

Dear Client,

As the data received, we have 4 sheets; CustomerDemographic, CustomerAddress, Transactions, NewCustomerList. After we explore the data, we have recommendations as below.

For CustomerDemographic, we found 6 fields should be fixed.

Field	Issues	Action
gender	Have 6 value; F, Femal, Female, M, Male, U which may have the same meaning	Define into 3 values; Female, Male, U
DOB	customer_id=34 may have incorrect DOB on 1843-12-21 and some data are missing	Verify this row if there is mistaken in gathering data process
job_title	Have missing values	Change missing to "n/a" or "unidentified" to make it has more meaningful
default	Have unreadable data	Investigate the source data, does it collect in the proper format
owns_car	Data does not consistent as other field	Change "Yes/No" to "Y/N"
tenure	Have missing values but is the data really missing or it is zero?	Verify this field if missing row, fill as "n/a". If it is zero, fill with 0.

For CustomerAddress, we found 2 fields should be fixed.

Field	Issues	Action
customer_id	There is no id 3, 10, 22, 23	Verify on these id
state	Some states are in the full name	Change to abbreviation

For Transactions, we found 9 fields should be fixed.

Field	Issues	Action
product_id	There is product id 0, which should start at 1. For [standard_cost] field, each product id usually has the same cost but product id 0 have varied cost	Verify on id 0, is it the same item or it may be missing value
Online_order	Have 17 missing values	Recheck the source data
brand, product_line, product_class, product_size, standard_cost, product_first_sold_date	Have 197 missing values in the same row	Check the product which have these missing, may have no description
product_first_sold_date	Does not in the date format	Change to date format

For NewCustomerList, we found 10 fields should be fixed.

Field	Issues	Action
past_3_years_bike_related_purchases	This field store in text format	Change data type to number
DOB	Does not be collected in the same format as [DOB] in Table CustomerDemographic (DD/MM/YYYY)	Change format to DD/MM/YYYY
job_title	Have missing values	Change missing to “n/a” or “unidentified” to make it has more meaningful
owns_car	Data does not consistent as other field	Change “Yes/No” to “Y/N”
property_valuation	Does not be collected in the same format, some are text, some are number	Change data type to number
There are 5 fields (between [property_valye] and [rank])	No field name	Add field name to make it has more meaningful