Hello,

This is Abhishek .J.N 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 the 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 customersDOB.
* We can assign a uniform last name of customers, which values aremissing.
* Replace gender ‘U’ with reference to the customer name and make aconsistency.
* 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 ofdata.
* Eliminate the blank orders considering fake orders.

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

# Customer Demographic (Total records 5000)

|  |  |
| --- | --- |
| **FIELD NAME** | **ERRORS** |
| DOB | 104 records Blanks |
| last\_name | 154 records Blanks |
| Gender | 88 records gender ‘U’  Values are not consistence M, Male, F, Female, Femal, U |
| job\_title | 612 records Blanks |
| job\_industry | 821 records mention ‘N/A’ |
| Default | ‘special characters’ includes null and Blanks, has no meaning |
| Tenure | 87 records Blanks |
| Address | 4 records Blanks |
| Postcode | 4 records Blanks |
| State | 4 records Blanks |
| Country | 4 records Blanks |
| Property\_valuation | 4 records Blanks |
| Rank | 4000 records Blanks |
| Value | 4000 records Blanks |

**Transactions (Total records 20000 )**

|  |  |
| --- | --- |
| **FIELD NAME** | **ERRORS** |
| Online\_order | 360 records Blanks |
| brand | 197 records Blanks |
| product\_line | 197 records Blanks |
| product\_class | 197 records Blanks |
| product\_size | 197 records Blanks |
| standard\_cost | 197 records Blanks |
| product\_first\_sold\_date | 197 records Blanks ,  Which has some numeric values instead of dates. |

Regards,

KPMG (Data Analytics Team)