Gathering Data

1. Twitter Archive File(df_1) : I downloaded this file on Udacity, which is basically including tweet ID, timestamps, source, etc
2. Image Prediction File(df_2) : They provide what breed of dog according to a prediction algorithm. This file had to be downloaded using the url. You need to use the 'requests' library that is usually used to download a file from the web using the file url.
3. Twitter API(df_3) : You should use 'Tweepy', which is an open source python package. It allows you to access the Twitter API. By using this convenient package, I queired the Twitter API for each JSON data.

Assessing Data : In this process, I looked through all the three tables and found out quality and tidiness issues that I should handle.

- Quality issue
    1. column timestamp datatype
    2. weird name
    3. unnecessary column
    4. some variable having 0 value
    5. typo in column name
    6. None value in dog stage
    7. tweet id datatype
    8. column p1, p2, p3 have upper/lowercase letter at the same time
- Tidiness issue
    1. Merge table
    2. 4 types dog stage

Cleaning Data : According to assessing process, I fix the issue

- Quality issue
    1. column timestamp datatype : convert data type into datetime
    2. weird name : Searched the weird name(a, the, his, etc) and remove
    3. unnecessary column : We don't need to use all the columns, so I dropped some of the columns that I'm not going to use to analyze the data
    4. some variable having 0 value : remove the variable having 0 value
    5. typo in column name : Based on 'dogtionary' we should edit the typo in the dog stage
    6. None value in dog stage : I changed the None value into the empty space

7. tweet id datatype : This column should be converted into string type. Make sure to change data type throughout all the three tables. Because it should be merged into one table at the end.
8. column p1, p2, p3 have upper/lowercase letters at the same time : It's a little bit messy writing upper or lower case letters. So I set the all the letters to the lower case

- Tidiness issue
    1. Merge table : All three tables should be merged into the one table that allows you to analyze in an easy way. I used 'tweet_id' to merge the table.
    2. 4 types of dog stage : We don't need to keep the dog stage separately. I grouped this into 'dog_stage' column

Analyzing/Visualizing Data

- What's the most popular dog breeds? : Bar plot
- What's the correlation between Retweet Count, Favorite Count, Ratings : Correlation heatmap
- What's the relationship between favorite counts and retweet counts : Scatter plot
- Which dog stage appeared the most? Bar plot
- Where did you get this dog rate? def function