Dear [Sprocket Central Pty Ltd],

Thank you for providing us with three datasets. The below table highlights summary of our finding.  
Please let us know if the figure are not corrected with your records.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table Name** | **No. of records** | **No. of feature** | **Date Recieved** |
| CustomerAddress | 3999 | 6 | 4/10/2022 |
| CustomerDemographic | 4,000 | 13 | 4/10/2022 |
| NewCustomerList | 1,000 | 23 | 4/10/2022 |
| Transactions | 20,000 | 13 | 4/10/2022 |

We also found data quality issue with missing data and incorrect data type

* **customer\_ids not found in CustomerDemographic table.**

Data will not synced with each other which will affect our analysis.

* **Multiple missing value in records.**There are less than 1% of transactions missing but they can be remove with minimal effect for training dataset.
* **Incorrect data type.**Such as date is not in datetime format.
* **Unexplained feature**A lot of unnamed feature in NewCustomerList table.  
  There is default feature in CustomerDemographic which is not clear.

We will continue with data cleaning, standardization and transformation process for model analysis.  
The process will be documented. We are hoping to spend some time with your data SME to ensure that all assumption are in line with your understanding.

Kind regards,  
[Name]