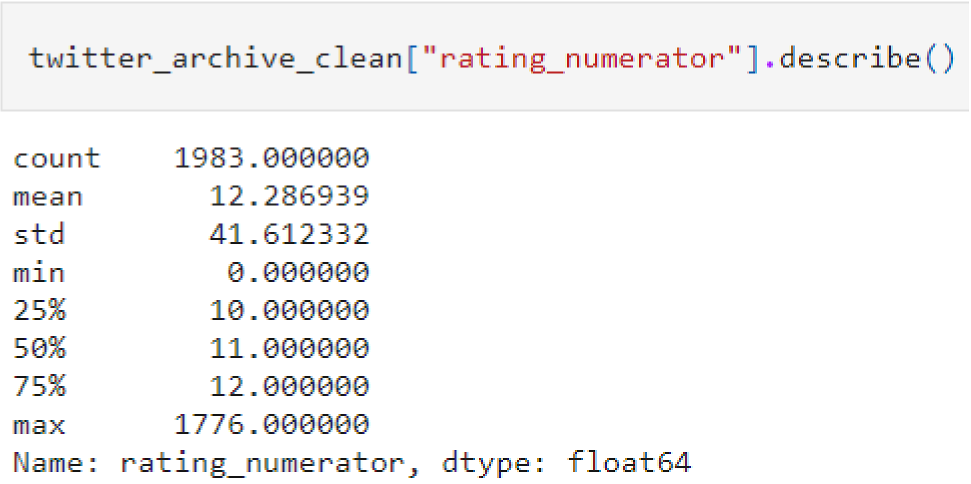
```
twitter_archive_clean.groupby(["dogs_stage"]).mean()[["favorite_count","retweet_count"]].plot(kind = "bar");
```



From above analysis and visualization, we have the following insights:

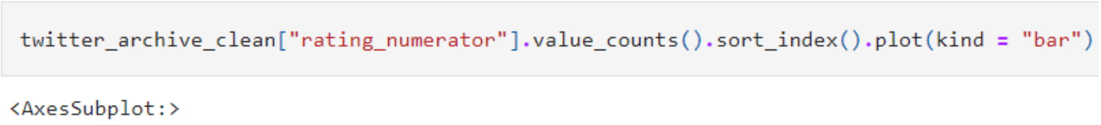
1) "puppo"and "doggo" pictures are most liked and have more retweets.
2) “pupper” pictures are the lowest liked and have more retweets.

4- What is mean rating of in tweets?



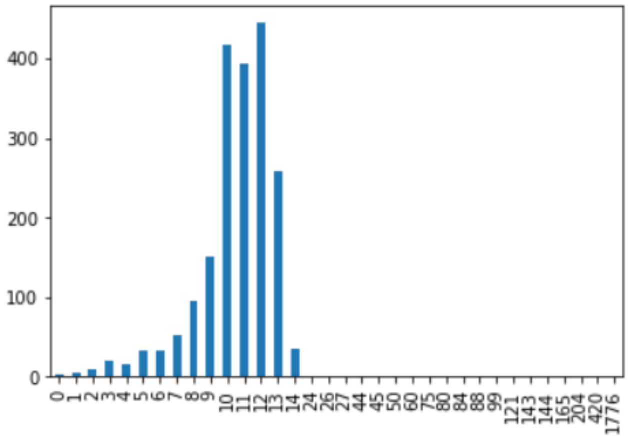
```
twitter_archive_clean["rating_numerator"].describe()
```

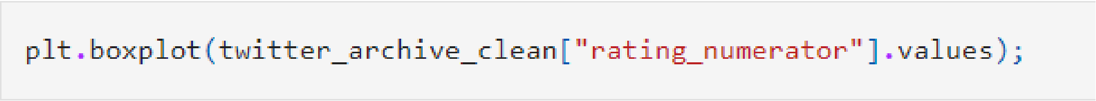
```
count    1983.000000
mean       12.286939
std        41.612332
min         0.000000
25%        10.000000
50%        11.000000
75%        12.000000
max      1776.000000
Name: rating_numerator, dtype: float64
```



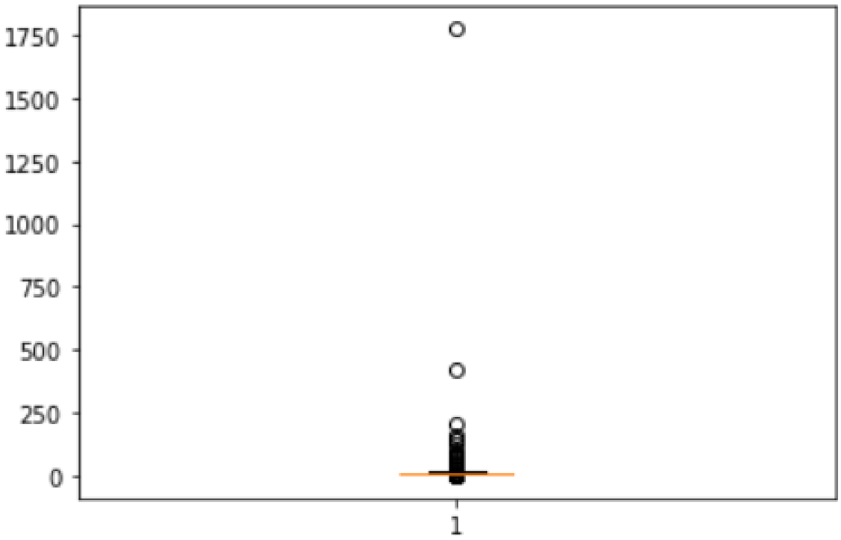
```
twitter_archive_clean["rating_numerator"].value_counts().sort_index().plot(kind = "bar")
```

```
<AxesSubplot:>
```





```
plt.boxplot(twitter_archive_clean["rating_numerator"].values);
```



From above analysis and visualization, we have the following insights:

1- **Average rating is 12.28.**
2- **"rating_numerator" has some outliers(e.g. max value = 1776).**