Dear [Name of the Client],

We have received the three raw datasets from SP rocket central Pty Limited. As a preliminary, we have analysed the quality of the raw data and found numerous issues that need addressing as highlighted below. We also provide Recommendations to mitigate the issues and improve effective usage of the data.

1. **Outliers**

**Issue:** some of the data values are outliers and can lead to disruption of dataset and errors when presenting reports. An example is the customer ID “34” with the name of Jephthah Bachmann DOB: 1843, meaning that he is 175 years old. This is an error in the data contained in the Customer Demographic Table.

**Recommendation**: Remove this particular data so it doesn’t skew the distribution of the dataset.

1. **Entries inconsistency across the dataset**

**Issue**: there are more entries in the customer ID field of the transactions table than that of the customer address and customer demographic.

**Recommendation:**  analysis will be performed only on the synchronized data of all the three customer tables across the customerID.

1. **Missing values**

**Issues:** in the Customer Demographic table “Job Title”, “Job\_Indusry\_Category” and “Tenure” some of the records are missing in these attributes.

The attributes “Product Line”, “Product Class”, “Online Order”, “Brand Name”, “Product Size”, “Standard Cost”, and “product\_first\_sold\_date” in the Transactions table had blank values.

**Recommendation:** the records occupy a very little percentage of the dataset, they can be removed without affecting our analysis.

1. **Inconsistent values in the same column.**

**Issue**: In the “State” field of the Customer Address Table multiple incoherent values were found such as “VIC” & Victoria, “NSW” & “New South Wales”.

Also, in the “Gender” column of the Customer Demographic Dataset.

**Recommendation**: use full spellings of states or abbreviations across all the records to ensure consistency of the addresses. For Gender Column, the records “U” can be imputed with the appropriate spellings to understand what it stands for.

1. **Invalid Data Types.**

**Issue:** in the Transaction table, in the field “Standard Cost” there are some records with special string characters ($) which causes inconsistency in the dataset.

**Recommendation:** Remove the special characters from the records and convert all the characters into numeric data to ensure consistent data types for effective data manipulation.

Kindly look into the quality issues mentioned above along with the recommendations to ensure the consistent quality of the dataset across all the tables. If all the suggestions are matched, we can proceed swiftly with further analysis of the data to derive some great insights for the company.

Warm Regards,

Sanni Muhaemin.