**LAB 2 of DSA230/CIS492 - Cleveland State University**

**2:00pm-3:15pm of 11/3/2022**

**Due time: 11:30 pm of 11/8/2022**

*[100 points in total] Please finish this lab assignment independently. You may use notes, slides or internets for your references. If you have any questions, you can ask the TA for help. If you cannot finish all the tasks, do not worry and just submit whatever you have already written for partial credits. Lab submission late within 24 hours will receive 20% deduction for late submission. Lab submission late more than 24 hours but within 48 hours will receive 40% deduction for late submission. Lab submission late more than 48 hours is not accepted except a significant emergency. Code must be written in Python3 and submitted via Blackboard. Sorry about no email attachment for submission.*

**Objective**

Create a python3 program named ‘data\_cleaning.py’to clean data in different CSV files.

**Specification:**

In data science, we may always need to clean data in different CSV files and finally combine and save the cleaned data. Four database files **where CSU ID is the primary key** are shown as follows:

student.csv: Graphical user interface, table, Excel

Description automatically generated

gpa.csv: Table

Description automatically generated

balance.csv: Table

Description automatically generated

hometown.csv: Table

Description automatically generated

Please write Python3 program to combine the above 4 CSV files so as to automatically generate a file named ‘result.csv’ for the combined clean data, which should be **highly similar** to the follow:

Table

Description automatically generated

**Note:**

* The ‘student.csv’ file is perfect without noisy data.
* Other CSV files might have some noisy data (i.e., the rows whose CSU IDs never show up in ‘student.csv’ file), for example, CSU ID 77777 and 88888 in ‘hometown.csv’. Please ignore the noisy data when generating the ‘result.csv’ file.
* The student order of other CSV files might be different from that of ‘student.csv’ file.
* **You are required to** **use the Nested Dictionary as introduced in the class** **to implement your program**. The advanced Python packages (e.g., NUMPY, PANDAS, SCIPY, etc.) are NOT allowed in this lab.
* Manual hard coding will receive 0 points.
* The student information (id, name, gpa, balance, hometown, status) in CSV files is fake for this lab only.

**Grading:**

Please submit a single zipped file including your code (‘data\_cleaning.py’)and your automatically generated file (‘result.csv’) on Blackboard.