PYTHON AND R LAB PROJECT

ANIME DATABASE

During the first part of our Python and R lab course we have been asked to collect data from an API and create a data frame and to convert it to a csv file so that our fellow students could use it for their own purposes in the second part of the course. Out of all the possibilities we had we selected the MyAnimeList API since we really like anime and it gave us the impression of being a lot interesting.

After importing the necessary libraries for retrieving data from the Jikan.moe website and manipulating the json file we also imported the pandas library for creating the data frame and converting the json file to a csv file. In order to do so we initialized a list l2 and created an empty data frame with pandas, subsequently we created a list of values ranging from 0 to 3000 and we performed a for loop in that range for retrieving all the animes info one by one according to their IDs and append them to the empty list l2. Unfortunately not all of them had all the IDs in the range of our list so in order to handle the value error we came up with a smart trick consisting in an if condition. If the response of the requests is = ‘404’ the code keeps executing and passes to the next value of the list completely ignoring the bad response of the API. We also included a line of code of code from the time library which is time.sleep() whose purpose is that of applying a delay that,according to the documentation of the API of Myanimelist ,should be of 4 seconds. Later we defined a function called create\_dataframe that iterates through the elements of the list l2 (which basically are all json files) and retrieves that data needed appending them to the empty data frame. Since we got many KeyErrors due to the fact that there are many missing values we decided to handle those errors with try and except assigning to every KeyError a Null value called None. However the only attributes which resulted in null values where aired\_to (see the table below) and Episodes (# of episodes) since many anime have not finished airing and some others are movies as can be seen by the type attribute in the data frame. We also got many instances without the anime\_name attribute and so, in order to help our colleagues ,we decided to clean the dataset up and implement a function clean() that drops all the instances with null values in that column and drops also all the duplicates. Finally we converted everything in a csv called ‘anime project.csv’ that we uploaded on Github.

|  |  |
| --- | --- |
| FEATURES | DESCRIPTION |
| anime\_name | Name of the anime |
| anime\_id | Id on the Myanimelist website |
| type | Typology of the anime (TV,Special, movie,OAV,…) |
| episodes | Number of episodes  (1 for OAVs and Movies) |
| score | Score assigned by the Myanimelist website |
| rank | Position in the Myanimelist ranking |
| popularity | How much popular is anime among the users (Numbers represent the position in a ranking of popularity) |
| scored\_by | Number of users who voted for that anime |
| likes | Number of likes from the users |
| aired\_from | The date in which the anime started airing |
| aired\_to | The date in which the anime finished airing |
| duration | Mean of the duration of an episode |
| Storyline | Plot of the anime |