**TASK 1**

**DATA QUALITY CHECK**

**Customer Demographic Table:**

Removed the top row which gives a note on internship.

Scrolled down to see there are 4000 customers with 4000 IDs. Made all the data into a table named ‘Customers’.

First name check – All the entries have the first name. There are no blanks.

Last name check – There are 125 customers who did not provide a last name and it’s blank. But I assume this won’t be a problem.

Gender Check – There are many entries like F, Female, Female, M, Male, U. I change entry with M as Male. Also, entries with F and Femal to Female.

There are 89 customers whose gender is U. If we look into gender wise, then this might rise and issue.

DOB check – 88 entries are blank. There is one entry with a customer ID 34 having DOB as 1843 which is illogical. This needs to be sorted out. Might be it’s a type – could be 1943. Waiting for your confirmation.

Job title check – 507 entries are blank

Job industry category check – 657 entries have n/a. This could be filled in manually based on the job title. If that is also missing, then I am waiting for your confirmation to proceed.

Default check – There is a column named default. It has miscellaneous characters which I believe it’s non relevant. I delete the whole column.

Tenure check - 87 entries are blank.

**Customer Address Table:**

Created a table from this dataset and named as Address.

Looks like there are 4003 IDs entered in this table whereas there are only 4000 IDs in Customers table. What happened to the customer details in Customers table?

State check – There are entries like NSW and New South Wales. Changed as New South Wales. Also Changed VIC to Victoria for consistency

Googled QLD and found it to be Queensland. So I replaced QLD with Queensland for consistency.

**Transactions Table:**

Created a table named Transactions with the Transactions dataset. Looks like there are 20000 transactions in total for the whole year 2017.

Product Id check - Not sure if there is a product id ‘0’.

Customer ID check - Customer ID 5034 is not there in the other two tables. But there are 3 transactions found in the transaction table. Need to verify.

Online order check – there are 361 blank entries. Need to verify and update as this is important to check if we have more online customers.

Order status check – There are 180 entries which are cancelled. Need to know the reason for cancel orders.

There are 198 blank entries about the product they have got and all the blank entries shows product ID as ‘0’. Curious about this product details. Is it really available?

List price check – The list price for the product ID ‘0’ is not consistent. Need to verify and require more inputs.

Product first sold date check – All the entries are in numbers, not date.