

Hello,

This is Kelvin from the Data Analytics Team. As per your voicemail, We checked the datasets you provided for quality in accordance with your voicemail. To make sure that the data is prepared for analysis in phase two, we must correct and rectify some problems and errors that we discovered during analysis.

We have some suggestions and recommendations that can help reduce the inconsistencies and enhance the overall quality of the data because Data Quality Assurance is a crucial step in analysis.

- Since there is overlap between the three datasets, it is preferable to combine them all to create better data for analysis.
- We can substitute missing places for missing DOB values of customers using the mode of all other customers.
- It is advised that, in cases where tenure values are missing, the mean of the tenures be used to fill in the gaps, ensuring consistency.
- The 'product_first_sold_date' column in the Transactions dataset needs to be double-checked to make sure it contains accurate data.
- The 'default' column in the Customer Democratic Table should be removed as it is considered unnecessary.
- The best values for the gender column are "Male," "Female," and "Unidentified," as they ensure greater relevancy and readability.
- For better analysis during phase two, it is also advised to try and fill in the blanks that are missing from the data.
- The blank missing values can also be removed because they are regarded as fictitious or imperfect entries.

Some of the methods that can be used to address the data quality problems that were found in the provided datasets are those that were just mentioned.

Regards,

Kelvin Simon

Data Analytics Team