

Subject: Assessment of Data Quality and Recommendations for Sprocket Central Pty Ltd

Dear Manager,

I hope this email finds you well. Thank you for providing the datasets from Sprocket Central Pty Ltd. Our team is excited to work on assessing the quality of the data you've shared. As part of this process, we will also be drafting recommendations to address any issues identified and ensure the data is ready for our upcoming analysis in phase two.

Upon reviewing the datasets—Customer Demographic, Customer Addresses, and Transaction data from the past three months—we have identified several data quality issues that may impact the accuracy and reliability of our analysis. I would like to share these findings with you and propose strategies to mitigate these concerns.

Transaction Data:

- Productid column: Some entries have a product id "0", and there is a mix of brand, product line, product class, for product ids 1, 2, 4, 5, etc.
- transaction date: column should be in date format.
- online order: Blank values present.
- brand, product line, product class, product size: Blank values detected.
- list price, standard cost: Different values for the same product.
- Product first sold date: Varied formats and blank values.
- Remove the first row.

New Customer List:

- Remove the first row.
- past_3_years_bike_related_purchases: Incorrect data format.
- DOB: Incorrect format and blank values.
- job title: Blank values.
- Postcode, property valuation: Incorrect format.
- Extra column With no name

Customer Demographic:

- Gender: Variations in entries (e.g., "f," "Female," "Femal," "M," "Male").
- DOB: Blank entries and anomalies (e.g., 1843).
- job title: Blank values.
- Default: Unrelated values.

Customer Address:

- Address: Some entries have a '0' in front of the address.\

To mitigate these issues and ensure the data's quality for our analysis, we recommend the following actions:

- Standardize data formats, especially for dates (e.g., transaction date, DOB).
- Fill or correct missing or incorrect values in categorical variables (e.g., gender, job title).
- Remove rows with anomalies or unrelated values.
- Validate and correct entries with formatting errors (e.g., productid, past_3_years_bike_related_purchases).
- Regularize address entries without leading zeros.

Please find attached a Data Quality Framework Table for your reference, which includes criteria and dimensions considered during the assessment. If you have any questions or require further clarification, feel free to reach out. We look forward to collaborating with you to ensure the data meets the highest standards for our analysis.

Thank you for your attention to this matter.

Best regards,

Devanapalli Ruthwik,

Data Analyst

KPMG