Dear Customer

We received your data and have here some highlights are being listed about them after running a Quality Assessment check. Kindly go through the feedback about this data and we have also listed some techniques which can help to reduce the errors and shortcomings in the datasets, hence increasing the analysis accuracy.

|  |  |  |
| --- | --- | --- |
| **File Name** | **Number of Records** | **Distinct Records** |
| Transactions | 20000 | 20000 |
| NewCustomerList | 1000 | 1000 |
| CustomerDemographics | 4000 | 4000 |
| CustomerAddress | 3999 | 3999 |

* **CustomerAddress has some extra ‘customer\_id’ which are absent in CustomerDemographics table.**

The data collected is not from the same timeline and is inconsistent.

*Mitigation: Please check the data entry error and check if data tables are from same timeline.*

* **Missing values.**

Except the NewCustomerList table, all tables had missing values in several columns, so we had to assume some systematic values for these missing fields.

*Mitigation: Look into data entry errors and if some information is missing, delete that instance from all tables.*

* **Inconsistent values for some attributes.**

Like at some places ‘Female’ is typed as ‘F’ and at others as ‘U’. This leads the system to treat them differently and hence may cause errors in result.

*Mitigation: Keep some consistent values as options to choose from for the customer.*

* **CustomerAddress has total data of 3999 customers, while CustomerDemographics has data of 4000 customers.**
* **Transactions table has data of only 3494 unique customers.**

This may lead to under performance of our final model, since the behaviour of all diverse customers can’t be analysed, also this may indicate that these customers were inactive during the timeline under consideration.

* **In some attributes there was inconsistency.**

Like ‘default’ column in Customer Demographics had some random string values as input. Also some columns have both string and integer as inputs.

*Mitigation: Follow some standard for data entry e.g. a fixed format for entering data-time fields.*

Kindly adhere to these guidelines, so that we can continue serving to your best interests in future. Feel free to contact the under mentioned for any further assistance.

Regards,

Salil Manocha (Analyst)