Hi Team,

This is Regarding Data Quality Analysis on the given Data. Below are the data quality issues identified in the given datasets.

**Transactions:**

* **Missing Values:** brand, Product\_line, product\_class, product\_size,standard\_cost, product\_first\_sold\_date has almost 1% of the data missing which is negligible and all the columns have missing values for same records

online order has 1.79% missing values, as it is categorical variable, we can make it as other new level "unknown"

* **product\_first\_sold\_date**: The column has date values in form of ordinal values, which has to be converted to date format.

**NewCustomerList:**

* There is no customerid to uniquely identify each customer.
* There are five columns in this table without name of the column, however these columns are present in the CustomerDemogrphic and CustomerAddress table so these may not be required for present scope.

**CustomerDemographic:**

* 1 customer with id 5034 from Transaction table doesn't have information in CutomerDemographics table.
* The DOB has values from 1843 which is more than 100 years a person more than hundred years cannot be alive. only one customer 34 has DOB as 1843 which can be treated as outlier.
* This column has Data accuracy issue the Female level is represented in different levels like 'F','Femal','Female' same with male level, There is another level 'U' which can be considered as unknown level.
* Almost every customer has deceased\_indicator 'N'.
* Column ‘default’ has different values and different data which seems like noise.

**CustomerAddress**:

* Except customer\_id(3,10,22,23) all the customers have CustomerAddress information.