Dear(relevant person),

I hope this finds you well. Thank you for providing us with the three datasets; customer demographic, customer address, and transactions. After careful analysis, we came across some issues with the dataset. These include;

Completeness

There were blank spaces across the datasets. Product id, online orders, and brand columns in the transactions dataset had blank rows. The following columns had bank rows in the customer demographic dataset; customer id, last name, date of birth, and job title. In the customer address dataset, the customer id column has blank columns.

To mitigate this error, the blank spaces were removed from the three datasets.

Consistency

Data in some of the datasets were inconsistent. For example, the state column in the customer address dataset had inconsistent data. The state of New South Wales has been written in its complete and short forms. The gender column in the customer demographic dataset also has inconsistencies. Females and males have been inconsistently spelled throughout the dataset.

To mitigate these errors, the New South Wales state name was abbreviated to NSW, and the inconstant genders were corrected to be male and female only.

Accuracy

There are inaccurate data in the customer demographic dataset. The date of birth column has an invalid date of birth.

To mitigate this error, the invalid date of birth was removed since it was only one value.

Relevance

There is an irrelevant column in the customer demographic dataset. The default column is not relevant to any manipulation or analysis that will be conducted.

To mitigate this error, the column will be deleted. The deceased indicator column in the customer demographic dataset also has irrelevant data. Deceased individuals who cannot be targeted are included in the data.

Accuracy

The list price and the product first sold date column have been inaccurately entered. The product first sold date is entered as a number, while the list price does not have the currency symbol.

To mitigate these inaccuracies, the product first sold date has been changed to a short date, and the appropriate symbols have been added to the values in the list price column. As we advance, the team will continue to clean and transform the data documenting any questions and concerns.

Regards,

Imelda Atieno.