**Assignment Problems: Day 2**

The following 3 problems require access to the file named ‘Dataset\_Day2.csv’, provided with this assignment. [**Use Numpy/Pandas packages**]

Some data info:

**Content**

Fields in the dataset:

* Name: Name of cereal
* mfr: Manufacturer of cereal
  + A = American Home Food Products;
  + G = General Mills
  + K = Kelloggs
  + N = Nabisco
  + P = Post
  + Q = Quaker Oats
  + R = Ralston Purina
* type:
  + cold
  + hot
* calories: calories per serving
* protein: grams of protein
* fat: grams of fat
* sodium: milligrams of sodium
* fiber: grams of dietary fiber
* carbo: grams of complex carbohydrates
* sugars: grams of sugars
* potass: milligrams of potassium
* vitamins: vitamins and minerals
* shelf: display shelf
* weight: weight in ounces of one serving
* cups: number of cups in one serving
* rating: a rating of the cereals (Possibly from Consumer Reports?)

Problems to solve:

1. Find all unique values of columns: “name”, “mfr”, “vitamins”   
   and store them in separate numpy arrays.
2. Create a new dataframe with all columns for which,

*‘sodium’* is greater than 100 AND ‘*protein’* is less than 3

Name this dataframe: *df\_HighSodLowProt*

1. From the dataframe in <2.> print the average ‘*calories’* by ‘*mfr’* and also print the ‘mfr’ with the highest average ‘calories’