MIS0855: Data Science

In-Class Exercise on Wednesday, Jan 21 – Building a Data Dictionary

Objective: Build a data dictionary.

Learning Outcomes:

- Identify pre-assigned data types from an existing Microsoft Excel spreadsheet
- Deduce the purpose of individual data fields based on context and sample values
- Develop a plan for finding additional information about the meaning and units of measurement for data fields

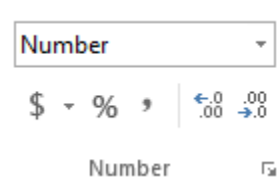**Step 1: Download and look at the spreadsheet (5 minutes)**

1) Download the spreadsheet "2015 Car Fuel Econ.xlsx" from the Community Site. This spreadsheet contains ratings from the Environmental Protection Agency for 729 cars from the 2015 model year.
2) Open the file. There are three tabs:
    a. FEguide – the main spreadsheet (all the data is there – note that "FE" means "Fuel Economy")
    b. Data Dictionary – this is what you're going to complete
    c. Credits – just some information about where the data comes from
3) Each column at the top of the FEguide tab is listed in the Data Dictionary

Step 2: Complete the Data Dictionary (20 minutes)

Your task is to create a useful data dictionary so future users of the file can quickly understand its contents.

1) Record the data type for each field (column) in the spreadsheet. You can find the data type by selecting any cell in the column and looking at the data type drop-down box in the "Number" section of the ribbon.

    For example, here is what you see when you select a cell in the Model Year column:

2) Record the units for each field. Sometimes this is obvious: the units for Model Year is obviously "year."

   However, sometimes it is less clear, such as engine displacement. For this, you may have to do a little research – for example try Googling "car engine displacement units."

3) Record a brief description for each field. Again, if this is not obvious you may need to do a little Internet research.

Step 3: Discussion (15 minutes)

We'll discuss your experience putting together the Data Dictionary. We'll address the following questions:

- What fields were the most difficult to complete? Why?
- Are you still not sure about one or more of the fields? If so, how would you go about getting enough information so that you felt confident?

Submit your Excel file on Blackboard by 10:00AM.