Dear [Recipient's Name],

I hope this email finds you well. Thank you for providing us with the three datasets from Sprocket Central Pty Ltd.

The below table highlights the summary statistics from the three datasets received. Please let us know if the figures are not aligned with your understanding.

Table name	No. of records	Distinct Customer IDs	Date Data Received
Customer Demographic	4,001	4,000	19/07/2023
Customer Address	4,004	4,003	19/07/2023
Transaction Data	20,000	3,494	19/07/2023

Here are some data quality issues that I found and my recommendation to avoid the re-occurrence of data quality issues and improve the accuracy of the underlying data used to drive business decisions.

- **Additional customer_ids in the 'Transactions table' and 'Customer Address table' but not in 'Customer Master (Customer Demographic)'**
  Mitigation: Please ensure that all tables are from the same period. Only customers in the Customer Master list will be used as a training set for our model.
  This indicates that the data received may not be in sync with each other which may skew the analysis results if there are missing data records. Please refer to excel file 'data_outliers.xlsx' for the list of outliers between tables.
- **Various columns, such as the brand of a purchase, or job title, have empty values in certain records**
  Mitigation: Use regular expression to replaced extended values into abbreviations to ensure consistency across addresses.
  Recommendation: Enforce a drop-down list for the user entering the data rather than a free text field.
  Gender records where 'U' have been replaced based on the distribution from the training dataset.
- **Inconsistent data type for the same attribute**
  Mitigation: Convert selected records in characters to numeric. Remove non-numeric characters from string.
  Recommendation: Ensure that fact tables in the given database have constraints on data types. Appropriate data transformations are made to ensure consistent data types for a given field.

I believe that by proactively addressing data quality issues and implementing effective mitigation strategies, we can enhance our decision-making processes, improve operational efficiency, and drive overall success. Questions will be raised along the way and assumptions documented. After we have completed this, it would be great to spend some time with your data SME to ensure that all assumptions are aligned with Sprocket Central's understanding.

Best regards,

Nguyen Van Dong