Subject: Data Quality Assessment and Recommendations

Dear Client,

I hope this message finds you well. We appreciate your trust in Sprocket Central Pty Ltd for your data analysis needs. Following our initial meeting, we have thoroughly reviewed the datasets you provided and have conducted a comprehensive assessment of their data quality. In this email, we will outline the identified data quality issues and present strategies to mitigate these concerns.

Data Quality Issues:

1. Customer Demographic Dataset:

- Missing Values: We observed missing values in the "Age" and "Gender" columns. This can affect our analysis by limiting the scope of demographic insights.

- Inconsistent Data Types: In the "DOB" (Date of Birth) column, there is a mix of date formats, making it challenging to standardize and analyse this information efficiently.

- Duplicate Records: We identified duplicate entries for some customers, potentially skewing our analysis results.

2. Customer Addresses Dataset:

- Incomplete Addresses: Several records lack complete address information, making it difficult to accurately determine geographical patterns or perform location-based analysis.

- Inconsistent Formatting: Address information shows inconsistent formatting, which may hinder the ability to geocode addresses and accurately map customer locations.

3. Transaction Data in the Past Three Months:

- Data Gaps: There are gaps in the transaction data, with some days missing entirely. This may affect our ability to analyse daily trends or seasonality accurately.

- Outliers: We noticed outliers in the "Transaction Amount" column, which could potentially skew statistical analysis and insights.

Strategies to Mitigate Data Quality Issues:

1. Customer Demographic Dataset:

- Missing Values: We recommend imputing missing values in the "Age" and "Gender" columns using appropriate techniques such as mean, median, or predictive modelling.

- Inconsistent Data Types: Standardize the date format in the "DOB" column to ensure uniformity and ease of analysis. Remove or correct entries with invalid formats.

- Duplicate Records: Deduplication procedures should be applied to remove redundant customer records.

2. Customer Addresses Dataset:

- Incomplete Addresses: Reach out to customers to collect missing address details, ensuring that we have complete and accurate address information.

- Inconsistent Formatting: Standardize the address format to facilitate geocoding and mapping.

3. Transaction Data in the Past Three Months:

- Data Gaps: Fill in the missing transaction data with appropriate values or use interpolation methods to estimate missing data points.

- Outliers: Identify and address outliers by either excluding them from analysis or applying robust statistical techniques.

We will use the provided "Data Quality Framework Table" as a guide to ensure that we address all relevant criteria and dimensions in our data quality improvement process.

Additionally, please note that we might need to clarify certain assumptions or data-related queries with you during this process to ensure the best possible outcomes for your analysis.

Thank you for entrusting us with your data analysis needs. We will diligently work on resolving these data quality issues and keep you updated on our progress. If you have any questions or require further information, please do not hesitate to contact us.

Kind regards,

Natwar Upadhyay

Data Analyst

Sprocket Central Pty Ltd