

To whomsoever it may concern,

Thank you for sending across the dataset for us to perform some analysis to improve the business of Sprocket Central. The following table represents the data as received by our team. Please reach out if there are any misconceptions regarding the data.

DATASET	SHAPE	DISTINCT CUSTOMER ID
Transactions	(20000, 13)	3494
NewCustomerList	(1000, 18)	-
CustomerDemographic	(4000, 13)	4000
CustomerAddress	(3999, 6)	3999

After obtaining the dataset, we conducted an analysis and obtained the results to identify any data quality issues.

#### **Data Quality issues**

##### **→ Completeness of the data: Null values**

Transactions- online\_order, brand, product\_line, product\_class, product\_size, standard\_cost, product\_first\_sold\_date are the columns with missing values.

NewCustomerList- last\_name, DOB, job\_title, job\_industry\_category

CustomerDemographic- last\_name, DOB, job\_title, job\_industry\_category, tenure

##### **→ Relevancy Check**

NewCustomerList

1)Gender- Deleted rows with the value 'U' since its irrelevant (Row values removed - 17)

CustomerDemographic

1)Gender- Deleted rows with the value 'U' since its irrelevant (Row values removed - 88)

2)deceased\_indicator - deleted rows with the values 'Y' since we need the data of only the customers who are alive.

3)Deleted the column 'default' since it is not meaningful

##### **→ Consistency**

CustomerDemographic- Changed the values of F and Femal to 'Female' and M to 'Male'.

Updated column has the values of Female,Male

CustomerAddress - Changed the values of 'New South Wales' to 'NSW' and Victoria to 'VIC'

##### **→ Accuracy**

Transactions- In the column 'product\_first\_sold\_date' the date was in the form of numbers. This was formatted back to the form of a date. But this column seems to be wrong since it shows the same date with different timestamps.This looks inaccurate.

**Recommendations:**

- 1) If the datasets are updated regularly the completeness of the data would not be a major problem. Performing sanity checks would help in avoiding any human errors.
- 2) Format data types for better numerical accuracy.

We believe that addressing the quality control issues outlined above will assist Sprocket Pvt Ltd in expanding their operations and delivering improved outcomes. KPMG will continue to offer our expertise in the ongoing data cleaning and preprocessing efforts for use in predictive analytics. Please do not hesitate to reach out to us if you have any inquiries regarding the same.

Kind regards,  
Malavika Venkatesh