INF 550: Data Science at Scale
Homework 1, Fall 2019
Due: Sep. 16, 2019. 3:30 PM

This assignment must be done independently.

The purpose of this assignment is to provide you with some experience exploring and analyzing data **without** using tools or programs. See the "HW1-dataset" file (that can be imported into Excel) about cereal products. The data set should be pretty self-explanatory. The Manufacturer is a one letter code with the expected mapping (Q-Quaker Oats, P-Post, G-General Mills, K-Kellogs, R-Ralston Purina, N-Nabisco) and Type is C (cold) or H (hot). You should explore and analyze this data using Excel or simply by hand (drawing pictures is fine), but do not use any analytics tools. Your goal here is to perform an exploratory analysis of the data set, to better understand the data set and its characteristics, and to develop insights about the cereal data.

Using the statistics and data exploration knowledge we studied at the class, describe what you find in this data exploration task. What you turn in should consist of three things.

1) First, list (bullet list of items) five "insights", chunks of knowledge, or deeper questions that you either encountered or gained while exploring the data. An insight could be some understanding of the data and its characteristics that is not relatively obvious or intuitive. It is something that most people might not realize initially (simple stat results such as means, finding wrong values are not appropriate insights). Note that an insight or knowledge chunk simply may be a deeper question that arose in your mind while exploring the data. And your analysis may not have been sufficient to answer the question.
2) Second, write one paragraph (don't be too long) about the process you used to do the exploration and analysis. Did you load the data into Excel, work manually, or do both? What did you do in Excel? Did you draw pictures? Just tell me (briefly) what you did.
3) Third, write one paragraph about challenges or problems that you encountered in doing the analysis this way. Did anything limit or frustrate you? If nothing did, perhaps there was something that was more difficult than you thought it should be. Nothing is perfect, so you should be able to list some potential issues here.

So, to sum up, your assignment should have a bullet list of five items followed by two paragraphs.

IMPORTANT: Turn in your writing in PDF format (max 2 pages) using the Blackboard. Emailing the assignment will not be accepted.

Grading: We will evaluate the quality of the insights you listed. We are looking for things that we find interesting or perhaps unexpected. Unique and interesting insights might get extra scores.