**Data Quality Framework Table**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Customer Demographic | Customer Address | Transaction past 3 months |
| Accuracy | * Dob = 1 data inaccurate * Default = All data inaccurate | √ | √ |
| Completeness | * dob = 87 Blanks * job title = 506 Blanks * industry category = 656 Blanks * default = 240 Blanks * tenure = 87 Blanks | √ | * Online Order = 360 Blanks * Brand = 197 Blanks * Product line = 197 Blanks * Product Class = 197 Blanks * Standart Coss = 197 Blanks |
| Consistency | * Gender = have data (F, Femal, Female, M, Male, and U) | * State have inconsistent data | * Standart Cost have 2 data with wrong format |
| Currency | √ | √ | √ |
| Relevancy | √ | √ | √ |
| validity | √ | √ | √ |
| Uniqueness | √ | √ | √ |

For a better understanding regarding the issues in the data in the table above, we have written an explanation below, this explanation also contains several recommendations to make the data cleaner so that it can be easier and produce accurate visualizations.

# **Issues and Recommendation**

## **Customer demographic data**

**The first issue** that this dataset has is **Accuracy**, which we can see clearly in **default column**, this column has values that are abstract, unclear, and cannot be analyzed, besides that **the dob column** also has 1 data that is incorrect or inappropriate, namely 1843-12-21, which should be in the format 12/21/1843.

With the aim of getting better data or information, below are suggestions or input from us:

* Removing the Default column
* Correcting the wrong format in the Dob column
* Create age column and group\_of\_age column in order to widen the analysis process

**The second issue is that this dataset has is Completeness,** in this table there are 5 columns that have null values, namely : dob column, job title column, industry category column, default column, and tenure column, with detailed amounts as shown in the table above.

With the aim of getting better data or information, below are suggestions or input from us:

* Changing the null values in a text or string column with “N/A” or Not Available.
* Use a replacement value such as (mean, mode, or median) for null values in numeric column, with a note that the number of null values in that column is small.

**The third problem this dataset has is Consistency**, we can clearly see in gender column that the gender labeling male or female is inconsistent, sometimes female and male, sometimes f and m, there are also typos like female

With the aim of getting better data or information, below are suggestions or input from us:

* Make clear decisions for gender labeling, using only initials or using full words

## **Customer address data**

**The only problem in this dataset is Consistency**, we can clearly see in state column that the state labeling is inconsistent, sometimes using initials sometimes using full state name.

With the aim of getting better data or information, below are suggestions or input from us:

* Make clear decisions for state labeling, using only initials or using full words

## **Transaction past 3 months data**

**The first issue is that this dataset has is Completeness,** in this table there are 5 columns that have null values, namely : Online Order column, Brand column, Product line column, Product Class column, and Standard Cost column, with detailed amounts as shown in the table above.

With the aim of getting better data or information, below are suggestions or input from us:

* Changing the null values in a text or string column with “N/A” or Not Available.
* Use a replacement value such as (mean, mode, or median) for null values in numeric column, with a note that the number of null values in that column is small.

**The second problem this dataset has is Consistency**, we can see some inconsistency when we use filter on standart cost column, this column should fill with value with format currency, but theres value still use general format which is 270.299, 312.735, and 667.400

With the aim of getting better data or information, below are suggestions or input from us:

* Change the wrong value format with currency format.

Moving forward, we will analysis your dataset, so we can get insight from the datasets provided, and it can be used for future decision making. Once again, thank you for having us as your partner. It is nice to work with you.

Best Regards,

Dimas Afrizal