Dear Sir,

Thank you for providing us with the three datasets from **Sprocket Central Pty Ltd**. The below table highlights the summary statistics from the three datasets received. Please let us know if the figures are not aligned with your understanding.

	Accuracy	**Completeness**	**Consistency**	**Currency**	**Relevancy**	**Validity**
Customer Demographic	1.DOB: inaccurate Age: missing	1.Job title: blank 2.Customer id: Incomplete	1.Gender: Inconsistency	1.Deceased customer: filterd out	1.Default column: delete	
Transaction	1.Profit: missing	1.Customer id: incomplete 2. Online Order: blanks 3. Brand: blanks				1.Product sold date: format
Customer Address		1.Customer id: incomplete	1.States: inconsistency			

Below are more in-depth descriptions of data quality issues discovered and methods of mitigation used.

Accuracy Issues

- **DOB was inaccurate for ‘‘Customer Demographic’’ and missing an age column: missing a profit column for ‘’ Transaction’ ’missing a full name column**

  Mitigation: Filter outlier in DOB.
  Recommendation: Create an **age column and full name**, allowing for more comprehensible data and easier to check for errors. Create a **profit column** in ‘’**Transactions**’’ to check the accuracy of sales.

  Creating additional columns for age, name, and profit will allow for easier identification of errors. The **profit column** will assist in future monetary analysis.

Completeness

- **Additional customer_ids were inconsistent among ‘‘Customer Demographic’’, ‘‘Customer Address’’ and ‘‘Transaction’’**

  Mitigation: Filter all customer_ids from 1 to 3500
  Recommendation: Maintain data to ensure tables are up to date. According to our model, only **3500 customer_ids** customers are completely recorded that can be used in further analysis.

  The data recorded inside ever spreadsheets may not be in sync, with incomplete data the analysis result would be skewed. We encourage you to cross check spreadsheets or utilize database management.

- **Blanks in job_title for ‘‘Customer Demographic’’ and ‘’New_Customer_List’’, in online_order and brand_column for ‘‘Transactions’’**

  Mitigation: Filter out ‘‘blanks’’ for job_title, online_order, and brand_column.
  Recommendation: Provide drop-down option for **job_title, online_order** and **brand_column**.

  Blanks are treated as incomplete data and can skew further analysis results. The addition of drop-down options will allow more accurate data and will contribute to more dependable and accurate results.

Consistency

- **Inconsistency in gender for ‘‘Customer Demographic’ and ‘‘Customer Address’’ respectively**

  Mitigation: Filter all ‘M’ under category of ‘Male’, filter all ‘Femal’ and ‘F’ under ‘Female’ for **gender**. Filter all ‘New South Wales’ to ‘NSW’ and ‘Victoria’ to ‘VIC’ for **states**.
  Recommendation: Create dropdown options for ‘Male’, and ‘Female’ in **gender**. Create dropdown options for all **state** abbreviations.

  Dropdown options minimize manual entry and human error. Allows for an increase in the consistency of terminology. Be careful when creating options for gender since it is a sensitive topic for analysis.

Currency

- **People that are ‘Y’ in deceased_indicator are not current customers for ‘‘Customer Demographic’’**

  Mitigation: Filter out customers with **‘Y’** tag in **deceased_inducator**.
  Recommendation: When can be sure the condition of deceased customers should update data immediately and accordingly.

  Deceased customers are not current customers, removing them from data will enhance currency of data and will leave more accurate estimates in further analysis.

Relevancy

- **Lack of relevancy or comprehensibility in default_column for ‘‘Customer Demographic’’.**
  Mitigation: Deleted Metadata in default_column.
  Recommendation: Check for incomprehensible Metadata or give an explanation.

  Irrelevant or comprehensible information can skew the analysis.

Validity

- **Format of list_price, product_sale_date for ‘‘Transactions’’**

Mitigation: Format **product_sale_date** to short date format, format **list_price** to currency.
Recommendation: Set up columns so that formats such as price and decimals are already in place when entering new data.

Allowable values can make data to be interpreted more easily. Formatting into the price and allowing for either 2 or 3 decimals placed consistently will increase readability. This would be an important factor for the speed and accuracy of business analysis.

We conducted and summarized the data quality issues from these six perspectives. The mitigation strategies suggested are effective ways of enhancing data quality for future business analysis. These suggestions can make the data completer and more reasonable, which will also help KPMG or hired analytics teams to analyze the data efficiently.

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

Kind regards,
Peiqing Lian