Instructions for Running the Python Insertion Programs

YELP DATABASE DESIGN & PROTOTYPING PROJECT

Python scripts file:
https://drive.google.com/file/d/1x_jU0s0kBbsAEtgKpc-RUUIL_KN7X602/view?usp=sharing

To load the Yelp dataset into the MySQL database, the insertion scripts must be executed in a Python environment with the appropriate dependencies installed. The project was developed and executed using **PyCharm**, but the steps apply to any IDE or command-line environment with minor adjustments.

1. Required Packages and Environment Setup

Before running any insertion script, ensure that the following Python packages are installed:

```
pip install mysql-connector-python
pip install pandas     # optional, (as this is JSON files)
```

The scripts rely on:

- **mysql-connector-python** for database connectivity
- The built-in **json** and **datetime** libraries for parsing and transforming data

No additional external dependencies are required.

You must also have:

- A running **MySQL server**
- A database created according to the project schema (e.g., `yelp_analysis`)
- **The Yelp JSON files are placed in the same directory as the insertion scripts**

2. Running the Scripts in PyCharm (recommended)

As these scripts were written using PyCharm, using this IDE to recreate the work is recommended.

In PyCharm:

1. Open the project folder containing the insertion scripts.
2. **Place each dataset file (e.g., `yelp_academic_dataset_user.json`) in the same directory as its corresponding script.**
3. Open the insertion script (e.g., `insert_user.py`).
4. Update the database credentials in the connection block:

```
conn = mysql.connector.connect(
    host="localhost",
    user="root",
    password="YOUR_PASSWORD",
    database="yelp_analysis"
)
```

5. Click **Run** in PyCharm to execute the script.
6. Wait for the console message confirming completion.

Because the dataset files are large—REVIEW (5.34 GB) and USER (3.36 GB), some scripts, especially **insert_user.py** and **insert_review.py** may take a long time to run.

3. Running the Scripts in Other IDEs (VS Code, Spyder, Jupyter, Terminal)

If using VS Code, Spyder, or any other Python IDE:

- The scripts run the same way, but the execution path may differ.
- **Ensure the working directory contains the JSON files.**
- In VS Code, you may need to explicitly set the Python interpreter and terminal environment using the Command Palette (`Ctrl+Shift+P` → "Python: Select Interpreter").
- In Jupyter Notebook, use the `%run insert_user.py` command or run script logic in cells (not recommended for multi-GB files).
- In a standard terminal, navigate to the folder containing the script and run:

```
python insert_user.py
```

Just make sure the JSON file is located in the same folder or adjust the file path in the script.

4. Order of Execution

Because of foreign key dependencies, the scripts should be executed **in the following sequence**:

1. `insert_user.py`
2. `insert_business.py`
3. `insert_review.py`
4. `insert_tip.py`
5. `insert_checkin.py`

Running them in this order ensures that referenced parent rows exist before child rows are inserted, preventing integrity errors.