Hello,

This is Sharyar Virani 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 customers DOB.
- We can assign a uniform last_name of customers which values are missing.
- Replace gender 'U' with reference to the customer 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 error encountered in the data set.

Customer Demographic (Total records 4000)

FIELD NAME	ERRORS
DOB	01 record 1843 87 records Blanks
last_name	125 records Blanks
Gender	88 records gender 'U' Values are not consistence M, Male, F, Female, Femal, U
job_title	506 records Blanks
job_industry	656 records mention 'N/A'
Default	3317 records value 'special characters' includes null and Blanks
Tenure	87 records Blanks

Transactions (Total records 20000 -past 3months)

FIELD NAME	ERRORS
Online_order	94 records Blanks
brand	48 records Blanks
product_line	48 records Blanks
product_class	48 records Blanks
product_size	48 records Blanks
standard_cost	48 records Blanks
product_first_sold_date	48 records Blanks

Regards,
KPMG (Data Analytics Team)