Subject: <u>Strategies to Improve Data Quality for Sprocket Central Pty Ltd</u>

Dear Sir/Ma'am

I hope this email finds you well. We at KPMG's Analytics, Information & Modelling team appreciate your interest in our expertise and are excited to help Sprocket Central Pty Ltd grow its business by effectively analysing your datasets.

During our preliminary data exploration, we have identified some data quality issues that need to be addressed to ensure accurate and reliable analysis. As per our discussion, the four datasets provided by you namely **Customer Demographic**, **Customer Addresses**, **Transactions** and **New Customers** data are essential for analysing and optimising your marketing strategy. However, data quality is critical to derive meaningful insights and drive business growth.

To help you mitigate these issues, we have identified some strategies based on the Data Quality Framework Table and other resources which we believe will improve the quality of your datasets. Below are the issues we have identified and the corresponding strategies to mitigate them:

Data Quality Issues Summary				
Data Quality Framework	**Customer Demographic**	**Customer Address**	**Transactions**	**New Customer**
Accuracy	No Issues	No Issues	No Issues	No Issues
Completeness	Missing Values in columns last_name, job_title, job_industry, tenure	No Issues	• Missing Values in columns online_order, brand, product_line, product_class, product_size, standard_cost, product_first_sold_date • 3500 out of 4000 customers transactions available (87%)	Missing Values in columns last_name, DOB, job_title, job_industry
Consistency	Inconsitent column format gender	Inconsitent column format state	Inconsitent column format product_first_sold_date	No Issues
Timeliness	No Issues	No Issues	No Issues	No Issues
Relevancy	Corrupted column default	No Issues	Cancelled orders need to be filter out	Irrelevant unnamed columns
Uniqueness	100% unique customer_id	100% unique customer_id	100% unique transaction_id	100% unique
Validity	No Issues	No Issues	Inconsitent column product_first_sold_date	No Issues

In the Customer Demographic and Customer Address datasets, missing values are present in some columns and inconsistent column formats are observed. In the Transactions dataset, missing values, inconsistent column formats and irrelevant columns are present. Moreover, canceled orders need to be filtered out. These are some of the issues that were encountered in primary data exploration, but this list is not exhaustive.

Below are some suggestions as per Data Quality Framework to avoid the occurrence of data quality issues:

1. Accuracy: The data must be accurate i.e. it should reflect the correct values without any errors. Inaccurate data can lead to incorrect analysis and hinder business decisions.

- Strategy: We suggest conducting regular data checks and implementing validation rules to ensure data accuracy. Data cleansing and standardisation techniques can also help in improving the accuracy of your datasets.

2. Completeness: The data must be complete i.e. all relevant data fields should be filled and there should be no missing data. Incomplete data can lead to biased analysis and inaccurate insights.

- Strategy: We recommend conducting a thorough analysis of your datasets to identify missing values and filling them in. You can also set up data entry rules to ensure complete data is captured at the time of entry.

3. Consistency: The data must be consistent i.e. the same data element should have the same value across all datasets. Inconsistent data can lead to incorrect analysis and wrong insights.

- Strategy: We suggest implementing standardisation rules for data entry which will ensure consistent values for the same data element across all datasets. Regular data checks can also help identify inconsistencies and rectify them.

4. Timeliness: The data must be up-to-date i.e. it should reflect the current status of the business. Outdated data can lead to incorrect analysis and wrong insights.

- Strategy: We recommend setting up a data governance framework to ensure regular updates of your datasets. You can also use automated tools to collect real-time data which will ensure timely and accurate analysis.

5. Validity: The data must be valid i.e. it should conform to the defined rules and standards. Invalid data can lead to incorrect analysis and hinder business decisions.

- Strategy: We suggest implementing validation rules for data entry and regularly checking for conformity to defined rules and standards.

We believe implementing these strategies will help improve the quality of your datasets and enable us to provide you with accurate and reliable insights. We look forward to working with you to optimise your marketing strategy and drive business growth.

If you have any questions or concerns, please do not hesitate to reach out to us.

Thanks & Regards

Manish Kumar

KPMG's Analytics, Information & Modelling team