

Dear [Recipient],

I would like to extend my gratitude to you for providing us with the dataset. Following a comprehensive analysis of the data, we have identified several data quality issues that need to be addressed. I would like to highlight these concerns and propose strategies to mitigate them:

a. Completeness: One prominent issue observed across all datasets is the presence of missing values. Specifically, we have identified missing values in the following columns:

- Transactions dataset: Online\_order, Brand, Product\_line, Product\_class, Product\_size, Standard\_cost, and Product\_first\_sold\_date. Notably, all missing rows in columns G, H, I, J, L, and M correspond to this issue.
- NewCustomerList dataset: Last\_name, DOB, Job\_title.
- CustomerDemographics dataset: Customer\_id, First\_name, Last\_name, Past\_3\_years\_bike\_related\_purchase, Gender, DOB, Job\_title, Job\_industry\_category, Wealth\_segment, Deceased\_indicator, Owns\_car, Tenure.
- Customer\_address: Address, Postcode, Country, and Property\_valuation.

To address this problem, we could explore the possibility of obtaining the missing values through additional data collection.

b. Accuracy: We have identified an incorrect value in the "DOB" column of the CustomerDemographics dataset. To rectify this, we suggest verifying the accuracy of the provided data and correcting any discrepancies by cross-referencing with reliable sources or contacting the individuals directly.

c. Consistency: Lack of consistency has been observed in the "gender" column of both the CustomerDemographics and NewCustomerList datasets. To ensure data consistency, we propose conducting a thorough review and standardization of the gender values across the datasets.

d. Relevancy: The presence of the "defaults" column in the CustomerDemographics dataset has been identified as irrelevant. We suggest removing this column to streamline the dataset and enhance its overall coherence.

e. Data validity: The "list\_price" column and the "product\_first\_sold\_date" column exhibit formatting issues. To address this, we recommend conducting a data validation process to ensure that these columns adhere to the required format and standards. This may involve reformatting the data or converting it into the appropriate data types (Currency & date format) for consistency and analysis purposes.

In conclusion, we understand the importance of data quality and its impact on the reliability and validity of our analysis. To mitigate the aforementioned issues, we propose implementing the strategies outlined above.

Thank you once again for providing us with this opportunity. We are committed to addressing these data quality concerns promptly and look forward to working collaboratively to achieve the best possible outcomes.

Yours sincerely,

Nneamaka Uzoegwu