Overview:

This final project is a basic data science project that use much of the knowledge gained so far.

Description:

The United States, home to approximately 320 million citizens, is a large country made of 50 states and the nation's capital, the District of Columbia (also called Washington D.C.). In this project, you will explore some of the economic statistics of each state and produce a report. No data on U.S. territories are included in the file.

Data will be read from a CSV (comma-separated values) file called Economic_Data_2010.txt. The first row of a csv file usually consists of labels that describe the individual fields. Since we have not yet covered csv files this row has been eliminated.

Each row in the file contains the following information on each state in the United States from 2010. Values are separated by commas.:

1) State
2) Region (defined by the Bureau of Economic Analysis)
3) Population (in millions). The total number of people living in the state.
4) GDP (in billions). Measure of the state's economic activity, a higher GDP means higher monetary value for goods and services within the state's boarder.
5) Personal Income (in billions). All incomes received by individuals and households.
6) Subsidies (in millions). Money granted by the state's government to help an industry or business.
7) Compensation of Employees (in billions). Pre-taxed wages paid by employers to employees.
8) Taxes on Production and Imports (in billions). Taxes chargeable to business expenses of producing and importing

Note that Python uses Zero-based indexing, meaning that when data is put in a list the first value (State) will be found at the 0th index of the list.

Also recognize that some values are in millions while others are in billions, so when using operations on two values (which you will) make sure to adjust them accordingly.

The Bureau of Economic Analysis contains the following regions. This information is already in the provided data file in the Region column.

- Far_West: Alaska, California, Hawaii, Nevada, Oregon, Washington
- Great_Lakes: Illinois, Indiana, Michigan, Ohio, Wisconsin
- Mideast: Delaware, District of Columbia (Washington D.C.), Maryland, New_Jersey, New_York, Pennsylvania
- New_England: Connecticut, Maine, Massachusetts, New_Hampshire, Rhode_Island, Vermont
- Plains: Iowa, Kansas, Minnesota, Missouri, Nebraska, North_Dakota, South_Dakota

- Rocky_Mountain: Colorado, Idaho, Montana, Utah, Wyoming
- Southeast: Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North_Carolina, South_Carolina, Tennessee, Virginia, West_Virginia
- Southwest: Arizona, New_Mexico, Oklahoma, Texas

This data is provided by the U.S. Bureau of Economic Analysis..

Objective:

You are to provide economic statistics by region as follows:

- Total population for the region (sum of the column value)
- Total GDP for the region (sum of the fourth column)
- Average population in the region (total population / number of states in region)
- Average personal income (Total personal income / total population). You need to derive a meaningful measurement for this statistic. See the example below.

Your code must provide the following:

- Prompt the user for a region name.
- Read the data from the file selecting only the data for the region.
- Build three dictionaries for the region for the population, GDP, and personal income. The key for each dictionary is the State.
- In the case where the user entered an invalid region, an appropriate error message should be displayed. Note, you will not know this until you've read through the file and found no data.
- Write a function called calc_total_pop is passed the population dictionary and returns the total population. You will obviously need to iterate through the dictionary to produce the total population.
- Write a function called calc_total_gdp that is passed the gdp dictionary and returns the total GDP.
- Write a function called calc_total_pi that is passed the personal income dictionary and returns the total personal income.

The output of your must be as follows:

- A sorted list (not python list) of states in the region.
- The Total population for the region.
- The total GDP for the region.
- The average population of the region.
- The average personal income in the region.

Include any appropriate captions and labels to describe the above date. Below is a suggest output format. Points will be deducted for an untidy report.

```
Economic statistics for the Great_Lakes region:
   States in Region:    Illinois, Indiana, Michigan, Ohio, Wisconsin
   Total population:    46.436  million
   Average population:  9.2872  million
   Total GDP:           1776.04  billion
   Average PI:          37669.85
```

You may want to investigate the round function to reduce the number for decimals in floating point numbers. The join function will provide handy for showing all States on one line.

Note since this is one problem in one notebook, use multiple cells to modularize your program.

Once the project is complete and the results verified, upload the notebook to Blackboard.