Dear Customer [Name of contact],

Thank you for providing us the customer dataset of Sprocket Central Pty Ltd, based on which, after examining, we have encountered some quality issues as following. On top of that, recommendations to these issues are provided to avoid data inaccuracy and inconsistency as well as aiming to better derive useful business insights out of the dataset.

* **Data incompleteness**:

Missing values are found in a number of columns of “Transaction” and “CustomerDemographic” sheet such as blank cells in “Brand”, “Product Line”, “Product Class”, “Product Size”, “standard cost”, “Job Title”.

*Recommendation*: Please ensure that the data is filled in as much as possible. A valid reason should go along with wherever there is a missing item; thus, it is determined which methods should be adopted (filling in with substitute values, drop records…).

* **Data inconsistency:**

There is more than one format found in one column. For example, data in the field “DOB” are set in both text and custom format.

Value inconsistency is also found where two different values are recorded for the same attribute. For example, “F” and “Female” in “DOB” and “NSW” and “New South Wales” in “State”.

*Recommendation:* Please make sure that one standardized format is adopted throughout fields, and sheets of a dataset.

* **Data Inaccuracy**:

The data reflects inaccurate information. One customer is 178 years old which seems definitely wrong.

*Recommendation***:** Please set up rules or constraints to govern the accuracy of each field.

* **Data not coming from the same source**. 3 records from “Transaction” and 4 records from “CustomerAddress” do not come from the “CustomerDemographic” file.

*Recommendation:* Please make sure that data of Fact table should be all derived from a Dimension table (muster data).

Should you find any of the information above not aligned with your understanding, please do not hesistate to let us know, so that we could provide an explanation in more detail.

Best regards,