

Final Report

Following are the issues in the given three dataset/Excel sheet :

1. CustomerDemographic Sheet

a) Following Column has **missing/blank** values:

- last_name
- DOB
- job_title
- job_industry_category
- default
- tenure

All these Data fields must be completed to be able to drive a better insight from the data. (Completeness)

b) In **gender** Column we can see three issues :

- The values are not consistent, gender is either have to label Male/Female or M/F not both.
- The inaccuracy/incorrect data in the datafield, gender column is labelled as Femal
- Irrelevant value in the dataset which is value U in gender column.

c) In **DOB** column we can see that one of the year is labeled as 1843 it means they are 175 years old which is obviously not correct.

d) in **job_title** column we can see many job_title data has number in their job role which is not correct.

e) In **default** column we can see the non-validity of the data field because it has no meaningful data in it.

2. CustomerAddress Sheet

a) In **State** Column we can see data inconsistency i.e New South Wales and NSW. Both full name and codename used in the dataset.

b) In **Property_valuation** column no meaningful information derived from this field means that no currency sign/value is mentioned i.e we can't guess that 10 is either 10 million dollars or euro or if 10 billion.

c) After joining CustomerDemographic and Customer Address table we identify two issues:

- customer_id number **3, 10, 22, 23** records from **CustomerDemographic** table/sheet are not in **CustomerAddress** table/sheet.
- customer_id number **4001, 4002, 4003** records from **CustomerAddress** table/sheet are not in** CustomerDemographic** table/sheet.

3. Transactions Sheet

a) Following column has **missing/blank** values:

- online_order
- brand
- product_line
- product_class
- product_size
- standard_cost
- product_first_sold_date

All these Data fields must be completed to be able to drive a better insight from the data. (Completeness)

b) After considering all table/sheets we see that customer_id number **5034** made a transaction but its detail is not in our both (CustomerDemographic and CustomerAddress) sheets/table.

c) In **list_price** column no currency sign (dollar, euro etc) is mentioned.

d) In **standard_cost** Column we can see data inconsistency, some value has currency sign in it while some has not. Furthermore some value has 2 decimal value while some has more.

e) In **Product_first_sold_date** Column no information is derived from this data. Not correct date column.