ETL Project Notes

Toolbox for google to get data

Can attach a separate mysql file into github to turn in report for class

Looking at a JSON file- see what it is, see square brackets so in python is a list, then comes the curly brackets which makes it a dictionary, inside dictionary we have IDs

In terminal: history | tail or head

Make a duplicate copy of the original dataframe that came with the first step, when we are making a copy we only wanted the first 3 olumns

Why do we need the copy function? .copy()

* If you don’t give it a copy, PANDAS does not make a cop-y
* The original data will not be intact
* It may well be bc in pandas and python we don’t declare variables in the beginning so there is no memory allocation for that so the .copy command might be necessary
* If you don’t do the copy function then your before table and new table will both be pointing to the same memory location

Use PANDAS to read the JSON format

Now have to run SQL bc PANDAS is not capabale of doing the primary key- so the next part of the code has to be an SQL script

* Prepare it in the background
* Then continue with jupyter notebook
* Between clean and local database, use MYSQL