Assignment 3: Data Prep (Excel to Tableau Prep)

1. (30%) Replicate the cleaning steps in Excel demo, i.e., use Excel to (1) ensure the consistency in donor potential data (hi/high etc.), and (2) ensure donor name info consistency by extracting first / middle / last names from the donor names using Excel functions. The functions are written in the PPT file.

2. (25%) Create a Tableau Prep Builder Flow file (.tfl) that contains the following steps:
    a. Starts with the messy donations text file (from the Excel demo) as input
    b. Clean up the inconsistencies in the Donor Potentials field
    c. Create one calculated field to indicate whether the first name / last name of donor were reversed (flagRev), i.e., whether the name is in the format of John Doe or Doe, John
    d. Then, export the data into a .hyper or .twbx format (either one is fine).

    REMEMBER: You must **submit the TFL file (the flow) not the data file**! If you'd like you can submit the TFLX file which embeds the data, but you will be graded based on the flow you created. The TA will edit the connection to point to the messy CSV file, and run the flow. Your flow should be able to generate the expected output, with all those fields + flagRev + cleaned "Donor Potential" column.

3. (45%) Replicate Assignment 1 in Tableau Prep Builder, so that what you did there can be reproduced.
    a. Use the "Car data.xlsx" file provided to you here in this week's assignment.
    b. Use the two worksheets as two different tables, as inputs (one is car MPG, the other is gas price). Note that you do not need to do "unpivot" (reshape) as you did the previous exercise.
    c. Repeat the following steps:
        i. Start from the car MPG data, AGGREGATE to year – city MPG (you did this in Pivot Table in the first assignment)
        ii. Then use JOIN to link to the gas price table.
        iii. Then output your data to have three fields: year, avgMPG, gasPrice (use the "Gas Prices Adjusted for Inflation (In 2017 Dollars)")
        iv. The output should be a .hyper file.
    d. Submit both the TFL file and the HYPER file.