MSSP 607: Practical Programming for Data Science
Homework 1, Due October 28, 2019

Part 1:

This homework assignment requires you to investigate some data from the federal Department of Education's 2014 School District Finance Survey. The homework comes with two files, named `state_sizes.csv` and `district_statistics.csv.` In no less than 100 words, why might someone want to look at this data? Why would it matter?

Part 2:

Choose any four of the following six questions. Without changing the function names or parameters in `homework1.py`, implement the function matching each of the four questions you've selected, and generate an answer. You can comment out or delete the functions for the two questions you skipped.

1. In our dataset, what percentage of school districts nationwide received more than $10,000,000 in local funding?

2. In our dataset, among school districts with a total budget (expenses) of at least $1 billion, which has the shortest name? What state is that district in?

3. Hawaii has only one school district, which covers the entire state. Across the rest of the United States, how many school districts had greater expenses than Hawaii did statewide?

4. The budget deficit of a school district is equal to its expenses minus its revenue. Which school district in Pennsylvania had the largest budget deficit? And how large was it?

5. New England comprises the states Maine, Vermont, New Hampshire, Rhode Island, Massachusetts, and Connecticut. To calculate per capita statistics, you divide the total amount by the total population size. Which state in New England gave the most non-federal money (state and local combined) to schools? How much did they give per capita?

6. The mean size of a school district in a state is equal to the total population divided by the total number of school districts. Which state had the smallest mean school district size? How many people, on average, live in each school district in that state?

Part 3:

Which state receives the "best deal" in school funding from the federal government? Why? Provide the source code you wrote to generate any statistics supporting your argument, and provide your answer in a professional and appropriate format that backs up your reasoning.

Extra Credit:

In no less than 300 words, describe some other questions you think could be interesting when researching school districts, their sources of funding, and their expenses. Why are those questions interesting? Are the values in the homework file enough to answer those questions? If not, what additional columns would you need to collect in order to answer them?

Part 1 (4 Points Total)			
Your Score	**Maximum Score**	**Dimension**	**Description**
	4	Reasoning	You'll receive up to 4 points for the quality of your reasoning about why this data is interesting and what questions it could help data scientists to answer.

Part 2 (24 Points Total) *Note: If you're thoughtful about it, all of these questions could be answered in Excel. Resist the urge! Correct answers with no Python code attempted will receive no credit.*			
Your Score	**Maximum Score**	**Dimension**	**Description**
	2 x 4 = 8	Correctness	For each question you attempt, you will receive 2 points for calculating the correct answer. If you give an incorrect answer but made a reasonable mistake, you can receive up to 1 point.
	4 x 4 = 16	Technical Quality	For each question you attempt, you will receive up to 4 points for the quality of your source code. Functions with no comments can receive no higher than a 3/4.

Part 3 (12 Points Total)			
Your Score	**Maximum Score**	**Dimension**	**Description**
	5	Reasoning	You'll receive up to 5 points for the quality of your written argument for how you decided to calculate which state received the "best deal" and what your answer was based on the data.
	5	Technical Quality	You'll receive up to 5 points for the quality of your source code and calculations that led to your answer.
	2	Presentation	You'll receive up to 2 points based on how clearly and professionally you communicate your result.

Extra Credit (5 Points Total)			
Your Score	**Maximum Score**	**Dimension**	**Description**
	3	Reasoning	You'll receive up to 3 points for the quality of you reasoning about why the questions you thought of are interesting and what we might learn from them.
	2	Necessary Data	You'll receive up to 2 points for describing how you could calculate the answers to your questions, either with the existing data in the homework file, or by proposing other metrics that would be necessary.

Your Total Score:	**/ 40**