

Dear John Doe,

Thank you for providing us with the three datasets from Sprocket Central Pty. Ltd. The key quality issues that we discovered within the three datasets has been highlighted in the summary table below. Please send us a word if you have any queries regarding the issues presented.

	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
Customer Demographic	1)DOB: -inaccurate 2)age: -missing	1)Job title -blanks 2)Customer id: -incomplete	1)Gender inconsistency	1)Deceased customers filtered out	1)Default column : -delete	
Customer Address		1) Customer id incomplete	1)States: inconsistency			
Transactions	1)Profit missing	1) Customer id incomplete 2)Online order:Blanks 3)Brand:blanks			1)Cancelle d status order: -filtered out	1)List Price: -format 2)Product sale Date -format

Below are more in depth descriptions of data quality issues discovered and methods of mitigation used. Recommendations and explanations have also been elucidated to avoid such data quality issues in future.

The total no of records missing is mentioned in the tables below:-

Customer Demographic (Total records 4000)

FIELD NAME	ERRORS
<u>DOB</u>	01 record 1843 87 records Blanks
<u>last_name</u>	125 records Blanks
<u>Gender</u>	88 records gender 'U' Values are not consistence M, Male, F, Female, Femal, U
<u>job_title</u>	506 records Blanks
<u>job_industry</u>	656 records mention 'N/A'
<u>Default</u>	3317 records value 'special characters' includes null and Blanks
<u>Tenure</u>	87 records Blanks

Transactions (Total records 20000 -past 3months)

FIELD NAME	ERRORS
<u>Online_order</u>	94 records Blanks
<u>brand</u>	48 records Blanks
<u>product_line</u>	48 records Blanks
<u>product_class</u>	48 records Blanks
<u>product_size</u>	48 records Blanks
<u>standard_cost</u>	48 records Blanks
<u>product_first_sold_date</u>	48 records Blanks

ACCURACY ISSUES

- **DOB was inaccurate for “Customer Demographic” and missing an age column;Missing a profit column in Transactions.**

Mitigation:- filter out DOB

***Recommendation:-create an age column allowing for more comprehensible data and easier to check for errors.Check profit_column for future accuracy of sales
The profit column will assist in future financial analysis.***

COMPLETENESS

- **Additional customer_ids were inconsistent among “Customer Demographic”, “Customer Address”, “Transactions”.**

Mitigation: -Filter out all customer_ids from 1 to 3500.

Recommendation: -Ensure tables are up to date from the same period.

The data received may not be in sync across all spreadsheets. With incomplete data the analysis results would be deviated and inaccurate. This is a completeness issue so to prevent future occurrences it is advised to check across spreadsheets and sync data.

CONSISTENCY

- **Inconsistency in gender for “customer demographic”, in online_order and brand_column for transactions.**

Mitigation: -Filter all categories of ‘M’ under ‘MALE’, ‘F’ and ‘FEMAL’ under ‘FEMALE’. Filter all New South Wales to ‘NSW’, and Victoria to ‘VIC’ for states.

Recommendation: -Create dropdown options for ‘Male’, ‘Femal’, ‘U’ in Gender. Create dropdown option for all states.

Drop down entry minimizes human error.

CURRENCY

- **People that are “Y” in the deceased column are not part of the “customer demographic”.**

Mitigation: - filter out customers checked ‘Y’ in the deceased column.

Recommendation: -Can be difficult to check for deceased customers but one should update the data accordingly.

Deceased customers are not current customers. Removing them would increase the currency of data.

RELEVANCY

- **Lack of relevance in default column for “Customer demographic” and “order_status” for Transactions.**

Mitigation:-Deleted metadata in default column;filtered out “cancelled out” order status.

Recommendation:-Check for incomprehensible metadata and format to make it comprehensible.

“Cancelled order” is irrelevant for the future.

VALIDITY

- **Format of list_price , product_sale_date for “Transactions”**

Mitigation:- Format product_sale _data to short date format.Format list_price to currency.

Recommendation:-set up new columns so that when new data is entered the formats such as price and decimals are already in place.

The above content summarises all the data quality issues discovered in the 1st stage of data quality analysis and methods to mitigate such issues.The mitigation method suggested are simple and practical ways for enhancing the quality of data for future analysis.This will not only improve the analysis output that one can perform within the company but also increase the level of analysis that can be performed by KPMG.

Kindly ,let us know if you have any issues concerning mitigation or any data quality issues identified.

Kind Regards,
Aishani Aarya