Hello Sir/Madam,

Myself Rutuja Wardole 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 drive the better data analytics, results for your company.

1. We can take a mode year value for the missing records of customers DOB.
2. We can assign a uniform last name of customers, which values are missing.
3. Replace gender ‘U’ with reference to the customer 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 data.
5. Eliminate the blank orders considering fake orders.

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sheet name** | **Column name** | **Data quality issue** |  |  |  |
| Transactions | Online Order | Null values |  |  |  |
|  | Brands | Null values |  |  |  |
|  | Product first sold date | Relevency | convert int to date-time format | | |
| New customer list | last name | Null values |  |  |  |
|  | gender | consistency | U - unspecified | |  |
|  | job title | Null values |  |  |  |
|  | job industry | Null values |  |  |  |
| Customer Demographic | last name | Null values |  |  |  |
|  | gender | consistency |  |  |  |
|  | job title | Null values |  |  |  |
|  | job industry | Null values |  |  |  |
|  | Deceased indicator | Removed "yes" value |  |  |  |
|  | default | Relevency removed |  |  |  |
| Customer Address | State | consistency |  |  |  |

Regards,

Rutuja Wardole

KPMG -Data Analytics Team