Student Data Analysis Database

Code on GitHub

Objective

This program was created to turn the dataset of 1000 student responses into a SQLite database. While I was working on the Student Data Analysis project, this was a huge help to be able to quickly query the database for patterns in the data.

Steps

Preparing the data

Once the CSV file was loaded into the program. The data needed to be “cleaned”. While cleaned can mean a lot of things, in this case, each data was wrapped in double quotation marks. The numbers were also in a string format. The first part of the code cleaned the data properly.

In order for the data to be properly inserted into the database. The SQL command ‘executemany’ takes the connection object and can run several commands in a row if they are all of the same format. The ‘executemany’ command looks for question marks (?) and then uses those as placeholders for the data. To use this command the cleaned data needed to be converted into a list of tuples (each tuple containing the data for one row of the database).

Creating the database

The database was created with the data types in mind. Since, in the last step, all possible integer values were converted into integers, the integer type was specified in the creation of the database. This made it much easier to query.

Querying the database

With the completed database, I left space at the end to set up an SQL query and display the results in the command line with pandas. The Pandas library was used to visualize the queries that were made.