

Dear client,

Thank you for providing us with the three datasets from Sprocket Central Pty Ltd. The below observations are drawn from data which are to be rectified to further analysis by data preprocessing, feature scaling if it requires.

Table Name	No. of records	Distinct Customer IDs	Date Data Received
Customer Demographic	4000	4000	27/07/2023
Customer Address	3999	3999	27/07/2023
Transaction Data	20,000	3494	27/07/2023

Observations in data:

1. **Transactions** data contains many null values like in *online_order*, *product_line*, *product_class*, *product_size*, *standard_cost* and *product_first_sold_date*. Dropping these rows cause data loss for further analysis.
2. **Transactions** data contains *transaction_date* from 01/01/2017 to 30/12/2017.
3. **CustomerDemographic** data contains inconsistent *default* column default. Inconsistent data not good for analysis we have to drop. And it also contains null values we have to handle them to reduce data loss.
4. **CustomerAddress** dataset already quite good, we can proceed to further analysis.
5. **NewCustomerList** dataset's *gender* column has value as 'U', it represents unknown gender, we can represent in better way to understand like 'unknown'.
6. There are no outliers for all these datasets which saves time from preprocessing data.

“ Moving forward to continue with the data cleaning, standardisation and transformation process for the purpose of model analysis.”

Regards,
Dobbali Ravi Kumar