Subject: Feedback on Data Quality Review - Sprocket Central Pty Ltd Datasets

Dear Manager,

I hope this email finds you well. This is Ameen from the Data Analytics Team, writing in response to your voicemail regarding the review of the datasets provided by Sprocket Central Pty Ltd.

We have completed the review of the datasets and have identified several issues and errors that need to be addressed before we can proceed with the analysis in phase two. Our team is dedicated to ensuring that the data is of high quality and suitable for analysis. In order to improve the quality of the data, here are some of our findings and recommendations:

- The “product_first_sold_date” is of the wrong data type. If its of great importance to our analysis, it can be corrected from our end here.
- The Value count in the gender colum should be reduced to three (Male, Female and Others). Presently we have six and they have no meaning and could not be a pointer to anything in our analysis.
- We recommend dropping the “default” column in the customer demographic sheet as they are of no essence to our analysis.
- In the transaction sheet, there are lots of missing values in the following columns (`online_order: 360, brand: 197, product_line: 197, product_class: 197, product_size: 197, standard_cost: 197, product_first_sold_date: 197`) .

With regards to the missing values, here are some actions that can be taken:

1. **Deletion**: Deleting the rows or columns that contain missing values, but this method may result in loss of information.
2. **Imputation**: Replacing missing values with statistical measures such as mean, median, mode, etc.
3. **Interpolation**: Using statistical models or algorithms to estimate the missing values based on the available data.
4. **Prediction**: Building a predictive model to predict the missing values based on the available data.

The choice of technique will depend on the amount of missing data, the reason for the missing data, and the desired outcome of the analysis. In some cases, a combination of techniques may be used to handle missing values effectively.

We hope these suggestions and recommendations will be helpful in mitigating the data quality issues and improving the overall quality of the data. If you have any questions or concerns, please don't hesitate to reach out.

Thank you for your time and consideration.

Best regards,
Ameen
Data Analytics Team