

**Subject: Data Quality Assessment and Mitigation Plan**

Dear Sprocket Central Pty Ltd Team ,

I hope this message finds you well. We have completed a comprehensive assessment of the data provided by your company and would like to present our findings regarding data quality, along with recommendations to address the identified issues.

**Data Quality Assessment :**

We conducted an in-depth analysis of the three datasets you shared with us: Customer Demographic, Customer Addresses, and Transaction Data. Here are our key observations:

**1. Data Accuracy and Completeness:**

**Transaction Data:** We found no missing values or duplicate records in the transaction dataset. This dataset appears to be highly accurate and complete.

**Demographic Data:** We observed missing values in several columns, most notably in the "last\_name," "DOB," "job\_title," and "job\_industry\_category" fields. Approximately 16.4% of the records have missing values in the "job\_industry\_category" column. We also detected some data inconsistencies, such as the "gender" column containing multiple variations of gender labels, which have been standardized for consistency.

**Customer Data:** Similar to the Demographic Data, the Customer Data contains missing values in various columns, including "online\_order," "brand," "product\_line," "product\_class," "product\_size," "standard\_cost," and "product\_first\_sold\_date." These missing values account for less than 5% of the dataset.

**2. Data Consistency :**

We have identified inconsistencies in the "gender" column of the Demographic Data, which have been rectified by standardizing the gender labels.

**Recommendations for Data Quality Improvement:**

To ensure that the data is of the highest quality for our analysis, we propose the following strategies:

**1. Address Missing Values :**

We recommend working with your team to address the missing values in the "Demographic Data" and "Customer Data" datasets. One option is to gather missing data from reliable sources, if available. Another approach is to consider data imputation techniques, such as filling missing values with means or modes, based on the nature of the data.

**2. Data Standardization :**

Maintaining consistent data standards is essential. We have already standardized the "gender" column. We suggest applying a similar standardization process to other categorical data fields to enhance consistency.

### 3. Quality Control :

Implement quality control measures to monitor data entry and validate data accuracy in real-time. Regular data audits can help identify and rectify data inconsistencies.

### 4. Data Governance :

Consider establishing a robust data governance framework to ensure data quality control and management throughout your organization.

### 5. Client Communication:

Maintain open lines of communication with clients to address any potential data issues proactively. This can help identify and resolve data quality concerns more effectively.

### 6. Currency :

Regarding the data quality assessment, we have identified that the oldest transaction date in the 'customer\_data' dataset is '2017-01-01 ', while the most recent transaction date is '2017-12-30 '. Additionally, we have successfully converted the 'product\_first\_sold\_date' column to a datetime format.

Please note that we are only interested in data for the last three months in our dataset for analysis but the previous dates would also prove so helpful for our analysis.

### Data Quality Framework :

For reference, we used a data quality framework to assess the datasets, which includes dimensions like accuracy, completeness, consistency, and timeliness. These dimensions guided our evaluation.

Please feel free to reach out if you have any questions or require further assistance. We are committed to supporting Sprocket Central Pty Ltd in optimizing your data for informed decision-making.

Thank you for entrusting us with this critical task.

Sincerely,

Martin Dudi

Data Analyst.

KPMG Data Quality Assessment Team