Subject: Request for Quality Data for Analytical Purposes - Data Quality Assessment Findings and Recommendations

Dear **[Stakeholder's Name]**,

I hope this email finds you well. I am writing to provide you with the findings of a comprehensive data quality assessment conducted during my internship with The Forage, where we analyzed the dataset provided by Sprocket Central Pty Ltd. The objective of this assessment was to evaluate the accuracy, completeness, consistency, currency, relevance, uniqueness, and validity of the data. I have identified certain challenges within the dataset that require attention, and I would like to request your assistance in resolving these issues to ensure the availability of high-quality data for analytical purposes.
Below are the findings and recommendations for each table within the dataset:

Table 1 - Customer Data:

1. Missing Last Name: There are 125 records with missing last names. [Completeness]
2. Inconsistent Gender Values: The "gender" column contains various representations of gender, such as "F," "femal," "M," "U," "female," and "male." [Consistency]
3. Null Values in Job Title: 506 records do not have a job title. [Completeness]
4. "n/a" Job Industry Category: 656 records have "n/a" as the job industry category. [Validity]
5. Inadequate Date Format: The date format in the "DOB" column needs to be reviewed and standardized. [Accuracy, Consistency]
6. Garbage Values in Default Column: The "default" column contains garbage values, and there are 240 blank values. [Accuracy, Completeness]
7. Blank Values in Tenure: There are 87 records with blank values in the tenure column. [Completeness]

Table 2 - Address Data:

1. Duplicacy in State Values: The "state" column contains both "New South Wales" and "NSW" as state representations. [Consistency]

Table 3 - New Customer Data:

1. Null Values in New Customer List: There are 29 records with null values in the "new customer list" column. [Completeness]
2. Inconsistent Gender Value: The "gender" column contains the value "U." [Consistency]
3. Blank Values in Job Titles: There are 106 records with blank values in the "job_title" column. [Completeness]
4. "n/a" Value in Job Industry Category: There are 165 records with "n/a" as the job industry category. [Validity]
5. Float Values in Property Values: Some property values are represented as floats instead of integers. [Validity]
6. String Format in Postal Code, Property Value, and Past 3 Purchase Columns: The "postcode," "property_valuation," and "past_3_years_bike_related_purchases" columns are represented as strings instead of integers. [Validity]
7. Unlabeled Columns: Columns Q, R, S, T, and U have no titles or labels, despite containing data. [Consistency]

Table 4 - Transaction Data:

1. Null Values in Online Order: There are 360 records with null values in the "online_order" column. [Completeness]
2. Missing Values in Brand, Size, Line, Class, Standard Cost, and Product Purchase Date: There are 197 records with missing values for these attributes. [Completeness]
3. Inadequate Date Format in Product Purchase Date: The "product_first_sold_date" column contains date values in a format that may not be suitable for comprehensive analysis. [Accuracy, Consistency]

Based on the assessment findings, I have marked the suggested changes in yellow to indicate where modifications have been made, and in red to highlight areas where changes are required. This visual representation will help identify specific areas in need of attention.

To address these challenges, I **recommend** the following actions:

1. Ensure complete data collection for missing last names, job titles, and tenure. [Completeness]
2. Standardize gender values as "Male" and "Female" for consistency and accuracy. [Consistency, Accuracy]
3. Revise data collection procedures to capture accurate job industry categories. [Validity]
4. Validate and standardize date formats across all relevant columns. [Accuracy, Consistency]
5. Cleanse the "default" column from garbage values and collect accurate data. [Accuracy]
6. Standardize state representations in the "state" column as either "New South Wales" or "NSW" for consistency. [Consistency]
7. Investigate and rectify missing values in the "new customer list" and "online_order" columns. [Completeness]
8. Convert float values to integers for property values and relevant columns. [Validity]
9. Update data types to ensure appropriate numeric analysis in the "postcode," "property_valuation," and "past_3_years_bike_related_purchases" columns. [Validity]
10. Provide appropriate titles or labels for unlabeled columns. [Consistency]

By implementing these recommendations, we will enhance the **accuracy, completeness, consistency, and validity of the dataset**, resulting in more relevant and valuable insights for analytical purposes.

I would like to express my gratitude for your support in resolving these data quality issues. Your collaboration is vital in ensuring the availability of high-quality data, which is fundamental for driving effective decision-making and achieving successful outcomes.

Should you have any further questions or require additional information, please feel free to reach out to me. I am available to discuss these findings and recommendations in more detail. Thank you for your attention to this matter, and I look forward to your guidance on the next steps.

Kind regards,

Yash Sonkhiya

Data Analyst, Forage

yashsonkhiya2195@gmail.com