

Dear client ,

This is SARANYA N from the Data Analytics Team. first of all thank you for giving these dataset .

As per your voicemail, we have reviewed the datasets provided by you for its quality. During analysis we have encountered some issues and errors that have to be corrected and rectified to ensure that the data is ready for analysis in phase two.

As Data Quality Assurance is an important step in analysis, we have some suggestions and recommendations that can help to improve the overall quality of the data.

- ❖ As all the three datasets are relevant to each other, it is better to merge all of them to improve the data for analysis.
- ❖ We can take a mode year value for the missing records of customers DOB.
- ❖ For missing tenure values, it is suggested to take the mean of the tenures and replace the missing values with them in order to ensure consistency.
- ❖ In the Transactions dataset, the 'product_first_sold_date' column needs to be rechecked and ensured that it has relevant information.
- ❖ It is recommended to drop out the unnecessary columns like the 'default' column in the Customer Democratic Table.
- ❖ Inorder to ensure more relevancy and readability, it is better to use 'Male', 'Female', 'Unidentified' as values for the gender column.
- ❖ It is also recommended to try and fill in the missing blank values for improved analysis during the phase two.
- ❖ Eliminate the blank orders considering fake orders.

	No of records	Unique ids
Transaction dataset	20002	3495
Customers demographic	4002	4002
Customers address	4001	4001

The above mentioned are some of the strategies that can be used to mitigate the data quality issues that were discovered in the given datasets.

Regards,

Saranya N
Data Analytics Team.