Algorithms for Data Science
Assignment3
Data Specific Addition

This is the data specific addition to the assignment 3 specification.

The data we will be using a slightly modified file obtained from Zillow. The file is called Metro.csv and contains average housing price data from the major metro areas of the United States. If you open the file you will see the first line contains header information. And the second line contains averaged information for the entire country. Everything else is the data you will be using.

Each line contains information for a single metro area. We are interested in only the name of the metro area along with the most recent average (Jan 2020), and the increase over the last year (from Jan 2019 to Jan 2020). You will use this increase over the last year as your estimate of future gains for that city. The price you can buy houses for is the most recent price (Jan 2020).

You will use your DP algorithm to determine which metro areas to purchase house in. For this assignment, assume you can only buy a single house in each city. You are given 1 million dollars to use. Your DP algorithm should tell you how much estimated profit you will make, along with a list of the city names of your purchases.

You should make one modification to the algorithm as described in the spec document. Since 1 million dollars will create a table with one huge dimension and houses usually aren't bought down to the dollar, add a variable to your optimizeInvestments function. A described previously, it took the lists of investments and the dollar amount you have to spend. Add to this a third parameter that tell the smallest increment the dollar amount can have. For this particular data, it should be 1000. But don't hard-code the 1000 into the function – pass it in as a parameter so that you can get the original functionality of the algorithm by passing in a 1.

Lastly, I am not providing the results for this dataset for you to compare your answers to. That means, it is probably a good idea to create your own small dataset first and make sure your code works on that properly before trying it on the large housing dataset.