Hello,

Thank you for providing me with the datasets by Sprocket Central Pvt Ltd.

This is Anuj Arora from KPMG Data Analytics (Virtual Internship) team.

I have reviewed the data sets which were provided by the company and during the data quality analysis, I have found some issues in the datasets.

Here is the Summary of the issues found in the given datasets

1. **Transactions**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Accuracy | Completeness | Consistency | Relevancy | Validity |
| Profit : Missing | Online Order : Blanks  Brands : Blanks |  | Cancelled Status : Filter | List Price : Formatting  Product Sold Date : Formatting |

|  |  |
| --- | --- |
| **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 |

1. **Customer Demographic**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Accuracy | Completeness | Consistency | Relevancy | Validity |
| DOB : Inaccurate formatting  Age : Missing | Job Title: Blanks | Gender Inconsistency | Deceased Customers : Filter  Default Column : Delete |  |

|  |  |
| --- | --- |
| **FIELD NAME** | **ERRORS** |
| DOB | 01 record 1843  87 records Blanks |
| last\_name | 125 records Blanks |
| Gender | 88 records gender ‘U’  Values are not consistence 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 |

1. **Customer Address**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Accuracy | Completeness | Consistency | Relevancy | Validity |
|  |  | State Inconsistency |  |  |

The data quality analysis is the core phase and due to errors in the data set, I 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.
* 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.

Regards,

KPMG (Data Analytics Team)