

Python and R for Data Science.

For the first part of the assignment we have been asked to create a 1000 rows database using an API. We chose the OMDB API. To create the database (output_db.json and output_db.csv), we used the following functions:

Functions

- `generate_1000_ids()`: this function takes no positional arguments and returns a list of 1000 film ids. Thanks to the OMDB documentation we knew that the ids are composed as follows: every id starts with 'tt' and followed by 7 digits (e.g. the first film as the id tt0000001). To create those ids we first populated an array containing only the digital part, `suffix_list`. The first element is 0000001 and the last one is 0001000. Then we populated a second array, `ids_list` containing the prefix (i.e. tt) followed by the suffix created before.
- `populate_dataframe()`: this function takes one positional argument: a list of ids and returns a pandas dataframe. First, we initialize an empty pandas dataframe called `df`. Second, we loop through the list of ids and for each id we send a request to the server (the OMDB API). We store the response in a variable `response` and transform it into a json object with the `.json()` function. Then we add the response into our dataframe. Finally, we return this dataframe.
- `create_database()`: this function takes no positional arguments and output a *json file*. First, thanks to the two previous functions, we create a `data` variable containing our 1000 rows dataframe. Then, we export it as a json file. Finally, we print 'Completed!' to show the user that everything succeeded.

We wrapped the `create_database()` function in a `if __name__ == "__main__":` in order to be able to export each individual function, if needed, without creating a database everytime.

Variable

For this project we needed to instantiate 3 variables. Variables name are written with capital letter to respect pep 8 coding convention (constant variables must be name with capital letters).

- `API_KEY`: this is the personal API key received from the OMDB website.
- `API_ADDRESS`: this variable store the url needed to connect to the API endpoint.
- `INITIAL_IMDB_ID_PREFIX`: this variable store ids' prefix (i.e. 'tt').