Dear Client,

Thank you for providing all the three datasets and from the given data the following are the inferences.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table Name** | **No. of. records** | **Distinct Customer Ids** | **Date Data Record** |
| Customer Demographics | 4000 records | All are distinct ids | Date the data has been recorded is not provided |
| Customer Address | 3999 records | All ids are unique | Few records are not present in Customer Demographics |
| Transaction Details | 20000records | All transactions have unique ids but each customer has multiple transactions | All the transaction data was recorded on the same date |

There are notable issues in the tables that are listed below.

1. There are additional ‘customer\_id’s in the ‘Transactions’ table and ‘customer\_id’ s in ‘CustomerAddress’ table which are not present in ‘CustomerDemographic’ table ‘customer\_id’ column.
2. The ‘gender’ column in ‘CustomerDemographic’ table has different abbreviations for ‘Female’ and ‘Male’. Same abbreviations need to be maintained for the consistency of the data.
3. In the ‘CustomerAddress’ table abbreviations for ‘state’ should be consistent. Ex: ‘Victoria’, ‘Vic’ were given for same state which can be either given as ‘Victoria’ or ‘VIC’ not both can be maintained.
4. Proper data types for datetime are not maintained in ‘Transactions’ table for ‘product\_first\_sold\_date’ column.
5. ‘online\_order’ can be given either ‘0’ or ‘1’ which can be an integer. Float data type has been given to this column in ‘Transactions’ table.
6. Columns like ‘brand’, ‘product\_line’, ‘product\_class’, ‘product\_size’, ‘standard\_cost’ are ‘Nan’ or has no data.

Moving as a team we start working on the dataset for further process like data cleaning, standardization, drawing insights.

Thank you,

XYZ