Subject: Data Quality Assessment and Recommendations for Improvement

Dear Client,

I hope this email finds you well. As part of our ongoing collaboration, we have conducted a preliminary data quality assessment on your datasets. Our analysis focused on the following data quality dimensions: Accuracy, Completeness, Consistency, Currency, Relevancy, Validity, and Uniqueness.

**Transactions Data**

* **Completeness**: We observed missing values in multiple columns such as **online\_order, brand,** and **product\_line**. We recommend conducting an internal review to identify the cause of these missing values. It would be better if the date format is changed in the column “product\_first\_sold\_date” to “dd-mm-yyyy”, for better picture.
* **Consistency**: The data types for each field appear to be consistent and appropriate for the type of information they hold.
* **Relevancy**: All fields may not be relevant for all analyses. It might be beneficial to consult with the data collection team to eliminate any unnecessary fields.

**New Customer List**

* **Completeness**: The dataset has missing values in **last\_name, DOB, job\_title,** and **job\_industry\_category.** It is crucial to address these gaps for a comprehensive analysis.
* **Relevancy**: The dataset contains unnamed columns that might not be relevant for analysis and should be reviewed.

**Customer Demographic**

* **Completeness**: Missing values are present in several fields like **last\_name, DOB, job\_title,** etc. These missing values could hinder the depth and reliability of our analysis.
* **Validity**: The **default** column contains values that look like SQL commands or nonsensical strings. We highly recommend reviewing the validity of this field.
* **Accuracy**: The year of birth of name “Jephthah Bachmann” was given 1843 in the table, which means he is 175 years old. This violates Accuracy.

**Customer Address**

* **Completeness**: No missing values were found, which is a positive indicator.
* **Consistency**: Data types are appropriate, but it's worth noting that the **state** field has different formats. Standardization may be beneficial here.
* **Relevancy**: As all customers are from Australia, it would be good to remove the ‘country’ column.

**General Recommendations**

* **Accuracy & Validity**: The accuracy and validity of the data require domain expertise for thorough verification. We suggest conducting an internal audit to ensure all data points are accurate and valid.
* **Currency**: We recommend adding a last-updated field to maintain the currency of the data.
* **Uniqueness**: No duplicate records were found, which is a good indicator of data quality in this regard.

Please note that data quality is a shared responsibility and requires ongoing attention to maintain high standards. We are available to assist you in implementing these recommendations and look forward to continuing our collaboration.

Best regards,

Advait Manduva