Dear Kathleen,

Thank you for providing me with the three datasets from Sprocket Central Pty Ltd. The table below is a summary of the key quality issues i discovered within the dataset provided. Please let me know if you have any queries about the issues presented.

Summary Table

Datasets	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
Customer Demographic	DOB: inaccurate Age: missing	Job title: blanks Customer id: incomplete	Gender: inconsistent	Deceased customers: filtered out	Default column: deleted	
Customer Address		Customer id: incomplete	State: Inconsistent			
Transactions	Profit: missing	Customer id: incomplete Online order: blanks Brand: blanks				List price: format Product sold date : format

I have listed a more in depth description or data quality issues discovered below and methods of mitigation used. Explanations and recommendations have also been included to avoid further data quality issues in the future. The recommendations provided will help influence data-driven decisions of Sprocket Central Pty Ltd in the future.

Accuracy Issues:
- o **DOB was inaccurate for "Customer Demographic" and age_column was missing; profit_column was missing for "Transaction"**

  *Mitigation: filter out outliers in **DOB***
  Recommendation: create and age_column, it makes it easier to check for error and also make the data easier to read.
  create a profit_column in "Transaction" to check the sales accuracy and easier monetary analysis.

Completeness
- o **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.

  The data received may not be in sync across spreadsheets, with incomplete data, the analysis Results may be skewed. It's strongly recommended to cross check spreadsheet and sync data to Avoid future occurrence.

- o **Blanks in job_title for "Customer Demographic", in online_order and brand_column for "Transaction"**

  *Mitigation: Filter out 'blanks' for **job_title, online_order and brand_column**.*
  Recommendation: create a drop down option for job_title, online_order and brand_column.

  Blanks are treated as incomplete data and it can further skew the analysis result.

Consistency
- o **Inconsistency in gender for "Customer Demographic" and "Customer Address"**

  *Mitigation: Replace all 'M' in the **gender_column** to 'Male', replace all 'femal' and 'F' to 'Female'.*
  *Replace all 'NSW' to 'New South Wales' and all 'VIC' to 'Victoria' in the **state_column.***
  *Recommendation: create dropdown option for 'male', 'femal' and 'U' in **gender_column.***

  Dropdown option helps in data validation; it will also help to minimize manual entry and human error.

Currency
- o **People with 'Y' in deceased_indicator are not current customers for "Customer Demographic"**

  Mitigation: Filter out customers checked 'Y' in deceased_indicator.
  Recommendation: it can be difficult to check for deceased customers, but once this information is received, it should be updated accordingly.

  Deceased customers are not current customers, filtering them out from the data will increase data currency and it will make future analysis more accurate and current.

Relevancy
- o **Lack of relevancy in default_column for "Customer Demographic"**

  Mitigation: Deleted the default_column
  Recommendation: Ensure that all the information entered into the spreadsheet is relevant and comprehensible.

Validity
- o **Format for list_price, product_sale_date for"Transaction"**

  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.

That is the summary of all the data quality discovered in the first stage of data quality assessment. The mitigation strategies suggested are simple and effective ways to improve the data quality for future analysis. They will help improve the analysis that one can perform in the company and also the analysis that can be performed by KPMG and other hired analysis teams.

Kindly reach out if you have any question about the mitigation or any data quality issues.

Kind regards,
Nofisat Adegbenro.