**Observations from the dataset -**

**Transactions dataset**

1. Product ID –

There is a product ID number ‘0’ that appears with no corresponding brand name, product line, product class and product size. Please help me understand this product details.

1. Missing records – here is a list of all the missing records here.

Table

Description automatically generated

1. How is a product’s first sold date have some impact on this analysis?

**New Customers list**

1. Property valuation data type changed
2. Postal code data type needs to be changed
3. Missing records –

Table

Description automatically generated with medium confidence

1. Addresses starting with ‘0’ followed by a street name looks incorrect.

**Customer demographics**

1. Gender column has different forms of mentioning females and males(eg. F, Femal, Female, M, Male)
2. DOB Column – year mentioned as 1843? This looks incorrect.
3. Blank records –

A screenshot of a computer

Description automatically generated with medium confidence

1. Default column is not clear

**Customer address**

1. State names are ambiguous. Need to be changed to New South Wales, Queenslands, and Victoria.
2. Address columns have addresses staring with 0

Remember to take note of any assumptions or issues we need to go back to the client on. As well as recommendations going forward to mitigate current data quality concerns.

STANDARD DATA QUALITY DIMENSIONS

|  |  |
| --- | --- |
| Correct Values | Accuracy |
| Data fields with values | Completeness |
| Values free from contradiction | Consistency |
| Values up to date | Currency |
| Data items with values meta-data | Relevancy |
| Data containing allowable values | Validity |
| Records that are duplicated | Uniqueness |