Hello,

This is Bhakti Kulkarni from KPMG Data Analytics (Virtual Internship) team. We have received three datasets from sprocket Central pty LTD and have reviewed it. During Quality Analysis we have come across some errors in the data set.

Table Name	Number of Records	Distinct Customer ID	Data date received (dd-mm-yyyy)
Transactions	20,000	3496	03-07-2020
Customer Demographic	4000	4000	03-07-2020
Customer Address	3999	3999	03-07-2020

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.

- 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.
- We can take mode year value for missing records of customer DOB.
- We can assign uniform last_name of customer which value is missing.

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

Customer Demographic

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

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

Customer Address

FIELD NAME	ERRORS
state	Value are not consistence New South wales, NSW, QLD , VIC ,Victoria.

Regards,
 KPMG (Data Analytics Team)