Hello,

The Manager,

Sprocket Central Pty Ltd.

Trust this mail finds you well,

Thank you for contacting Tony Smith(partner) in KPMG's Lighthouse and Innovation Team and providing us with the three datasets from your organization. A summary table was created to highlight key quality issues that we discovered within the three data sets provided. Do not hesitate to call our attention if you have any oppositions regarding the issues outlined.

SUMMARY TABLE

	Accuracy	Completeness	Consistency	Validity	Relevancy	Currency
Customer Demographic	Inaccurate D.O.B Missing Age	Blank Job Titles Incomplete Customer ID	Gender Inconsistency		Deletion of default column	Deceased customers Filtered Out
Customer Address		Incomplete Customer ID	State Inconsistency			
Transactions	Missing Profit	Incomplete Customer ID Blank Online Order Blank Brand		Formatting of List price and Product sold date	Cancelled Status Order Filtered Out	

According to the above summary table drafted, it is seen that the data quality is lagging and needs adjustment for better analysis that will improve your organization sales. Below are more in-depth descriptions of data quality issues discovered and methods of mitigation used. This approach will help avoid further data quality issues in future.

ACCURACY ISSUES

- **Inaccuracy of D.O.B for the "customer demographic" and the absence of age column; profit column for "Transactions".**

    The inaccurate input of D.O.B can be filtered out to maintain accuracy for the "customer demographic" dataset, while the creation of a new column for profit and age in the "Transactions" dataset help to check sales accuracy. This ill also enable future monetary analysis.

COMPLETENESS ISSUES

- **Additional customer_ids were inconsistent among "Customer Demographic", "Customer Address" and "Transactions". This makes the column for customer_id incomplete.**

    The customer_id column is filtered from 1-3500, this is due to the data completeness for the customers in the range of the IDs. Data incompleteness can affect its analysis, this can be avoided by cross checking and syncing data across spreadsheets before been sent out for analysis.

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

    The blanks are filtered out for the columns and a dropdown list is created for each column to streamline data been inserted to ensure completeness of data in the spreadsheet. The presence or absence of blanks affect the accuracy if the dataset analysis.

CONSISTENCY ISSUES

- **Inconsistency in gender for "Customer Demographic" and "Customer Address" respectively.**

    All inconsistent gender inputs are filtered out, such that "M" is filtered into "Male", while "F" and "Femal" are filtered into "Female" categories. Filter all "New South Wales" to "NS" and "Victoria| to "VIC" for states.
    A dropdown option can be created for bot gender and states categories to minimize manual entry and human errors.

VALIDITY ISSUES

- **Format of List_price, Product_sale_date for "Transactions".**

List_price is formatted to currency and Product_sold_date is formatted to short date format. This will enable consistent and readable input data that is easily interpreted, therefore, an increment in speed and accuracy of future analysis for profitable business decisions.

RELEVANCY ISSUES

- **Lack of relevancy in default_column for "Customer Demographic" and Order_status for "Transactions".**

Metadata in Default_column is to be deleted and "cancelled" in Order_status is filtered out. "Cancelled" Order_status is an irrelevant information and may cause inaccuracy in future analysis.

CURRENCY ISSUES

- **People that are "Y" in Deceased-indicator are not current customers for "Customer Demographic".**

Filter out customers marked "Y" in the Deceased_Indicator, deceased customers are not current customers, removing them from data will increase data currency and accuracy of future analysis.

This is a summary of all data quality issues discovered through the first phase of the data quality analysis. Data quality can be improved effectively for future analysis if the suggested simple mitigation strategies are adhered to. This will improve the analysis to be done within the organization, by KPMG and other hired analysis teams.

Please let us know if you have questions concerning the data quality issues identified and the mitigation strategies been offered.

Kind regards,

Obodo Favour Onyebuchi.