To,
Sprocket Central Pty Ltd

Subject - Data Quality issues and strategies to mitigate these issues.

Respected Sir/Ma'am,

Below are the information of the data quality issues and strategies to mitigate these data problems.

Sheet Name	Unique id cols	Shape of datasets	Shape of datasets (After cleaning)
Transactions	transaction_id	(20000, 13)	(19445, 13)
NewCustomerList	No Unique id	(1000, 17)	(715, 17)
CustomerDemographic	customer_id	(4000, 13)	(2630, 13)
CustomerAddress	customer_id	(3999, 6)	(3999, 6)

1. **Fields with missing values:**

Transaction's sheet contains null values to columns online order, brand, product line, product class, product size, standard cost, product first sold date.
NewCustomerList contains null values to the columns DOB, job title, job industry contains null values. CustomerDemographic → DOB, job title, job industry contains null values.

Mitigate Steps -
Now we will focus on data cleaning, fill null values as per the importance.

2. **Relevancy (Irrelevant values)**
   - NewCustomerList datasets has gender columns where gender value "U" Unknown we can use methods to replace with Male or Female or we can remove if value "U" is in less amount.
   - CustomerDemographics has gender column which has value "U" is unknown and F and Female is same value and M and Male is the same value in the column

Mitigate Steps -
The Now we will make the data into a use case format to manage the gender columns and another if required

3. **Consistency (Up to date values)**
   - Customer Demographics datasets- Here we will replace the value of the column gender where the value F to Female and M to Male
   - CustomerAddress datasets has the state column which has the value NSW and new south Wales is the same and Victory ad VIC is the same.

Mitigate Steps -
Same here we will handle and replace the value of the state where two value is same and we will make it in useful format.

Mitigate Steps -
The Data Handling team should Check the data type of columns to mitigate this type of issue.

Kind regards,
[New Internee]