

While gathering the data from the twitter API I was stuck for several days on different problems but mainly on an error returning:

```
Object of type Status is not JSON serializable
```

After investigating for a while and trying many different angles, I finally discovered the following block post in udacity's knowledge section.

<https://knowledge.udacity.com/questions/2986>

Adding 'parser=tweepy.parsers.JSONParser()' got rid of the problem. Also switching wait_on_rate_limit on fixed the problem of the twitter connection closing because of rate limitation which I first tried to fix by running the 'for id in twitter_archive['tweet_id']' several times with the missing_tweet_ids.

After adding the code ('wait_on_rate_limit=True') the loop needed to run ~45 minutes to finish but I was getting the results I was looking for.

Because I lost so much time on this gathering part I had limited time for the rest of the project so I decided to keep the cleaning and analyzing to a minimum.

For cleaning I focused on some of the most obvious issues but there are still a few issues that should be addressed including other names that could be wrong inside twitter_archive. I only fixed the ones that I casually spotted.

In addition some work could be done with the image_predictions tries in different columns. Would be nice to only store the ones where the result is a dog breed and has the highest confidence level. As a consequence of the missed wrangling efforts on this part I also excluded it from any analysis in the final part of the project.

I also have to acknowledge that some of the code especially in the cleaning part is not very "beautiful" and according to what are considered "good" coding standards. But as my time was limited and it got the job done I didn't bother to invest time in cleaning up my code.

Overall it was fun to work in python again after working on R for the past Udacity project.