Hello,

This is Piyush Gupta from KPMG Data Analytics (Virtual Internship) team. We have reviewed the data sets which were provided by your company and during the data quality analysis, we have found some errors in the data sets.

The data quality analysis is the core phase and due to errors in the data set we 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’s 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 the error encountered in the data set.

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

|  |  |
| --- | --- |
| **Field Name** | **Error** |
| Online Order | 360 records blank |
| Brand | 197 records blank |
| Product\_line | 197 records blank |
| Product\_class | 197 records blank |
| Product\_size | 197 records blank |
| Standard\_cost | 197 records blank |
| Product\_first\_sold\_date | 197 records blank |

**Customer Demogrpahic (Total records 4000)**

|  |  |
| --- | --- |
| **Field Name** | **Error** |
| Last\_name | 125 records blank |
| gender | 88 records ‘U’ Values are not consistent M, Male, F, Female, Femal, U |
| DOB | 87 records blank  01 record with ‘1843-12-21’ |
| Job\_title | 506 records blank |
| job\_industry | 656 records mention N/A |
| Default | Inconsistent records |
| Tenure | 87 records blank |

Regards,

KPMG (Data Analytics Team)