Dear \_\_\_ ,

We have received and analysed three tables (Transactions, CustomerDemographic and CustomerAddress) from SP rocket central Pty Limited and identified some issues with the data. We have provided recommendations on how to deal with any data problems. Here are the findings:

1. **Incorrect values**

**Issue**: Some of the data values are outliers and can disrupt the whole dataset. For example, date of birth for customer ID No.34 is 1843, meaning that he is 175 years old which is an error in the data in the Customer Demographic Table.

**Recommendation**: Remove the incorrect data as it may skew the distribution of the dataset.

1. **Missing Values.**

**Issue**: Missing values in three of the tables.

**Recommendation:** As the percentage of missing values in the datasets is low as compared to the whole dataset we can go proceed by removing them.

1. **Additional customer\_ids in the ‘Transactions table’ and ‘Customer Address table’ but not in ‘Customer Demographic’.**

**Issue**: The skewed data cannot be used if there are any missing records.

**Recommendation:** We will perform the analysis (model building) on the merged data of all the three customer tables on the customer\_ID.

1. **Inconsistent values.  
   Issue:** The column “Standard Cost” in the Transaction table need to be rounded to the same amount of decimal places. There is also a column “default” in CustomerDemographic table that has special characters, its role is unclear.

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

We will continue with the data cleaning and transformation. Please look into the above-mentioned quality issues along with the recommended changes to ensure the consistent quality of the dataset across all the tables. If all the suggestions are matched we can proceed with further analysis.

Regards,

Mykhailo