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

Dear Sprocket Central Pty Ltd,

I hope this email finds you well. As discussed in our recent voicemail, I wanted to provide you with an update on the data quality assessment of the datasets received. Our aim is to ensure that the data is prepared for our upcoming analysis in phase two of the project.

Upon careful examination of the datasets, we have identified several data quality issues that require attention. Below, I have detailed these concerns and proposed strategies for their mitigation.

**Customer Demographic Data:**

* **Consistency**: In the "State" column, we have observed discrepancies due to the use of abbreviations and variations in state names. To promote uniformity, we recommend standardizing state values.

**Customer Address Data**:

* **Consistency**: The "Gender" column exhibits inconsistencies with the use of abbreviations and variations in gender names. Standardizing gender values, along with referencing 'U' for 'Unknown,' is advisable.
* **Completeness**: Notably, the "Job Title" and "Job Industry" columns contain missing values. To address this, we suggest the application of data imputation techniques, particularly mode imputation based on existing data.
* **Accuracy**: There are instances of inconsistency and incorrect formatting in the "DOB" (Date of Birth) column. Enhancing accuracy necessitates validation and standardization of date of birth entries.
* **Relevance**: We have identified the "Default" column as irrelevant to our analysis. We recommend its removal to ensure a clean dataset.

**Transaction Data in the Past Three Months**:

* **Completeness**: The "Online Orders" and "Brand" columns reveal missing values that require attention. We recommend a thorough review of the dataset with a focus on completing these entries.
* **Validity**: The "Product First Sold Date" column appears to be incorrectly formatted. It is crucial to ensure that the entries adhere to the correct date/time format.

**To effectively address these data quality concerns, we propose the following actions:**

* Verify the correct formatting of column entries.
* Identify and address missing entries using appropriate methods.
* Enforce consistency checks across the dataset.
* Identify and remove irrelevant columns to ensure dataset cleanliness.

We greatly value your cooperation in ensuring the data's preparedness for our phase two analysis. Your feedback and insights regarding the recommended strategies are invaluable as we proceed.

Thank you for your attention to this matter, and we eagerly anticipate your response.

Best Regards,

Mahender Kore,

Data analyst team, KPMG.