Subject: Data Quality Assessment and Strategies for Enhanced Analysis.

Dear Sprocket Central Pty Ltd Team,

I trust this message finds you well. We have conducted a comprehensive data quality assessment of the datasets provided, and we are writing to address several key issues that require immediate attention to ensure the accuracy and reliability of the data for our upcoming analysis in phase two. Our assessment has identified the following data quality issues:

Initial records received for Customer Demographics,Customer Addresses, Transactions are 3,998, 3,999, 20, 000 records respectively.

The customer IDs don’t tally int the datasets since there are additional values in Customer Address and Transactions.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset | Accuracy | Completeness | Consistency | Currency | Relevancy | Validity | Uniqueness |
| Customer Demographic |  | Blanks in last name, DOB, job title, job industry category | Inconsistent gender values | DOB lacked currency ”1843-12-21”, deceased | Default lacked relevance | DOB format |  |
| Customer Address |  |  | Inconsistent State values |  |  |  |  |
| Transactions |  | Blanks in online order, brand |  |  |  | product first sold date in wrong format |  |

**Customer Demographic Dataset Mitigation**

* If the blank values present is a small percentage of the dataset and the missing data does not significantly affect the overall analysis, consider removing the incomplete records.
* Choose a standardized format for representing gender, such as "Female" and "Male" or "F" and "M." You may want to convert all values to a consistent format to ensure uniformity and clarity and also create data validation rules to restrict the input options for the gender field.
* Consider removing or correcting data points that fall outside a reasonable range like “1843-12-21” and utilize regular expressions or specific formatting tools to ensure that all DOB entries adhere to the specified date format.
* Regularly monitor the data for any new missing values and establish protocols to handle them promptly, ensuring ongoing data integrity and completeness on the ‘n/a’ entries.

**Customer Addresses Dataset Mitigation**

* Choose a standardized format for representing the states such as NSW, VIC, QLD. Convert all values to a consistent format to ensure uniformity and clarity. Also, create data validation rules to restrict the input options.

**Transactions Dataset Mitigation.**

* filter out blanks if it’s a small percentage of the general dataset.
* Make sure that the format of the fields especially dates and currency are in order.

To mitigate these data quality concerns, we have implemented comprehensive data cleaning and standardization processes, ensuring that the datasets are prepared for robust analysis in the next phase. We remain committed to delivering the highest standards of data quality and integrity for our analysis.

Please feel free to reach out if you have any questions or require further clarification on these data quality enhancements. We look forward to collaborating with you to achieve the best outcomes for Sprocket Central Pty Ltd.

Thank you for your attention to this matter.

Best regards,

Eunice Okiya.

Junior consultant.

KPMG’s Lighthouse & Innovation Team.