Hello Sir / Ma'am,

This is Shruti Tandon from KPMG Data Analytics team. We have reviewed the data sets which were provided by your company Sprocket Central Pty Ltd, and during the data quality analysis, we have 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 mitigates 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 of customers DOB. We can assign a uniform last name of customers, for which values are missing. Replace gender 'U' with reference to the customer's name and get 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. We also need to covert the columns containing date, from Integer to datetime.

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

Transactions (Total records 20000)

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	197 records blank and Datatype=Integer

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
Unnamed Columns	5 unnamed columns, either drop them or name them

FIELD NAME	ERRORS
DOB	87 records blank
last_name	125 records blank
gender	88 records gender 'U' Certain categories are not correctly titled i.e., values are not consistent M, Male, F, Female, Femal, U. The names in these categories are to be re-named.
job_title	506 records blank
job_industry_category	656 records blank
default	302 records blank, column contains special characters and inconsistent values, hence drop the column.
tenure	87 records blank

Customer Address (Total records 3999)

There are no null and duplicate values. All the columns appear to have consistent and correct information.

Regards,
KPMG (Data Analytics Team)