***Databases and Big data project: Instructions for our project:***

ADVISED METHOD (IT IS PREFERABLE TO USE THIS METHOD)

First, open the folder named NBA\_project in your IDE, then find the file named main.py which is located in the following path: database\_project\_final\Project\NBA\_project\NBA\_project. To run our code, it is enough to run the python script (main.py) provided only AFTER having inputted your MySQL password 2 times in the required parts of the python code (line 31 + line 41). The first one is to connect your SQL to python, while the second one is to connect the database and to be able to populate it.

The python file you are running contains the whole process from:

* Creating a data frame for each csv file (each represents one of our entities: total of 6)
* Connecting SQL to Python (MySQL Workbench)
* Creating the tables based on the ER diagram (logical design from ER diagram) and adding their respective attributes (taking care of constraints)
* Inserting values from the csv files into our tables (that we created)
* Inputting our queries (in SQL format)

As a result, you will see the output of our queries (in the python terminal) after having ran the file.

ALTERNATIVE NUMBER 1:

Alternatively, even though the code in our python (for creating database, tables, filling the tables) is written in SQL format already, we thought that it would be better, following the guidelines you provided in the project description (where you asked for a separate python file and a separate SQL code) to separate the queries so that you can run them directly from MySQL Workbench. For this, we included a file named “Queries.sql”.  
 However, to be able to run this file successfully, you must first run the python file, commenting the part from line 256 onwards. This way, the python file will take care of everything except the queries. Only after running the python script can you proceed to running the SQL script we provided (IN YOUR LOCAL INSTANCE).

ALTERNATIVE NUMBER 2:

We decided to take it one step further, and not only implement the queries in SQL, but also the whole process except for the part where we insert the values from the csv to populate the tables (you can find this SQL file named “SQLCODE.sql” in our project folder. But before running this script, you should comment out the following in the python (main.py) file:

* Line 36
* Lines 53 to 148

After doing that, follow these steps:

* First, in the aforementioned SQL file, select rows 1 to 70 (basically everything except the queries) and run them.
* Second, run the python file (commenting out parts mentioned above) to populate the tables with values from csv files
* Third, run the rest of the SQL file, so basically the queries (lines 72-149) AFTER HAVING REFRESHED THE SQL FILE
* Note: If you want queries to appear only in SQL and not in your python terminal, you can also comment out lines 256 onwards.

General Note: If you want to test all our methods, be sure to drop schema (in MySQL workbench) before testing next method.

UNRELATED

Additional general side note:

* In our project folder, we provided 2 ER diagram: a full one where all attributes are present, and a simplified one with essential attributes that allows better flow and readability.
* To see our csv files, you must follow this path: database\_project\_final\Project\NBA\_project\NBA\_project\NBA\_data