

70-511: Statistical Programming  
Programming Assignment 5 – Data Preparations

### Introduction

The file `energy.csv` (attached) contains the values of energy generated for each country in Terawatt Hours for different years (from OECD Factbook 2011: Economic, Environmental and Social Statistics). The data contains missing values (denoted by `..`) and also aggregate rows for the EU, OECD, and World countries.

### Requirements

You are to create a program in Python that performs the following:

1. Loads the `energy.csv` file (assume it's in the current directory) and create a `DataFrame` object from it.
2. Replaces the missing values with the mean energy production for that country.
3. Removes the data for the aggregate values (EU27, OECD, World)
4. Adds a new column called 'Continent' and fills it in with the continent that corresponds to each country.
5. Creates a `DataFrame` that contains the continent name as the index and the following columns:
  - a. 'num\_countries' = number of countries for this continent
  - b. 'mean' = mean energy production of the countries in this continent
  - c. 'small\_production' = 1 if continent mean is less than the mean production of all countries minus one standard deviation; 0 otherwise
  - d. 'avg\_production' = 1 if continent mean is greater than the mean production of all countries minus one standard deviation, but less than the mean production of all countries plus one standard deviation; 0 otherwise
  - e. 'large\_production' = 1 if continent mean is greater than the mean production of all countries plus one standard deviation; 0 otherwise
6. Display the new `DataFrame` to the screen

**Additional Requirements**

1. The name of your source code file should be `DataPrep.py`. All your code should be within a single file.
2. You need to use the pandas `DataFrame` object for storing data.
3. Your code should follow good coding practices, including good use of whitespace and use of both inline and block comments.
4. You need to use meaningful identifier names that conform to standard naming conventions.
5. At the top of each file, you need to put in a block comment with the following information: your name, date, course name, semester, and assignment name.

**What to Turn In**

You will turn in the single `DataPrep.py` file using BlackBoard.

### Sample Program Output

---

70-511, [semester] [year]

NAME: [put your name here]

PROGRAMMING ASSIGNMENT #5

	num_countries	mean	small	avg	large
Europe	25	161.572326	0	1	0
Asia	7	607.983830	0	1	0
North America	3	1560.438095	0	0	1
South America	2	199.148901	0	1	0
Australia	1	218.021429	0	1	0
Africa	1	213.846154	0	1	0
Oceania	1	39.378571	0	1	0

---