We Rate Dogs

Wrangling Documentation by Ifedayo Ayilaran

The We Rate Dogs wrangling project is the second project of the Udacity DAND. It is certainly a project to remember. In theory, the tasks were easy:

- Gather data from different sources
- Assess the data
- Clean the data
- Analyze the data

These four steps are supposed to showcase skills gained from the Data Wrangling learnings in the program.

In the gather phase, data is gathered from three different sources. The first source is a csv file of tweets and tweet ids about dogs from the We Rate Dogs twitter account. Next step was to use the *requests* library to download a TSV file about image predictions of the dog species.

The last source of data was to gather retweet and like counts for the tweet ids in the provided file. For this, I had to apply for a Twitter Developer account to access the twitter API using the Tweepy library.

I had issues with hiding my API keys since such information is sensitive and should be hidden from the code. I finally made use of environment variables to do this. I learnt new things like *dotenv*, '*.gitignore*,' and a lot of what looked like gibberish at first.

The next step is to connect to the API using the keys and to retrieve the tweets corresponding to the tweet ids by using a for loop. After retrieving (which goes on for a while), I stored the tweets in a '.txt' file using the *with* method and *json.dumps*.

Using the pandas '*.Dataframe'* method, I converted the file into a data frame ready for use in the notebook.

That's the data gathering phase!

For the assessing and cleaning, I used numerous pandas methods to visually and programmatically assess the data and document quality issues as well as tidiness issues. I then cleaned the data using the issues as a guide.

The define-code-test method made the cleaning process a lot easier. I stored all the data into a master data frame.

After storing, I made use of the matplotlib and seaborn libraries to produce a few visualizations about the data. Since the project was mainly about data wrangling, I did not have to do much for the EDA.

The project was tasking and equally rewarding. I look forward to learning and practicing in further projects.