Dear Sprocket Central Pty Ltd,

Below is the summary of the data analysis from the data sets your provided, based on Standard Data Quality Dimensions:

Correct Values (Accuracy):

Table CustomerDemographic:

1. Issue: Gender column should only contain three values: Female, Male and Unknown, but there are rows which include unstandardized values, such as F, M, Femal and so on.

    Solution: I unified the unstandardized values, so that gender column only has Female, Male and Unknown.

2. Issue: The data type in the dob column should be Date, but there are some rows which are not correctly formatted.

    Solution: I re-formatted then entire column to have the Date data type.

Table CustomerAddress:

1. Issue: State column should only contain three values: New South Wales, Queensland and Victoria, but there are rows which include unstandardized values, such as VIC NSW, QSL and so on.

    Solution: I unified the unstandardized values, so that state column only: New South Wales, Queensland and Victoria.

Data Fields with Values (Completeness):

Table CustomerDemographic:

1. Issue: There are 125 rows whose last_name field is null.

    Solution: These rows need to be filled with corresponding data later.

2. Issue: There are 506 rows whose job_title field is null.

    Solution: These rows need to be filled with corresponding data later.

3. Issue: There are 65 rows whose job_industry_category field is null.

    Solution: These rows need to be filled with corresponding data later.

4. Issue: There are 87 rows whose tenure field is null.

    Solution: These rows need to be filled with corresponding data later.

5. Issue: All columns after tenure column have only null data.

   Solution: I removed all of them.

Table CustomerAddress:

1. Issue: All columns after property_valuation column have only null data.

   Solution: I removed all of them.

Table Transactions:

1. Issue: There are 360 rows whose online_order field is null.

   Solution: These rows need to be filled with corresponding data later.

2. Issue: There are 197 rows whose brand, product_line, product_class, product_size, standard_cost and product_first_sold_date fields are null.

   Solution: These rows need to be filled with corresponding data later.

3. Issue: All columns after product_first_sold_date column have only null data.

   Solution: I removed all of them.

Values Free from Contradiction (Consistency):

Table CustomerInformation (Left Joined from CustomerDemographic and CustomerAddress):

1. Issue: There are 4 rows with customer_id = 3, 10, 22 and 23in CustomerDemographic table, which do not have corresponding data in CustomerAddress table.

   Solution: These rows need to be filled with corresponding data later.

Data Containing Allowable Values (Validity):

Table CustomerDemographic:

1. Issue: Default column has weird data.

   Solution: I removed it.

Kind regards,

Ruohao Yin