Hello,

This is Riddhi Chodvadiya from the Data Analytics Team. 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 mitigate the inconsistencies and 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.
- For the missing DOB values of customers, we can replace the missing places using the mode of all other customers. Or just convert the DOB format into Age, Month and Year and drop the DOB column to ensure more readability and consistency.
- 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 as well as Unnamed column 16 to 20 in New Customer List Table.
- In order 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.
- The blank missing values can also be eliminated, considering them as fake entries or incomplete entries.

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,
Riddhi Chodvadiya
Data Analytics Team