HW1

Muhammad Rashid

Student # 0975150

MS-ADS

Five insights of cereal datasets:

1. Most of the item is of type C, where it can be suggested that C type is mostly preferred by the end consumer.
2. As the fiber content increases in the items attributes, the potassium content also increases in amount.
3. Some of the items weight went to negative which does not make sense and can be discarded, similarly other attributes (sugar, etc.) negative values can be either discarded or put zero(null) instead.
4. Most of the cereals are produced by manufacturer (K) and secondly by G. Vitamins less than 25% can clearly be explored by sorting for not FDA recommended.
5. Most of the cereals are in shelf 3 here.

Exploration and analysis (paragraph):

By first looking at the dataset I realized that most of the items are of cold (C) type and this I realized by just seeing very few examples of hot (H) type. By doing simply numeric sort in excel, I saw that fiber content and potassium are in direct proportion. As the content of fiber increases for cereal the potassium content increases as such. As I was doing the sorting I saw negative values for weight, cups and sugar which does not make sense in this example. K manufacturer produced the most items in terms of cereals. We can looking for variety of cereals in shelf 3.

Challenges and problems:

The first problem that I saw with this dataset is that it contains certain negative values. This dataset must be normalized. Other challenges with this example dataset are that data is imbalance and will need more instances of data to better visualize and analyzed.