Hello,

This is Alberta Johnson from KPMG Data Analysis Virtual internship team. I have reviewed the dataset which were provided by Company and during the data quality analysis, I have found some errors in the data the dataset.

The data quality analysis is the core phase and due to the errors in the dataset, I suggest the following in order to improve the data quality, which will eventually help us to derive a better analysis result for your company.

1. We can take a mode year value of the missing records of customer DOB.
2. We can assign a uniform last name of customer, which values are missing.
3. Replace gender “U” with reference to the customer’s name and make a consistency.
4. 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 the data.
5. Eliminate the blank orders consisting fake orders

The following are the details of errors encountered in the dataset.

Transaction (Total records 20000 – past 3 months)

|  |  |
| --- | --- |
| **FIELD NAME** | **MISSING VALUES** |
| Online\_order | 360 |
| Brand | 197 |
| Product\_line | 197 |
| Product\_class | 197 |
| Product\_size | 197 |
| Standard\_cost | 197 |
| Product\_first\_sold\_date | 197 |

Regards,

KPMG (Data Analysis Team)