Subject: Data Quality Review and Recommendations for Sprocket Central Pty Ltd. Datasets.

Dear Client,

I hope this email finds you well.

Following our receipt of the datasets from Sprocket Central Pty Ltd, we have conducted a thorough review of the data quality as part of our initial assessment. We have identified several data quality issues and implemented preliminary fixes to ensure the data is ready for analysis in phase two. Below, we have detailed the issues discovered, the actions taken, and recommendations for mitigating these issues in the future.

Transactions Data

Issues Identified:

1. 'Product First Sold date' column was in the wrong data type.
2. 'List price' column had incorrect data type.
3. 'Transaction date' column had incorrect data type.
4. Some values in the 'Standard Cost' column were in the wrong format.
5. Columns such as 'Online Order', 'Brand', 'Product Line', 'Product Class', 'Product Size', 'Standard Cost', and 'Product First Sold Date' had empty rows.

Actions Taken:

- Converted 'Product First Sold date' to 'Long Date' data type.
- Changed 'List price' column data type to 'Currency'.
- Changed 'Transaction date' column data type to 'Long Date'.
- Corrected values in 'Standard Cost' column to 'Currency' data type.
- Replaced empty rows in the identified columns with 'N/A'.

Customer Address Data

Issues Identified:

1. 'States' column had abbreviated state names.

Actions Taken:

- Changed abbreviated state names to their full names for consistency.

Customer Demographic Data

Issues Identified:

1. 'Default' column contained irrelevant information.
2. Empty cells in the 'Last Name' column.
3. Inconsistent and abbreviated entries in the 'Gender' column.

4. Duplicate entries in 'First Name' and 'Last Name'.
5. Empty rows and incorrect format in the 'DOB' column.
6. Empty rows in the 'Job Title' column.
7. Empty cells in 'Job Industry Category' column.
8. Blanks in the 'Tenure' column.
9. Abbreviations in the 'Deceased Indicator' column.

Actions Taken:

- Removed 'Default' column.
- Replaced empty cells in 'Last Name' with 'N/A'.
- Corrected spellings and replaced abbreviations in the 'Gender' column.
- Made minor adjustments to handle duplicate entries in 'First Name' and 'Last Name.'
- Replaced empty rows in 'DOB' with 'N/A', corrected a value, and changed the datatype to 'Long Date'.
- Replaced empty rows in 'Job Title' with 'N/A'.
- Replaced empty cells in 'Job Industry Category' with 'N/A'.
- Replaced blanks in 'Tenure' with 'N/A'.
- Changed abbreviations in 'Deceased Indicator' to full words.

New Customers Data

Issues Identified:

1. Empty cells in the 'Last Name' column.
2. Abbreviations in the 'Gender' column.
3. Incorrect data type in 'Past 3 years Bike Related Purchases' column.
4. Empty cells and incorrect format in 'DOB' column.
5. Blanks in the 'Job Title' column.
6. Abbreviations in the 'Deceased Indicator' column.
7. Incorrect data type in the 'Postcode' column.
8. Inconsistent state names in the 'States' column.
9. Incorrect data type in the 'Property Valuation' column.
10. Extra irrelevant columns.

Actions Taken:

- Replaced empty cells in 'Last Name' with 'N/A'.
- Replaced abbreviations in 'Gender' column with full words.
- Corrected 'Past 3 years Bike Related Purchases' column to 'Number' data type.
- Replaced blanks in 'DOB' with 'N/A', created a new column with correct format, and replaced the old one.
- Replaced blanks in 'Job Title' with 'N/A'.
- Removed abbreviations in 'Deceased Indicator' column.
- Changed 'Postcode' column data type to 'Number'.
- Corrected state names in the 'States' column.
- Changed 'Property Valuation' column data type.

- Removed irrelevant columns and adjusted data types for 'Rank' and 'Value' columns to 'Numbers'.

Recommendations for Future Data Quality Assurance:

1. **Standardization:** Ensure all data entries follow a standardized format, especially for dates, currencies, and categorical variables.
2. **Validation Rules:** Implement validation rules at the point of data entry to prevent incorrect or inconsistent data.
3. **Regular Audits:** Conduct regular data audits to identify and rectify any emerging data quality issues promptly.
4. **Documentation:** Maintain detailed documentation of data entry guidelines and quality control processes.
5. **Training:** Provide ongoing training for staff involved in data entry to ensure they understand the importance of data quality and adhere to best practices.

We believe these steps will significantly enhance the quality and reliability of your data for future analyses. If you have any questions or require further assistance, please do not hesitate to contact us.

Thank you for your attention to this matter.

Best regards,

Shashank S.