Hello,

This is Kajal Jain from KPMG Virtual Data Analytics team. I have thoroughly reviewed the dataset provided by your company and during the data exploration and visualization, I came across below mentioned missing information in the dataset.

I have used various NumPy and pandas in Python programing Language for data cleaning process and will suggest below mentioned changes to be done for driving the company growth. The cleaned dataset will eventually help us to mitigate certain issues and provide better insight of data.

For the given dataset, data about 'NewCustomerList' is not useful for us to make any important decision for improving and enhancing growth of company. We have focused on cleaning three datasheets and described changes made to them below.

Details of error encountered:

Customer Demographic Sheet: (4000 entries)

FIELD NAME	ERRORS
DOB	01 record 1843 87 records Blanks NOTE: Age is created as new column with numeric value that denotes age of person.
last_name	125 records Blanks
Gender	88 records gender 'U' Values are not consistence M, Male, F, Female, Femal, U NOTE: Duplicates records are converted and duplicacy is removed.
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

Transaction Sheet (20000 entries)

FIELD NAME	ERRORS
Online_order	94 records Blanks Note: Removing 'N' online order from dataset while cleaning data.
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

Profit	A new column is created for calculating overall profit for sold items.
Order_Status	Note: Keeping information about only Approved order, as declined order status will not help us to make any decision.

Customer_adress Sheet (3999 entries)

State	Duplicate entries have been removed.
Country	Duplicate entries have been removed.

Regards,
KPMG (Data Analytics Team)