Dear Client,

Thank you very much for the dataset provided.

However, I have analyzed the dataset and below are my findings and possible recommendations we need to apply going forward, in other to achieve accuracy in our analysis.

**Incomplete data**

To determine customers who made online purchases, the ***online\_order*** column must contain complete data. The following columns below contain missing data;

* **Transactions**
* *Brand*
* product\_line
* product\_class
* product\_size
* standard\_cost
* product\_first\_sold\_date
* **The NewCustomerList sheet**
* job\_title
* job\_industry\_category
* **CustomerDemographic**
* DOB
* job\_title
* job\_industry\_category
* tenure

Looking at the data accuracy, the **list\_price** column (Transactions sheet) should contain currency symbol and the **product\_first\_sold\_date** column should be in date format. Also, the DOB cell with 1843-12-21 (under CustomerDemographic sheet) is a wrong data collection.

For data accuracy and consistency, the gender column (NewCustomerList and CustomerDemographic sheets) should not contain the letter “U” as this does not denote a proper gender.

The column “default” (CustomerDemographic sheets) is absolutely irrelevant in the dataset because it contains invalid data.

In other to mitigate all of these, then there should be proper data collection/data cleaning after that, we should be getting accurate result upon completion of our analysis.

Regards,

Ifeanyi Michael K.

Data Analyst