**Phase 1- Data Cleaning.**

There are four sheets in the Excel file- Transactions, NewCustomerList, CustomerDemograhic, and CustomerAddress.

In the CustomerDemographic sheet, I sorted the date of birth from Oldest to newest, and there was a date of birth with the year 1843. This is obviously an error in the dataset, so I proceeded to delete the entire row.

Across the entire dataset, there are several missing values, so I proceeded to delete the blamk cells.

In the customerAddress sheet, the state column had some entries like QLD, VIC, NSW, New South Wales. This shows inconsistency in the entries. So I proceeded to write the abbreviations in full- QLD to QueensLand, VIC to Victoria, NSW to New South Wales.

It is the same issue with the NewCustomerList sheet, as the state column also had the same issue.

Also, the values in the Gender column of the ‘CustomerDemographic’ data are not consistent. There are male, female, F, M and U. I corrected the abbreviations to full- M to Male, F to female, and I left U as it is.

**Phase 2: Data Exploration.**

The tool of choice is R programming. I imported the dataset,

* Customer Address data:
* I converted the data type of customer ID, postcode and property valuation to numeric.
* Distinct number of Customer IDs is 3999
* Customer Demographic data:
* I converted the data type of tenure, past\_3\_years\_bike\_related\_purchase to numeric, and the date of birth to date data type.
* Distinct number of Customer IDs is 3137.
* New Customer List:
* I converted the DOB to date data type, past\_3\_years\_bike\_related\_purchase to numeric, tenure column to numeric, postcode column to numeric, and property valuation to numeric.
* I extracted the years from the dates of birth, and calculated their age, and created a new column for it.
* Distinct number of names is 759.
* Transaction data
* Transaction\_ID, product\_ID, customer\_ID, list\_price, standard\_cost to numeric data type. Transaction date to date data type, online\_order to numeric
* Distinct number of Customer IDs is 3317.

Then I proceeded to Tableau to Visualize.