Dear Client,

First of all, we would like to thank you for giving these datasets, and during the processing on datasets, we get the following overview. Please let us know if the figures are not aligned with your understanding

**Summary Table**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table Name** | **Accuracy** | **Completeness** | **Consistency** | **CURRENCY** | **Relevancy** | **Validity** |
| **Transaction** | * Profit: missing * Transaction month: missing | * Online Order: Blanks * Brands: Blanks |  |  |  | * List Price: Format * Product Sold Date: Format |
|  |  |  |  |  |  |  |
| **Customer Demographic** | * Age: missing | * Job title: Blanks | * Gender: inconsistency | * Deceased customers: filter out | * Default column: delete |  |
|  |  |  |  |  |  |  |
| **Customer Address** |  |  | * States: inconsistency |  |  |  |

Below are more in-depth descriptions of data quality issues discovered and methods of mitigation used. Recommendations and explanations have also been included to avoid further data quality issues in the future. The following recommendations will improve the accuracy of data used to influence business decisions.

**Accuracy Issues**

* **Missing age column for “Customer Demographic”, Transaction month and profit column for “Transactions”**

Recommendations: create an **age column**, allowing for more comprehensible data and easier to check for errors. Create **Transaction month** and **profit column** in **“Transaction"** to check accuracy sales.

**Completeness**

* **Blanks in job\_title for “Customer Demographic”, Online\_Order and Brands for “Transactions”**

Mitigation: Filter out ‘blanks’ for **job\_title, online\_order and brands column.**

Recommendations: Simplify job\_title to another category such as **industry\_industry** or provide dropdown options for **job\_title** .Provide dropdown options for online\_orders and brand\_column

**Consistency**

* **Inconsistency in gender for “Customer Demographic” and states for “Customer Address”**

Mitigation: Filter all ‘M’ under category of ‘Male', filter all ‘Femal’ and ‘F’ under ‘Female’ for **gender**. Filter all ‘NSW’ to ‘New South Males’ and ‘VIC’ to ‘Victoria’ for **states**. Recommendations: Create dropdown options for ‘Male', 'Female’ and ‘U’ in **gender**. Create dropdown options for all **state** abbreviations.

Dropdown options minimizes manual entry and human errors

**Currency**

* **People that are ‘Y’ in deceased\_indicator are not current customers for “Customer Demographic”**

Mitigation: Filter out customers checked ‘Y’ in deceased\_indicator Recommendations: can be difficult to check for deceased customers, but once this information is received one should update the data accordingly.

Deceased customers are not current customers, removing them from the data will increase currency of data and will result in more accurate estimated in future analysis.

**Relevancy**

* **Lack of relevancy or comprehensibility in default\_column for “Customer Demographic”**

Mitigation: Deleted Metadata in a default column Recommendations: check for incomprehensible Metadata and delete or format to make comprehensible.

**Validity**

* **Format of list\_price , Product\_sale\_date for “Transaction”**

Mitigation: Format product\_sale\_date to short date format, format list\_price to currency. Recommendations: Set up columns so that formats such as price and decimals are already in place when entering the new data.

That summarizes all the data quality issues discovered through the first stage of the data quality analysis. The mitigation strategies suggested are simple and effective ways of improving the data quality for future analysis

Please let us know if you have any questions regarding mitigation or any data quality issues identified.

Kind regards, Aman Agrawal