TakharKuljit

February 24, 2025

```
<h2 style ="text-align:center; padding-top:5px;"> CS 101 - Foundation of Data Science and Engi
<p style="text-align:center;padding:5px; fontt-size:14px"><b> PSET-3 - Managing Data Excercise
```

0.0.1 This is an individual assignment. No collaboration is allowed.

0.0.2 Assignment Goal: Using Python connect to MySQL, create database, tables, and Load data from csv file.

Start by reviewing the provided file nj_teachers_salaries_pset3.csv. Examine the column names, data types of this data file. After reviewing this file please provide your solutions for the questions below.

Note: For this Homework you are not required to do any data cleaning. In your next Homework you will be required to clean the data before storing it into database table. You are also not required to use pandas.

0.1 Question 1 (5 pts)

0.1.1 Connect to your MySql database using your username and password. Name the cursor returned from the mysql connection object as mycursor.

[55]:
```
#import libraries
import mysql.connector as sq

mydb = sq.connect(
    host="localhost",
    user="cs101",
    password="dataisfun",
    allow_local_infile=True
)

# create cursor
mycursor = mydb.cursor()

print("Connected to MySQL database successfully!")
```

```
Connected to MySQL database successfully!
```

0.2 For Question 2-6, If you have tried running your sql statements in MySQL workbench, please make sure that the database has been dropped before executing your code here.

0.3 Question 2 (5 pts)

0.3.1 Create a new database called nj_state_teachers_salaries.

Be sure to verify that an empty database was created in MySQL before moving to next step.

[57]:
```
# create database
mycursor.execute("CREATE DATABASE IF NOT EXISTS nj_state_teachers_salaries;")
```

0.4 Question 3 (20 pts)

Create a table called teachers_salaries. (Tables are created within a database. In this step we are asking you to create your table within the database you created in Question-2).

Since data cleaning is not required, you may use VARCHAR or TEXT as your datatype while creating your table. In your next assignment you will be required to use appropriate data type for each column

Be sure to verify that the table was created in MySQL before moving to next step.

[59]:
```
# define database and table name
db_name = "nj_state_teachers_salaries"
table_name = "teachers_salaries"

# select database
mycursor.execute(f"USE {db_name};")

# create table
create_table_query = f"""
CREATE TABLE IF NOT EXISTS {table_name} (
    id INT AUTO_INCREMENT PRIMARY KEY,
    last_name VARCHAR(255),
    first_name VARCHAR(255),
    county VARCHAR(255),
    district VARCHAR(255),
    school VARCHAR(255),
    primary_job VARCHAR(255),
    fte FLOAT,
    salary INT,
    certificate VARCHAR(255),
    subcategory VARCHAR(255),
    teaching_route VARCHAR(255),
    highly_qualified VARCHAR(255),
    experience_district INT,
```

```
    experience_nj INT,
    experience_total INT
);
"""
mycursor.execute(create_table_query)
print(f"Table '{table_name}' created or already exists.")
```

```
Table 'teachers_salaries' created or already exists.
```

0.5 Question 4 (30 pts)

Using LOAD DATA statement (as discussed in Module 4 lectures) load the data from nj__teachers__salaries__pset3.csv to your table created in Question-3. Use of OPTIONALLY ENCLOSED BY clause , TERMINATED BY clause and ESCAPED BY clause is recommended.

You will find Module 4 lectures helpful for this part, as well as the additional resources under module 4 on configuring your system to allow MYSQL file upload.

[61]:
```
# define path to csv file
csv_file_path = '/var/lib/mysql-files/nj_teachers_salaries_pset3.csv'

# load data from csv file into table
load_data_query = f"""
LOAD DATA LOCAL INFILE '{csv_file_path}'
INTO TABLE {table_name}
FIELDS TERMINATED BY ','
OPTIONALLY ENCLOSED BY '"'
ESCAPED BY '\\\\'
LINES TERMINATED BY '\n'
IGNORE 1 ROWS
(last_name, first_name, county, district, school, primary_job,
fte, salary, certificate, subcategory, teaching_route,
highly_qualified, experience_district, experience_nj, experience_total);
"""

# execute load data query
try:
    mycursor.execute(load_data_query)
    mydb.commit()
    print(f"Successfully loaded data from {csv_file_path} into {table_name}.")
except sq.Error as err:
    print(f"MySQL Error: {err}")
```

```
Successfully loaded data from /var/lib/mysql-
files/nj_teachers_salaries_pset3.csv into teachers_salaries.
```

[ ]:
```

```

0.5.1 Question 5-6 - For these questions you are only required to run the cells. To get credit your code from Question 1-4 must have been successfully run, and executed. All your data should be loaded in the previous step. No credit will be awarded if data was loaded using MySQL workbench.

0.6 Question 5 (15 pts)

Run the cell below. The code checks if all the data rows were stored in the database.

The code below assumes that you named your cursor object as mycursor(As specified in Question-1). If you named it differently, you can rename mycursor to match the variable name.

[63]:
```
cmd = "select count(*) from \
                 nj_state_teachers_salaries.teachers_salaries"
mycursor.execute(cmd)
count = mycursor.fetchone()[0]

print(f"Number of rows in teachers_salaries table : {count}")
```

```
Number of rows in teachers_salaries table : 100003
```

0.7 Question 6 (5 pts)

Run the cell below. The code checks if all the data columns were stored in the database.

[65]:
```
cmd = """SELECT COUNT(*) \
                FROM INFORMATION_SCHEMA.COLUMNS \
                WHERE table_schema = 'nj_state_teachers_salaries' \
                AND table_name = 'teachers_salaries'"""
mycursor.execute(cmd)
count = mycursor.fetchone()[0]
print(f"Number of columns in teachers_salaries table : {count}")
```

```
Number of columns in teachers_salaries table : 16
```

0.8 Question 7 (20 pts)

Submit a pdf report with the following information:

1. Describe the challenges you encountered in loading the data and how did you solve the issues you encountered.

2. Add OPTIONALLY ENCLOSED BY ',' in your SQL command when loading the data from the CSV file to the MySQL database. Did you observe any differences with and without it?

3. Add OPTIONALLY ENCLOSED BY "'' in your SQL command when loading the data from the CSV file to the MySQL database. Did you observe any differences with and without it?

4. Add ESCAPED BY '\\' in your SQL command when loading the data from the CSV file to the MySQL database. Did you observe any differences with and without it?

Answers need to be in complete sentences. Please be sure to explain briefly what OPTIONALLY ENCLOSED BY ‘,’, OPTIONALLY ENCLOSED BY ’“ ’and ESCAPED BY '\\' does. This file should be one-page max.

[ ]:

0.9 Submission on Gradescope

On canvas left menu -> click on Gradescope

Submit the jupyter notebook, a pdf report for question 7, and a pdf version of this notebook.

To create a pdf of this notebook : In your browser open print, and save as pdf. Name the pdf LastNameFirstName_pset3.pdf example: DoeJohn_pset3.pdf

Name this jupyter notebook with the same format LastNameFirstName.ipynb

You can name your report LastNameFirstName_report.pdf

Make sure that your notebook has been run before creating pdf. Any outputs from running the code needs to be clearly visible. We need both .ipynb, and pdf of this notebook to assign you grades.

Drop all the files in gradescope under PSET 3: Managing Data Exercise 1.

[ ]: