

Hello!

This is Debojjal Bagchi from KPMG Data Analytics (Virtual Internship) Team. I Reviewed the Data Provided, and I have come across with some issues. I have listed down the various issues and my recommendations to handle the issues, mitigate the errors and eventually improve the quality of Dataset. This will help us to provide better analytics results for your company. Kindly go through them and provide me with an official go ahead so that I can move forward with further analysis & predictions.

## **Analysis of Dataset & Recommendations:**

### ***1. Transaction data in the past three months***

<b>Data Field with Issues</b>	<b>Issues</b>	<b>Recommendations</b>
online_order	360 Empty Columns	Mark Empty Rows as 0.5 to mark neither 0(not online) or 1(online)
brand	197 Empty Columns	Fill Empty Rows with 'Unknown'
product_line	197 Empty Columns	Fill Empty Rows with 'Normal'
product_class	197 Empty Columns	Fill Empty Rows with 'Unknown Class'
product_size	197 Empty Columns	Fill Empty Rows with 'Unknown Size'
standard_cost	197 Empty Columns	Fill Empty Rows with mean standard cost
product_first_sold_date	197 Empty Columns	Date format could not be understood. All Rows showed same date. Recommend dropping this data field for analysis.

### ***2. Customer Demographic***

<b>Data Field with Issues</b>	<b>Issues</b>	<b>Recommendations</b>
last_name	125 Empty Columns	Fill All Empty Rows, i.e assign a uniform last name to people without last name with 'last_name'
gender	Inconsistent data with male , female, m , f, u, femal	Use consistent Male/ Female/ Unspecified

DOB	87 Empty Columns. People with such DOB which make age more than 180 present.	Find Age of Each person. Create a Age column and fill data. Fill Empty age rows with mean age. If Age is more than 180 fill it with mean age of rest of the group
job_title	506 Empty Columns	Fill Empty Rows with 'Others'
job_industry_category	656 Empty Columns	Fill Empty Rows with 'Not Applicable'
default	302 Empty Columns, Data Field Could not be understood	Irrelevant Data. Recommend dropping this column

### 3. Customer Addresses

Data Field with Issues	Issues	Recommendations
No Issues with Data		

### 4. New Customer List

Data Field with Issues	Issues	Recommendations
last_name	29 Empty Columns	Fill All Empty Rows, i.e assign a uniform last name to people without last name with 'last_name'
gender	Inconsistent data with male , female, m , f, u, femal	Use consistent Male/ Female/ Unspecified
DOB	106 Empty Columns. People with such DOB which make age more than 180 present.	Find Age of Each person. Create a Age column and fill data. Fill Empty age rows with mean age. If Age is more than 180 fill it with mean age of rest of the group.
Unnamed: 16	The Type of information couldn't be understood	Recommend dropping this column
Unnamed: 17		
Unnamed: 18		
Unnamed: 19		
Unnamed: 20		

### Some Other Recommendations:

1. Customer Demographic Data & Customer Addresses Data together have same Information fields as the New Customer List Data. Recommend merging the

Demographic & Addresses Data and finally adding the merged Data to New Customer Data by assigning IDs to New Customers.

I have incorporated all the above changes & I have uploaded the files on my GitHub repository - [https://github.com/debojjalb/KPMG\\_Module\\_1](https://github.com/debojjalb/KPMG_Module_1)

Anticipating getting a positive go ahead from you soon

Stay Safe!

With Warm Regards,

Debojjal Bagchi

KPMG Data Analytics (Virtual Internship) Team

debojjalb@gmail.com