Hello,

This is Sounam Ghosh from KPMG Data Analytics (Virtual Internship). I have reviewed the data sets which were provided by your company and during the data quality analysis, I have found some errors in the data sets.

The data quality analysis is the core phase and due to errors in the data set, I suggest the following mitigates in order to improve the data quality, which will eventually help us to driven the better analytics, results for your company.

* Create an ***age\_column***, allowing for more comprehensible data and easier to check for errors.

Create a ***profit\_column*** in **“Transactions”** to check accuracy of sales, it will assist in future monetary analysis.

* We can assign a uniform last name of customers; which values are missing.
* Replace gender ‘U’ with reference to the customer name and make a consistency.
* 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.

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

**Customer Demographic (Total records 4000)**

|  |  |
| --- | --- |
| **FIELD NAME** | **ERROS** |
| 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)**

|  |  |
| --- | --- |
| **FIELD 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 |

**Customer Address**

|  |  |
| --- | --- |
| **FIELD NAME** | **ERRORS** |
| Customer id | Incomplete(1 to 3500 will be used) |
| States | inconsistency |

Regards,

Sounam Ghosh

KPMG (Data Analytics Team)