Dear Client,

Thank you for providing us with three datasets from Sprocket Central Pyt. Ltd. The summary table below highlights key quality issues that we discovered within the three datasets. Please let us know if you have any queries surrounding the issues presented.

Summary Table:

Dataset	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
Customer Demographic	1.DOB: Inaccurate 2.Age: missing	1.Job titles: Blanks 2.Customer id: incomplete	1.Gender: inconsistency	1.Deceased Customers: Filter out	1.Default Column: deleted	
Customer Address		1. Customer id: incomplete	1.State: inconsistency			
Transactions	1.Profit: missing	1.Customer id: Incomplete 2.Online Order: Blanks 3.Brand: Blanks			1.Cancelled Status order: Filter out	1.List Price: Format 2.product sold date: format

Below is more in-depth description of data quality issues discovered and methods of mitigation used. Recommendations and explanation have also been included to avoid further data quality issues in the future. Following recommendations will improve accuracy of data used to influence business decisions of sprocket central Pvt. Ltd in the future.

Accuracy Issues:

- **DOB was inaccurate for "customer Demographic" and missing an age column; missing a profit column for "Transactions"**

  Mitigation: Filter out outlier in **DOB**.

Completeness:

- **Additional Customer ids where inconsistent among "customer Demographic", "Customer Address" and "Transactions"**
  Mitigation: Filter all customer ids from 1 to 3500
- **Blanks in job title for "customer Demographic" in online order and brand column for "Transactions"**
  Mitigation: Filter out blanks for job title, online orders and brand column.

Consistency:

- **Inconsistency in gender for "customer Demographic" and "Customer Address" respectively.**
  Mitigation: Filter all 'M' under category of 'Male', 'Female' and 'F' under 'Female' for gender. Filter all 'New South Wales' to 'NSW' and 'Victoria' to 'VIC' for states.

Currency:

- **People that are 'Y' in deceased indicator are not current customers for "customer Demographic".**
  Mitigation: Filter out customers checked 'Y' in deceased indicator.

Relevancy:

- **Lack of relevancy or comprehensibility in default column for "customer Demographic" and order status for "Transactions"**
  Mitigation: Deleted metadata in default column and filter out cancelled order status.

Validity:

- **Format of list price, product sale date for "Transactions".**
  Mitigation: Format product sale date to short date format and format list price to currency**.**

That summaries all data quality issues discovered through the first stage of the data quality analysis. The mitigation strategies suggested are simple and effective ways of improving data quality for future analysis. They will not only improve the analysis output that one can perform within the company but will increase the level of analysis that can be performed by KPMG and Data Analysis Team.

Please let us know if you have any questions regarding mitigation or any data quality issues identified.

Regards,

Harshavardhan Borage

(KPMG Data Analytics Team)