Dear Client,

Thank you for providing Datasets : NewCustomerList, CustomerDemographic and Transactions Dataset. We have identified some gaps in the database in terms of missing data, descriptions, field values which is explained as below.

NewCustomerList Dataset: Shows irrevent column fields numbered from 17 to 21 which has no significance/meaning. Column description can not be numbers.

CustomerDemographic Dataset :Shows irrevant data with column description as default. It gives no meaning.

Transactions Dataset: Shows total 101 unique entries for product\_id. For every transation\_id it has to be different.

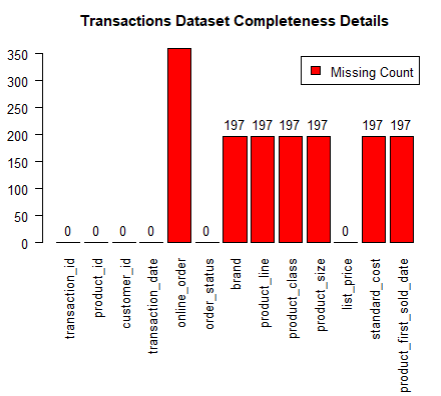
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Accuracy | Completeness | Consistency | Currency | Relevancy | Validity | Uniqueness |
|  | Correct Values | Total Data Fields with missing values | Values free from Contradiction | Values Up to Date | Data Items with Value MetaData | Allowable Values | Duplicate Data |
| Transactions |  | Details are mentioned below with visuals as there are many fields which are missing | Product\_id seems to have duplicate entries. Each Transaction id should have unique Product\_id value |  |  |  | Product\_id seems to have duplicate entries. Each Transaction id should have unique Product\_id value |
| NewCustomerList | Incorrect DOB |  |  | Incorrect DOB | Incorrect DOB |  |
| Customer  Demographic | Incorrect DOB  column with Default as name has no meaning | Default column as field name is confusing for understanding purpose |  | Incorrect DOB  Default column doesn’t reveal any understanding  column with Default as name has no meaning | Incorrect DOB  Default column? |  |
| CustomerAddress |  |  |  |  |  |  |

1. Accuracy
   1. Date of birth data is not accurate there are some outliers
      1. Mitigation: remove outliers
      2. Suggestions: outliers can be removed as it has very less impact as of now on volume of data
2. Completeness
   1. There is missing data in various fields of datasets
      1. Mitigation: apply missing value treatments, they can be either removed depending on its impact in analysis and they can be filled with mean values or neighborhood matchings
3. Consistency
   1. Inconsistency product\_id data in Transactions
      1. Mitigation: replace data to remove inconsistency or assuming product id for particular brand is same among all products
4. Validity
   1. There are some records with product\_id = 0
      1. Mitigation: check database whether there is 0 product\_id
   2. NewCustomerList has some columns named as 17,18,19,20,21
      1. Mitigation check database for correct name field
   3. Default column in CustomerDemographic
      1. It doesn’t reveal any understanding. More details are required for what its entries represent.

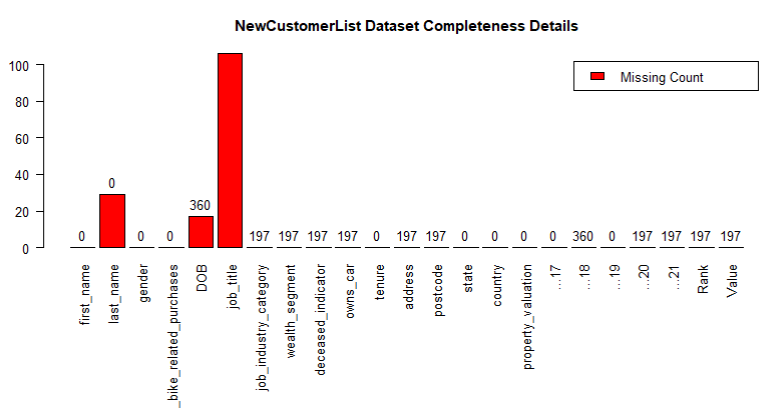
Overall data details are attached herewith for database entries.

**Completeness Details**

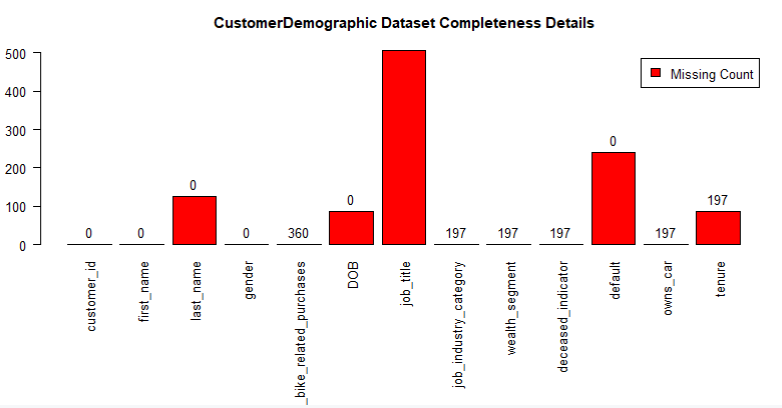
Transaction:



NewCustomerList :



CustomerDemographic



CustomerAddress

seems complete

