

To Whom It May Concern,

We thank you for providing the datasets for Sprocket Central Pty Ltd. The following statistics reflect the data as received by our team, please reach out if there is a misunderstanding:

Dataset	No. of Records	Distinct Customer IDs	Date Received
Customer Demographic	4,000	4,000	August 24, 2020
Customer Addresses	3,999	3,999	August 24, 2020
Transaction Data	20,000	3,494	August 24, 2020

Upon receiving the datasets, we went ahead and evaluated them as per standard data quality dimensions to identify any data quality issues. We have also identified certain strategies that shall help in mitigating these issues in the future that shall help in future analyses and growth of the company. The following table summarizes the various fields having issues that came to our notice:

	Customer Demographic	Customer Addresses	Transaction Data
Accuracy	<ul style="list-style-type: none">• DOB: Inaccuracy• Job Industry Category: Misspelling		
Completeness	<ul style="list-style-type: none">• DOB: Blanks• Job Title: Blanks• Job Industry Category: Blanks• Tenure: Blanks• Customer IDs: Not in Synchronisation	<ul style="list-style-type: none">• Customer IDs: Not in Synchronisation	<ul style="list-style-type: none">• Standard Cost: Blanks• Brand: Blanks• Product Line: Blanks• Product Class: Blanks• Product Size: Blanks• Product First Sold Date: Blanks• Customer IDs: Not in Synchronization
Consistency	<ul style="list-style-type: none">• Gender: Inconsistency	<ul style="list-style-type: none">• States: Inconsistency	
Currency	<ul style="list-style-type: none">• Deceased Customers: Filter Out		
Relevancy	<ul style="list-style-type: none">• Default: Exclude Field		<ul style="list-style-type: none">• Order Status: Exclude Cancelled
Validity			<ul style="list-style-type: none">• Product First Sold Date: Format

Detailed descriptions of the discovered data quality issues and the steps taken to mitigate these issues in later phases of our analysis is given in the following sections of this mail. Explanations and strategies to avoid these issues have been also included. The following recommendations will improve the quality of data used to take important future business decisions for Sprocket Central Pty Ltd:

- **Accuracy Issues**

Issues: ‘DOB’ field in the Customer Demographic dataset contains inaccurate values; the ‘Job Industry Category’ field contains misspelled category values.

Mitigations: Filtered outliers in the dataset. Fixed misspelled categories.

Recommendations: Include additional fields e.g. 'Age' in Customer Demographic dataset and 'Profit' in Transactions dataset for easier detection of outliers.

- **Completeness Issues**

Issues: 'DOB', 'Job Title', 'Job Industry Category', and 'Tenure' fields in the Customer Demographic dataset contain missing values; 'Standard Cost', 'Brand', 'Product Line', 'Product Class', 'Product Size' and 'Product First Sold Date' fields in the Transactions dataset contain missing values for a number of transactions.

Mitigation: Datasets were not in synchronization with join keys, 'Customer IDs' found to be inconsistent among datasets. Filtered out records containing incomplete data.

Recommendation: Ensure datasets are complete and updated regularly.

- **Consistency Issues**

Issues: 'Gender' field in the Customer Demographics dataset contains acronyms and misspellings of the major categorical values; the 'States' field in the Customer Addresses contains multiple versions to identify the same categories; Additional 'Customer IDs' in the Transactions and Customer Addresses datasets indicate missing records in the Customer Demographics dataset.

Mitigation: Replaced extended values and acronyms in the necessary cases using regular expressions. Customers with complete data shall be considered.

Recommendation: Ensure consistency of terminologies across datasets for categorical fields. Regularly run sanity checks on the datasets to reduce trails of human error.

- **Currency Issues**

Issues: The Customer Demographics dataset contains records pertaining to deceased customers.

Mitigation: Filtered out demographic data for deceased customers by checking the 'Deceased Indicator' field.

Recommendation: Can be difficult to track. Update dataset regularly and filter out records for more accurate analysis.

- **Relevancy Issues**

Issues: Incomprehensible 'Default' field in the Customer Demographics dataset; Cancelled orders present in the Transactions dataset which may not be relevant and may lead to inaccuracies and incorrect forecasts.

Mitigation: Dropped the 'Default' field and dropped records pertaining to cancelled orders in the respective datasets.

Recommendation: Delete or format customer metadata into comprehensible formats.

- **Validity Issues**

Issues: 'Product First Sold Date' incorrectly formatted as numeric data, may lead to errors in analysis.

Mitigation: Field converted to Date type.

Recommendation: Format data fields to correct data types and tweak precision of numeric data.

Acting upon the aforementioned recommendations made in our first phase of analysis shall allow Sprocket Central Pty Ltd to make better business decisions, grow revenue, and enhance operational capabilities. KPMG shall continue providing its expertise to provide a holistic view of your data in future phases of the analysis to enable your company to grow further.

Kind regards,
Sudipto Ghosh