**Subject: Data Quality Issues and Mitigation Strategies**

Dear Sir/Madam,

I am writing to address the data quality issues identified in our datasets and propose strategies to mitigate these issues. Below, I have outlined the data quality issues and provided steps to address them:

**1. Completeness (Fields with Missing Values):**

**Transactions Sheet:**

* + - Columns with missing values: online\_order, brand, product\_line, product\_class, product\_size, standard cost, product first sold date.
    - Number of Rows (Before Cleaning): 20,000
    - Number of Rows (After Cleaning): 19,946

**NewCustomerList:**

* Columns with missing values: DOB, job title, job industry.
* Number of Rows (Before Cleaning): 1,000
* Number of Rows (After Cleaning): 736

**CustomerDemographic:**

* Columns with missing values: DOB, job title, job industry.
* Number of Rows (Before Cleaning): 4,000
* Number of Rows (After Cleaning): 2,863

**CustomerAddress:**

* Columns with missing values: None
* Number of Rows (Before Cleaning): 4,000
* Number of Rows (After Cleaning): 4,000

**Mitigation Steps:**

The Data Handling team should attempt to fill null values by cross-referencing old data or relevant records.

**2. Relevancy (Irrelevant Values):**

**NewCustomerList:**

* Deleted value: "U" in the gender column since it's unclear if it represents male or female.

**CustomerDemographics:**

* Deleted value: "U" in the gender column since it's unclear if it represents male or female.
* Deleted value: "Y" in the deceased\_indicator column to retain only living customers.
* Deleted the "default" column as it is not decodable.

**Mitigation Steps:**

The Data Handling team should investigate the "U" values in the gender column to determine the correct values.

**3. Consistency (Up-to-Date Values):**

**CustomerDemographics:**

* Changed value: "F" to "Female" and "M" to "Male" in the gender column.

**CustomerAddress:**

* Changed value: "New South Wales" to "NSW" and "Victoria" to "VIC" in the state column.

**Mitigation Steps:**

The Data Handling team should ensure that data values are consistent and up-to-date, especially for fields like gender and state.

**4. Accuracy (Values in Incorrect Format):**

**Transactions:**

* Changed data type from number to date for the product first sold date column.

**Mitigation Steps:**

The Data Handling team should verify and correct data types to ensure accuracy.

Addressing these data quality issues and implementing the proposed mitigation strategies will help us maintain a cleaner and more reliable dataset. If you have any questions or require further clarification on these measures, please feel free to reach out.

Thank you for your attention to this matter.

Sincerely,

Piyush Singh