Respected Sir/Ma’am,

This draft is written in regards to the data set that I reviewed as a part of the virtual internship at KPMG. I have reviewed the data sets which were provided by Sprocket Central Pty Ltd and during the data quality analysis, we have found some errors in the data sets. This mail contains details in association to identifying the data quality issues and strategies to mitigate these issues

The suggestions provided by me are in reference to the ‘Data Quality Framework Table’.

* We can take a mode year value for the missing records of customers DOB.
* Follow a consistent pattern for country code and other data that requires government compliances.
* We can assign a uniform last name of customers, which values are missing.
* Replace gender ‘U’ with reference to the customer’s name and make a consistent Theme.
* For tenure values, we can take a mean of rest of the values and assign the mean value to the missing fields in order to maintain the consistency of data.
* Eliminate the blank orders considering fake orders.
* Make the data set free of duplications and blank space to ensure better analysis.

The following are the details of error encountered in the data set.

# *Customer Demographic (Total records 4000)*

|  |  |
| --- | --- |
| **ATTRIBUTE NAME** | **ERRORS** |
| DOB | 01 record 1843  87 records Blanks |
| Last name | 125 records Blanks |
| Gender | 88 records gender ‘U’  Values are not consistence M, Male, F, Female, Femal, U |
| Job title | 506 records Blanks |
| Job industry | 656 records mention ‘N/A’ |
| Default | 3317 records value ‘special characters’ includes null and Blanks |
| Tenure | 87 records Blanks |

# *Transactions (Total records 20000 - past 3months)*

|  |  |
| --- | --- |
| **ATTRIBUTE NAME** | **ERRORS** |
| Online order | 94 records Blanks |
| Brand | 48 records Blanks |
| Product line | 48 records Blanks |
| Product class | 48 records Blanks |
| Product size | 48 records Blanks |
| Standard cost | 48 records Blanks |
| Product\_first\_sold\_date | 48 records Blanks |

Regards,

Buddaraju Naveen Kumar Raju

Intern, KPMG (Data Analytics, Information & Modelling Team)