Data Cleaning for Mankind's Christmas Gift Co.

Dear Analyst,

The manager at Mankind's Christmas Gift Co. has tasked you with preparing the dataset for their end of year data analysis report.

You can access the dataset via the organization's link
https://github.com/MankindTheAnalyst/Data-Cleaning-with-Mankind-s-Christmas-Gift.git

Before diving into the analysis, it is crucial to clean the data and address any irregularities. Following the steps below:

1. Identify and remove repeated data entries so as to help in preserving unique data.
2. Inspect the "Gift_Item" column for irregularities in the texts. Implement a transformation to ensure all gift item names are written normally.
3. Investigate the dataset for any excess spaces in column values.
4. Generate a new column that combines the child's name and age, separated by hyphen (e.g., "James-11").

Documentation and Reporting:

✓ Devise a documentation strategy to record the changes made during the data cleaning process.
✓ Outline the key findings and improvements achieved through data cleaning.
✓ Discuss any challenges encountered during the cleaning process and propose solutions.

Remember to provide clear and detailed explanations for each step, ensuring that dataset is properly cleaned and ready for subsequent analysis.

We saw in your CV that you can work under high pressure, and so the deadline for this submission is 10AM tomorrow unless you like to wake up jobless.

Best of luck.

Hint

1. Remove Blank Rows

   - Select the entire data range

   - Go to the Home tab > Select Data > Remove Duplicates

   - Check the "Remove Duplicates" box and click OK to remove any fully blank rows

2. Handle Missing Values

   - Use the Find and Replace tool (Ctrl+H) to locate missing values like blanks or ?'s

   - Replace them with a consistent missing value indicator like "N/A" or "-"

3. Fix Inconsistent Data Entry

   - Use Find and Replace to standardize entries (e.g. kids -> child, tshirt -> t-shirt)

   - Create custom data validation lists for columns with pre-defined options

4. Split/Combine Columns

   - If names/addresses are smushed together, use Text to Columns to split them

   - If you have separate columns for first/last name, use CONCATENATE to combine

5. Remove Duplicates

   - Select the entire data range

   - Go to Data tab > Remove Duplicates

   - Check the columns you want to check for duplicates and click OK

6. Check for Outliers/Anomalies

   - Use conditional formatting to highlight numbers/dates outside expected ranges

   - Double check any outliers are not data entry errors

Be sure to create a backup of your original data before cleaning. With a few easy steps, your Christmas gift data will be shipshape!