

Dear Sprocket Central Pty Ltd,

Thank you for trusting KPMG for the task. Our team has reviewed the data set for quality assessment and has noted all the data issues. The summary of the data quality issues is listed in the table below.

TABLE	QUALITY ISSUES
Transaction	<ul style="list-style-type: none">• Completeness: missing values in 'online_order', 'brand' column• Relevancy: A new feature 'profit' would be more relevant• Validity: Text format of 'list_price', 'product_first_sold_date' are not valid
NewCustomerList	<ul style="list-style-type: none">• Completeness: missing values in 'last_name', 'DOB', 'job_title' column• Relevancy: A new feature 'age' would be more relevant• Validity: Text form of 'past_3_years_bike_related_purchases', 'post_code' and 'property_valuation' are not valid• Uniqueness: A duplicate column 'rank' is present
CustomerDemographics	<ul style="list-style-type: none">• Accuracy: inaccurate data is present in 'DOB' column• Completeness: missing values in 'last_name', 'job_title' column• Consistency: inconsistent data in 'gender' column• Currency: Deceased people data is present in 'deceased' column• Relevancy: 'default' column is not relevant, A new feature 'age' would be more relevant• Validity: Text format of 'past_3_years_bike_related_purchases' is not appropriate
CustomerAddress	<ul style="list-style-type: none">• Consistency: inconsistent data in 'state' column• Validity: Text format of 'Post_code' is not appropriate

Our team has identified the quality issues and come up with strategies to mitigate the issues. Some of the recommendations are shown below

Accuracy issues:

In the 'DOB' column of CustomerDemographics table date 1843-12-21 is not correct as it is 180 years from the current year. That row can be filtered out. Creating a new column 'age' can represent more comprehensible data.

Completeness issues:

There are missing values in the 'online_order', 'brand' column of Transaction table, 'last_name', 'DOB', 'job_title' column of NewCustomerList table, 'last_name', 'job_title' column of CustomerDemographics table. These missing values will affect analysis going forward so it should be filtered out.

Consistency issues:

In the gender column of CustomerDemographics table values are inconsistent. Male and M denote Male groups, similarly Female, Femal, and F belong to the Female groups and U may be Unspecified. We can create three groups of Gender Male, Female and Unspecified by replacing other data appropriately.

In the 'state' column of CustomerAddress table, NSW and New South Wales represent the same state, similarly Victoria and VIC are also the same state. Hence, we can replace New South Wales with NSW and Victoria with VIC for consistency.

Currency:

In the 'deceased' column of CustomerDemographics table, deceased people data are present which is no more required for analysis. The data corresponding to the "Y" value of the deceased column can be filtered out for the analysis.

Relevancy

The 'default' column CustomerDemographics table is not relevant. It should be removed from the worksheet. Again, column DOB is not much useful in providing data insight. Creating a new feature 'age' which is (current date-DOB) would be more relevant for the analysis.

In the Transaction table, the profit feature would be desirable. Hence a new column profit which is (list_price-standard_cost) can be added to the sheet.

Validity

The column 'past_3_years_bike_related_purchases', 'post_code' has text format which is inappropriate for the context. It should be converted to a number. The product_first_sold_date' should be converted to date format from the present text format. Again 'list_price' must be converted to currency format.

Uniqueness

There is a duplicate column 'rank' in the NewCustomerList table. This duplicate column can be deleted.

We believe, this quality assessment will be beneficial in moving forward with our objectives. Moving forward, the team will analyse the data to get an insight. We look forward to our continued collaboration and the positive outcomes.

Best regards

Subhransu S. Mallick

Junior Consultant, KPMG