

Dear Client,

I hope this message finds you in good health. I am reaching out to discuss the recent data quality assessment we conducted on the data provided for analysis.

I would like to address certain issues identified in the data and provide recommendations on how to address them effectively.

The assessment focused on various measures of data quality, including consistency, accuracy, completeness, uniqueness, and data validity.

Transactions Data:

1. The data is incomplete, with missing values.
2. The "product first sold" date appears as numerical values instead of actual dates.
3. The "list price" and "standard cost" fields lack relevance.
4. There is a missing "quantity" column.
5. The purpose of the "product class" column is unclear.

Customer Demographics:

1. The "gender" column contains multiple values and includes spelling errors such as "U", "F", "Femal", and "M."
2. The year values in the "date of birth" column need to be verified.
3. The "default" column is not easily interpretable.
4. Similar to the transactions data, there are missing values in the customer demographics.

Customer Addresses:

1. Some customer IDs present in the addresses data are not found in the customer demographics data.
2. The format of state names is inconsistent, with a mixture of short forms and full names.
3. The "property valuation" field lacks clarity regarding the unit of measurement

Based on these findings, I would like to offer the following recommendations:

1. Exercise caution when inputting date of birth values.
2. If you are using Excel, utilize filter options to ensure data consistency.
3. Double-check or cross-verify values entered into cells.
4. Establish clear formatting guidelines for each column, allowing data to be inserted only in specific formats.

5. Highlight any empty cells to draw attention to missing data.

Please feel free to reach out if you have any questions or require further clarification.

Best regards,

Sumanyadav

sumanyadavpodamekala@gmail.com