Data Cleaning Guidelines for Spreadsheets

1. **Create a backup copy in a separate workbook or sheet.**
    - This way you can always refer to the original dataset in case you make a mistake. You can rename the new sheet as “raw”.
    - Ensure that you note down the steps that you will take when performing data cleaning in a “Documentation” sheet.
2. **Ensure that the data is in a tabular format of rows and columns.**
    - Give the table a name.
    - Each column must have similar data (eg. all information on “last names”, “Color” or “Date of Birth” must be in their respective columns). This requires an understanding of what information is meant to be in each column.
    - All columns and rows must be visible.
    - There should be no completely blank columns and/or rows within the dataset.
3. **Fix any formatting issues that apply to the entire dataset.**
    - Apply the same font and font size.
    - Make sure the text is aligned correctly (align all columns to the left).
    - Make sure the text and the background fill are complementary. The standard is to use black text and a transparent/white background.
4. **Fix any formatting issues that apply to all elements of specific columns.**
    - Make sure that the formatting of the column is consistent (general, number, date, percentage, etc.).
    - Make sure that the correct type of formatting has been applied. Do not use a number format for years, or the percentage format for revenue figures.
5. **Make sure the information in each cell is easy to make sense of.**
    - If there is a lot of data, make sure the relevant rows and columns are frozen so you know what information you are reading. Freeze the first row.
    - Make sure column/row headers identify the information being presented in the cells.
    - Make sure each column/row has enough space to easily read the information presented in each cell. Increase the width/height of the column/row or wrap cells to allow this.
6. **Make sure there are no spelling mistakes, duplicates, and outliers.**
    - For now, a simple “Spelling & Grammar” check will do.
    - Remove any duplicates in the dataset.
    - Identify and explore outliers.