Dear Client,

Thank you for providing us datasets. There were 4 of them, we have reviewed and summarized the data quality issues in the following table. Please let us know if you have any queries on issues presented.

|  |  |  |  |
| --- | --- | --- | --- |
| **Sheet** | **Data Issue** | **Data Quality** | **Recommendations and Mitigation** |
| Transactions | Online order has Missing values | Completeness | Filter out 360 missing value rows as very few compared to total datasheet |
|  | Brand has missing values | Completeness | Filter out all the 196 rows of missing values. |
|  | Product\_line missing values | Completeness | Filter them out |
|  | Class missing values | Completeness | Filter them out |
|  | size has missing values | Completeness | Filter them out |
|  | Product\_first\_sold\_date has missing values | Completeness | Filter them out |
|  | Product\_first\_sold\_date is not in date format | Validity | Change format to date format |
|  | Standard cost for customer\_id 5034 | Relevancy | Change the number to currency format |
|  | List price format | Validity | Change to currency format |
| NewCustomerList | Gender has U values | Accuracy | Please specify more information |
|  | DOB has missing values | Completeness | Filter out or update 17 missing values |
|  | Job\_title has missing values | Completeness | Filter out or update 106 missing values |
| CustomerDemographic | Gender has incorrect values | Consistency | Correct the values in dataset. Mitigate using Drop down menu |
|  | DOB datatype | Validity | Change datatype of DOB to date |
|  | Missing values in DOB | Completeness | Filter out missing values. Mitigate by making it mandatory while entry |
|  | Outlier in DOB | Accuracy | Check for incorrect values by deriving a age column |
|  | Job\_title has missing values | Completeness | Filter out from dataset or Update |
|  | deceased\_indicator | Relevancy | Filter out deceased customers |
|  | Default column | Relevancy | Remove meta data column |
|  | tenure has missing values | Completeness | Filter out missing values or update |
|  | Customer\_id missing rows | Completeness | Few entries of customer\_id are missing which are found in other datasets |
| CustomerAddress | State has incorrect values | Consistency | Correct the values in dataset. Mitigate using Drop down menu |
|  | Customer\_ids missing | Completeness | Few entries of customer\_id are missing which are found in other datasets |

Moving forward, we will clean the data and transform for modeling. Assumptions will be documented separately. This summarizes the first phase of analysis.

Kind Regards,

Krupaul Mallepudi.