Dear Client,

It is our privilege to help you regarding the provided three data sets from Sprocket Central Pty Ltd.

We encountered the quality issues from the data sets, please refer summary table below.

Please let us know if any queries regarding the same.

Three data sets	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity	Uniqueness
Customer Demographic	• DOB: inaccurate • Age: missing	• Job title: blanks • Customer id: incomplete	• Gender: inconsistency	• Decreased customers: filter out	• Default column: delete		
Customer Address		• Customer id: incomplete	• State: inconsistency				
Transactions	• Profit: missing	• Customer id: incomplete • Online order: Blanks • Brand: Blanks			• Cancelled status order: filter out	• List price: format • Product sold date: format	

Below are more in-depth description of data quality issues discovered and methods of mitigation used. Recommendations and explanations have also been included to avoid further data quality issues in the future. Following recommendations will improve accuracy of data used to influence business decision of Sprocket Central Pty Ltd in the future.

- **Accuracy Issues:**
  DOB was inaccurate for "Customer Demographic" and missing an age column; also missing a profit column for "Transactions".

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

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

- **Completeness:**
  - **Additional customer_ids were inconsistent among "Customer Demographic", "Customer Address", and "Transactions".**

Mitigation: *Filter all customer_ids from 1 to 3500.*
Recommendation: Ensure tables are up to date (from the same time period). For our model, only customer_ids from 1 to 3500 will be used as they have complete data.

The data received may not be in sync across all spreadsheets, with incomplete data the analysis results may be skewed. This is a 'completeness' issue, to prevent future occurrences it is encouraged to cross check spreadsheets and sync data.

- **Blanks in job_title for "Customer Demographic", in online_order and brand_column for "Transactions".**

Mitigation: *Filter out blanks for job_title, online_order and brand_column.*
Recommendation: Simplify job_title to another category such as industry_industry or provide dropdown options for job_title. Provide dropdown options for online_order and brand_column.

Blanks are treated as incomplete data and can skew further analysis results. The addition of dropdown option will allow to have more complete data and will result more accurate analysis.

- **Consistency:**
  Inconsistency in gender for "Customer Demographic" and state for "Customer Address".

  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 option for 'Male', 'Female' and 'U' in gender. Create dropdown option for all state abbreviations.

  Dropdown options minimize manual entry and human error. Allows for increase of consistency of terminology. Gender identity can be a sensitive topic, proceed with caution when creating options.

- **Currency:**
  People that are 'Y' in decreased_indicator are not current customers for "Customer Demographic".

  Mitigation: *Filter out customers checked 'Y' in decreased_indicator.*
  Recommendation: Can be difficult to check for decreased customers, but once this information is received one should update data accordingly.

Decreased customers are not current customers, removing them from data will increase currency of data and will result in more accurate estimates in future analysis.

- **Relevancy:**
  Lack of relevancy or comprehensibility in default_column for "Customer Demographic" and order_status for "Transactions".

  *Mitigation***:** *Deleted Metadata in default_column. Filter out 'Cancelled' order_status.*
  Recommendation: Check for incomprehensible Metadata and delete or format to make comprehensive.

  'Cancelled' order_status is irrelevant information for future analysis, as it can skew data -for example total number of customers per annum will be an overestimate.

- **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 will make data to be interpreted more easily. Formatting into price and allowing for either 2 or 3 decimals placed consistently will increase readability. This will reflect positively on speed and accuracy for business decisions.

That summarises 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 other hired analysis team.

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

Kind regards,
Kaival Shah.