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 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 a 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**

|  |  |  |
| --- | --- | --- |
| **Data Field with Issues** | **Issues** | **Recommendations** |
| 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. |

1. **Customer Demographic**

|  |  |  |
| --- | --- | --- |
| **Data Field with Issues** | **Issues** | **Recommendations** |
| 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 |

1. **Customer Addresses**

|  |  |  |
| --- | --- | --- |
| **Data Field with Issues** | **Issues** | **Recommendations** |
| No Issues with Data | | |

1. **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>

Anticipating to get a positive go ahead from you soon

Stay Safe!

With Warm Regards,

Debojjal Bagchi

KPMG Data Analytics (Virtual Internship) Team

debojjalb@gmail.com