Dear Sprocket central pty ltd,

Thank you for providing us with the dataset from Sprocket, we have received the data and summarised the data quality issue with the dataset. We have given our comments about we have solved these data quality issue and laid out a plan to move forward with the data cleaning.

|  |  |
| --- | --- |
| Worksheet Name | Data quality issue |
| Customer demographic | Completeness, consistency and relevancy |
| New customer list | Completeness and consistency |
| Transaction | Completeness and relevancy |

This table outline a few data quality issues with Sprocket central pty ltd. We have taken necessary steps and given suggestions below to avoid such issues.

**1. Worksheet name Customer demographic which has inconsistency in column ‘gender’, there were values missing and irrelevant field with ‘default’**.

a. The column for ‘gender’ had inconsistency, there were spelling errors for female and some rows had abbreviations. This was changed to Male and female form. The columns also consisted an ‘U’ value which was discarded from the column. More clarity would be appreciated on the same.

b. Null values removed from ‘job\_title’ and ‘job\_industry’

c. Column called default was removed from the data since it had no relationship with the data.

d. Variable ‘y’ was removed from the column ‘deceased\_indicator’.

**2. Worksheet name NewCustomerList where we identified blank values, there was also inconsistent values for gender**.

a. Blanks values were present in sheet for the column ‘second\_name’ and more blank and null values were found in ‘job\_title’ and ‘industry’.

b. The column ‘gender’ had inconsistency, spelling mistakes and abbreviations were present in this column. This was changed to Male and Female entirely. This columns also consisted an irrelevant variable ‘U’ this variable was discarded since it had connection to the column.

**3.Worksheet name Transaction where we identified blank values for columns ‘online\_order’ and ‘brand’. The column for ‘product\_ first\_sold\_date’ was converted into a date/ time format.**

a. Rectified various blank values in the columns mentioned above this ensure completeness.

b. The column for ‘product\_first\_sold\_date’ was converted into date/time format.

We are moving forward for further process like data cleaning, standardization etc and it will be great to spend more time with your data

Kind Regards,

Pranav Padmanabhan