**Analytics queries**

* Why does the data comprise majorly of cold cereals?
* What are the top 5 cereals that have the maximum calorie content?
* Which cereal gives the maximum calorie content per serving?
* Are the ingredients linked to each other in any way?
* How are cereals arranged on shelves?

**Insights**

* Does a manufacturer exhibit some specific behavior?
* Is there some reason that a certain manufacturer has so many types of cereals in the market?
* Does a manufacturer has some cereal brands that have almost the same ingredients?
* Prices for the cereals are missing. How does the pricing affect the composition of a cereal could be a good insight?
* Which values have a high standard deviation and which ones have low? Why is such a behavior observed.

**Process you used to do the exploration and analysis**

I imported the data into Excel to analyze it. At the first glance, it looked like a data about all the cereals and their information content. It had columns for ingredient information, the calorie contents, manufacturers and the type of cereal. It also contained a column for ‘Shelf’; signifying that the data could belong to some supermarket store where the cereals are arranged in a certain way. Analyzing the values in the columns some values like ‘-1’ seemed ambiguous and I assumed that some data may have been wrongly entered or it did not pass the cleaning phase. I started observing the data column by column. Most of the data seemed to comprise of cold cereals and there was a dominance of two manufacturers who had most of the brands. Some columns had highly fluctuating values while some columns had consistent values. The deviation in the columns was also interesting. To analyze the data further, I added filters in the columns, which helped me to sort the data in ascending and descending order. I then created some line graphs to see the variation between the ingredients for every cereal. I was also interested in the shelve distribution of the cereals; grouping the data by shelves provided some information about the cereals. A good cereal is determined by the amount of calorie content per serving. To obtain this information I normalized the information of a serving size of 1 cup. I also tried to obtain a co-relation between the cereals using the inbuilt chart tools of excel. While analyzing the data, I was also wondering how can this information be of use to the consumer and how can it affect her buying options.

**Challenges and problems encountered**

The data was quite verbose and it covered all the ingredient information of a lot of cereals. There was a lot of things the data wanted to say but analyzing the data this way was not satisfying at all. I was looking at a lot of data right in front of me but I was unable to make out anything from it as it was not intuitive at all; I could hardly determine some strong observations from it. I got to know that certain manufacturers have a monopoly in the market but no reason to support the same. The data was also arranged in a random manner. I could not determine easily which cereal had the highest calorie count and which had the lowest. The data was also not represented clearly, it was unsure whether the weight given is in pounds or kg, what is serving cup size being used. Some values also had to be assumed, like calories were assumed in kcal, protein and carbohydrates were assumed in grams. Some false entries could be seen in the data for the value of cups as ‘-1’. This lead to a suspicion if the values given in Sodium column as ‘0’ are correct or a mere typo.   
The entire information given in the dataset was rather overwhelming, I could not make out anything from the data. How could this benefit to the user? If we had a column for a price, that would give an insight into how different manufacturers price their products and what is the composition at a given price. Also, it could help the consumer in making the decision of what product to buy. Clearly, no such information can be inferred from it. The data also does not give a quality measure about which cereal is good and which is not. It is confusing to interpret what composition would serve as a viable cereal. I believe that the data should have been able to tell me the details directly rather than me trying to find how to find those details.