**EMAIL TO SENIORS ABOUT DATA CLEANING**

Subject: Data cleaning of client Sprocket

Respected sir/mam,

I hope this email finds you well.

As we have discussed earlier regarding the Sprocket case, I have started with the work on it. I wanted to bring your attention to the crucial step of data cleaning. Before I carry on with the analysis, I would first have to clean the dataset to be clean and free of errors.

Data cleaning is the first step that we should start with in data analysis and it involves identifying and rectifying errors, inconsistencies, and inaccuracies within our datasets. By investing time and effort into this process, we can enhance the quality of our analyses,

Ensuring the cleanliness and reliability of our datasets is fundamental to obtaining accurate and meaningful insights as it minimizes potential biases, and allows us to make informed decisions based on reliable information.

Before I proceeded with the data cleaning, I converted the dataset to a table and arranged it in ascending order with respect to date (in sheets where there is a date column).

Here is the data cleaning activity I performed on the database using Excel:

1. Removed inconsistent and irrelevant formatting i.e., formatting date and currency columns.
2. Dropped rows with duplicates and blank values from the dataset in all tables.
3. Removing illogical values from the dataset. For e.g.- keeping Male, Female in the gender column, removing illogical date of births etc.
4. Dropping columns with irrelevant information. For example: The deceased column in New customer sheet and the country columns as it is obvious we are in Australia were removed.
5. Rounded up the values with a lot of decimals to just 3 values for ease in calculation.

If you have any questions or concerns regarding the data cleaning process or require assistance, please feel free to reach out to me. Together, we can ensure the reliability and integrity of our datasets, leading to more accurate and valuable analyses.

Thank you for your attention to this matter. I appreciate the commitment you have when it comes to maintaining high data quality standards.

Best regards,

Ishan

Data Analyst Intern

KPMG AU

PS: I have kept the old dataset separate as well in case you needed it. The new transformed dataset is in a separate excel file.