Hello Sir,

Greetings of the day..

Thanks for connecting the KPMG to grow your business. After analysing the dataset we found the following data issues:

1. There is one DOB which can’t be possible i.e. 1843-12-21 having customer ID 34
2. There is no requirement of complete address details of the customer, we can address them with their postal code which made easier to locate the fruitful area of business.
3. There is a column called product\_first \_sold\_date in Transaction sheet which should be in date format but there are some numbers only so it should be rectify or we can delete this column as this is not going to help to predict any insights
4. There are products whose list price & standard prices both are missing
5. In Customer Demographic Data, there are errors in gender column. There is spelling error for Female. Male & Female represented in two different ways which doesn’t make any sense. We can represent them either by M&F or by Male & Female. Also, there is one attribute ‘U’ which is unclear.
6. There is no use of column called ‘default’ in customer demographic data as it doesn’t contain any useful information to predict the area for growing, it can be removed from the data during data cleaning.
7. The property value column of New Customer List has two different formats for same value like for representing ‘8’ we have 8 & 8.00. We must similar the value format for every value.
8. In Transactions Data, there are duplicate customer id, it should be unique.
9. There are null values as well in every datasheet, if they are very less in number we can remove them otherwise we have to fill them by any means of method to have a better understanding.

I hope I am able to identifying the data quality issues and able to provide the strategies to mitigate these issues.

Looking forward to work with you further on your requirement and provide the best of our services with the best team. Do let me know if you have any queries or need any further assistance.