PSG College of Technology
Department of Applied Mathematics and Computational Sciences
MSc SS VIII Semester – Data Mining
DM Lab Assignment 1 Data Analysis– 30-12-20

A nutritionist at the Food and Drug Administration is studying the effects of cereal marketing on family meal choices. To gather data for her study, the nutritionist goes to the local grocery store and records data about cereal nutritional claims and shelf location for 77 cereals.

1. Consider the variables in the data set. Identify and list the variables that are qualitative and those that are quantitative.
2. Consider the variable *Shelf*. This variable is the shelf position of the cereal (bottom, middle, top) starting from the floor up. To see whether the shelf position is associated with one measure of nutritive value, the amount of sugar, look at the data for the variable *Sugars*. Compare the sugar content of cereals on each shelf by making a separate histogram for the sugar content of the cereals on each shelf: a total of three histograms. Use the sugar content values as they are - do not factor in the serving size (Use the same scales for your histograms so you can compare the data easily.Title each histogram and label the axes.)
3. Briefly describe the distribution in each histogram with respect to shape. Based on your histograms, which shelf position has cereals with the most sugar?
4. Find the five-number-summary, mean, and standard deviation of the variable "*Fiber*".
5. Draw a scatterplot of the two variabes, with *Calories* on the y axis and *Carbohydrates* on the x axis. Does there appear to be a linear association? If so, is it positive or negative? Strong or weak?

Check whether you can find some other interesting facts from the data.

The data set is provided.