

Hello,

This is Bhumika K R 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
<u>DOB</u>	01 record 1843 87 records Blanks
<u>last_name</u>	125 records Blanks
<u>Gender</u>	88 records gender 'U' Values are not consistence M, Male, F, Female, Femal, U
<u>job_title</u>	506 records Blanks
<u>job_industry</u>	656 records mention 'N/A'
<u>Default</u>	3317 records value 'special characters' includes null and Blanks
<u>Tenure</u>	87 records Blanks

Transactions (Total records 20000 -past 3months)

FIELD NAME	ERRORS
<u>Online_order</u>	94 records Blanks
<u>brand</u>	48 records Blanks
<u>product_line</u>	48 records Blanks
<u>product_class</u>	48 records Blanks
<u>product_size</u>	48 records Blanks

<u>standard_cost</u>	48 records Blanks
<u>product_first_sold_date</u>	48 records Blanks

Regards,
KPMG (Data Analytics Team)