This is Mashiyat Hussain from KPMG Data Analytics (Virtual Internship) team. I have reviewed the data sets which were provided by your company and during the data quality analysis, I have found the 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.

* We can take a mode year value for the missing records of customers DOB.
* 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

|  |  |
| --- | --- |
| **FIELD NAME** | **ERRORS** |
| last\_name | 125 data are missing |
| DOB | 87 data are missing |
| Gender | 88 records gender ‘U’  Values are not consistence M, Male, F, Female, Femal, U |
| job\_title | 506 data are missing |
| job\_industry\_category | 656 data mention ‘N/A’ |
| Default | 3317 records value ‘special characters’ includes null and Blanks |
| Tenure | 87 data are missing |

# Transactions

|  |  |
| --- | --- |
| **FIELD NAME** | **ERRORS** |
| Online\_order | 360 data are missing |
| brand | 197 data are missing |
| product\_line | 197 data are missing |
| product\_class | 197 data are missing |
| product\_size | 197 data are missing |
| standard\_cost | 197 data are missing |
| product\_first\_sold\_date | 197 data are missing |

# NewCustomerList

|  |  |
| --- | --- |
| **FIELD NAME** | **ERRORS** |
| last\_name | 29 data are missing |
| DOB | 17 data are missing |
| job\_title | 106 data are missing |
| Columns | 5 Unnamed columns whose name are given as  ‘Unnamed16’, ‘Unnamed17’, ‘Unnamed18’, ‘Unnamed19’, ‘Unnamed20’ |
| Gender | 17 records gender ‘U’ |
| job\_industry\_category | 165 data are missing |

Regards,

**Mashiyat Hussain**

**Email:- mashiyathussain2@gmail.com**

KPMG (Data Analytics Team)