

## Subject: Assessment of Data Quality Issues and Mitigation Strategies

Dear Client,

I hope this email finds you well. I wanted to bring to your attention some observations regarding the data quality of the recent dataset we've been working with. Our team has conducted a thorough evaluation, and we've identified several issues that require attention. Below is a summary of the data quality issues observed, categorized according to the Standard Data Quality Dimensions:

- Accuracy:

Outlier in Date of Birth: We noticed anomalies in the Date of Birth field (CustomerID#34), suggesting potential inaccuracies. This requires a closer examination to ensure the accuracy of this crucial information.

- Completeness:

Null values/Incomplete values: Several entries in the dataset exhibit missing or incomplete values. It is imperative to address these gaps to maintain the completeness of our data.

- Consistency:

Inconsistent gender and states category: The gender and states category appear inconsistent across entries. Achieving uniformity in this field will enhance the overall consistency of the dataset.

- Validity:

Wrong data type of Product First Sold Date: The data type for the Product First Sold Date field seems to be incorrect (Whole Number instead of Date). Validating and correcting this will ensure the dataset adheres to expected standards.

- Currency:

No currency on list price: The absence of currency indicators on list prices can lead to ambiguity. We recommend incorporating currency values to maintain clarity and consistency.

- Relevancy:

Irrelevant data on Default Column: Some fields contain irrelevant information, which may impact the dataset's relevance for our analysis. Streamlining the dataset by removing extraneous data will enhance its overall utility.

- Uniqueness:

Invalid character on Default column: An invalid character has been identified in the default column, potentially impacting uniqueness. Rectifying this will ensure the uniqueness and integrity of the dataset.

In light of these observations, our team recommends the following strategies to mitigate these data quality issues:

- Implement thorough data validation checks to identify and rectify inaccurate and invalid entries.
- Conduct a comprehensive review of data types, ensuring that each field aligns with its intended data format.
- Develop and apply data cleansing procedures to address null values, incomplete entries, and irrelevant information.
- Standardize gender categories and other categorical data to enhance consistency.
- Introduce currency indicators for list prices to maintain clarity and facilitate accurate analyses.

By addressing these issues, we aim to enhance the overall quality and reliability of the dataset, ensuring that it aligns with the highest data standards. If you have any questions or would like further clarification on any of these points, please feel free to reach out.

Thank you for your attention to this matter. We look forward to working collaboratively to optimize the quality of our data.

Best regards,

Arsie Mauricio

Data Analyst