

Greetings,
This is Nimrat Kaur from KPMG Data Analytics(Virtual Internship) team.
We have reviewed the files provided by your company and during the data quality analysis, we found few errors in the data set.

The data quality analysis is the most crucial phase and thus we strongly suggest the following mitigates in order to improve data quality which will help in carrying out better data analysis.

- We can make an age column from the DOB of customers.
- Replace U with unspecified for gender column in new_customer_lists table.
- Drop default column in customer_demographic.
- We can assign uniform last names to customers for whom value is missing.
- Eliminate blank orders, considering as fake orders.

CUSTOMER DEMOGRAPHIC TABLE(4000 RECORDS)

FIELD NAME	ERRORS
DOB	01 record 1843 87 records Blanks
last_name	125 records Blanks
Gender	88 records gender 'U' Values are not consistent 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,
Nimrat Kaur(Data Analytics Team)