Subject: Data Quality Assessment Report

Dear Sir,

We have conducted a comprehensive assessment of your data and identified key issues that may impact its quality and usability. Below is a summary of our findings and recommendations for improvement.

Customer Demographic Table

Completeness:

- Last Name column: 97% complete
- Date of Birth (DOB): 98% complete
- Job Title: 88% complete
- Tenure: 98% complete

Accuracy:

- The DOB column contains an anomaly: Jephthah Bachmann's age sums up to 175 years, which is unrealistic.

Consistency:

- The Gender column is inconsistent; expected values are "Male" and "Female," but variations such as "F" and "U" are present.
- The Gender column is 96% consistent.

Validity:

- Last Name column: 97% valid
- DOB: 98% valid
- Job Title: 88% valid
- Tenure: 98% valid
- The Default column is invalid and irrelevant.

Uniqueness:

- No duplicate entries were found.

Relevance:

- Columns 14 to 26 are deemed irrelevant to the dataset.
- The Default column is both invalid and unnecessary.

Transactions Table

Completeness:

- Several columns contain missing values.

Validity:

- The product_first_sold_date the column is in an incorrect format.
- The list price column should be formatted as a currency field.

Customer Address Table

Completeness:

- Several columns have missing values.

Accuracy:

The DOB column contains an incorrect value: Jephthah Bachmann's age sums up to 175 years, which is unrealistic.

Consistency:

- The State column contains inconsistent entries (e.g., "VIC" vs. "Victoria" and "NSW" vs. "New South Wales").
- Our Recommendations & Data Cleaning Approach

Recommendations & Data Cleaning Approach

To improve the overall quality of your data, we propose the following actions:

Handling Missing Data:

- We will address blank fields to ensure optimal reporting based on the available data.

Removing Irrelevant and Invalid Data:

- Unnecessary columns will be removed to optimize model size, maintain relevance, and improve analytical accuracy.

Ensuring Consistency:

- Inconsistent values will be standardized to ensure uniformity across all datasets.

Correcting Data Formats:

- Incorrect formats and data types will be corrected using our advanced data processing tools.

Eliminating Inaccuracies:

- Any inaccurate or erroneous data will be identified and removed.

How You Can Assist Us

To facilitate a smooth data cleansing process, we kindly request the following from you:

- Provide any missing data if available.
- Verify the incorrect data we identified, particularly Jephthah Bachmann's DOB.

We appreciate your cooperation and look forward to your feedback. Please let us know how you'd like to proceed.

Sincerely,

Amit Sharma

Data Analyst