Hello,

This is Shreyas V Achari from KPMG Data Analytics (Virtual Internship) team. We have reviewed the data sets provided by your company and during the data quality analysis, we 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 mitigate in order to improve the data quality, which will eventually help us to drive better analytics, and results for your company.

* We can take a mode year value for the missing records customer’s DOB.
* We can assign a uniform last name to customers, which values are missing.
* Replace gender ‘U’ as ‘unspecified’.
* For tenure values, we can take a mean of the rest of the values and assign the mean value to the missing fields in order to maintain the consistency of data.
* Eliminate blank orders considering fake orders.

**Employee Transaction (Total records: 2000)**

|  |  |
| --- | --- |
| **Field Name** | **Errors** |
| 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 | Values are of type float.  The date is the same for all the values.  197 records blank. |

**New Customer List (Total records: 1000)**

|  |  |
| --- | --- |
| **Field Name** | **Errors** |
| last\_name | 29 records blank. |
| DOB | 17 records blank. |
| job\_title | 106 records blank. |
| job\_industry\_category | 165 records blank. |
| gender | 17 records are unspecified. |

**Customer Demographic (Total records: 4000)**

|  |  |
| --- | --- |
| **Field Name** | **Errors** |
| last\_name | 125 records blank. |
| Gender | 88 records unspecified.  Certain categories are not correctly titled. (F, M, Femal, U) |
| DOB | 87 records blank. |
| job\_title | 506 records blank. |
| job\_industry\_category | 656 records blank. |
| default | 3698 records having special characters.  Also includes records that are blank. |
| tenure | 87 records blank. |

Regards,

KPMG (Data Analytics Team)