| **Mubashir Bello**  Data Analyst  **ㅡ**  (+234) 816 858 9208  bellomubashir4@gmail.com | horizontal line **Sprocket Central Pty Ltd**  ***KPMG***  Dear Sir,  I hope this email finds you well. I wanted to reach out to you regarding some data quality issues that we have identified in your company's dataset. As you know, data is a crucial component of any business operation, and ensuring its quality is vital for accurate decision-making.  During our recent analysis, we found that the following issues are affecting the accuracy and completeness of your data:   1. Inconsistent Data Entry: We noticed that there is a lack of standardization in the data entry process, leading to variations in how data is recorded across the dataset. The following are some of the lists of inconsistent data entries in the dataset    1. The gender data field (Customer Demography table) has a value that has no full meaning in the context of the field and the value is “**u**”. Also, gender has fields with similar meanings but typed in the system with different formats, we have ‘female’, ‘femal’, and ‘f’ which are all similar, and also ‘male’ and ‘m’ are similar.    2. DOB (Customer Demography table): Jephthah's date of birth is 1843-12-21 which is way beyond the average age and it seems to be a mistake because his age will be 180 as of December 2023.    3. Job\_industry\_categories (Customer Demography table) This field has a value ‘na’ which is allocated to some job whose category is already defined, for example ‘web developer’ which should be in the category of ‘IT’, 520 users filled na for their job.    4. The default field (Customer Demography table) has no significant meaning in the dataset.    5. The customer's address (customer address), has some inconsistent data, New souths wales and NSW is similar, also VIC and Victoria are similar.    6. The transaction customer\_id and customer\_address customer\_id have a value that doesn’t exist in the customer\_demography table, which should be looked into, or full data set should be provided. 2. Missing Data: Some fields in your database contain missing data, which can lead to incomplete records and inaccurate reporting. The following are fields with their tables with missing values.    1. The customer demography table has lots of missing data some filed with missing values are:       1. The last name field has 125 missing values which may occur because the field is not compulsory       2. The DOB field has 87 missing values.   To mitigate these data quality issues, we recommend the following strategies:   * Standardize Data Entry: Establishing a set of guidelines and protocols for data entry can help ensure consistency across systems and departments. This can include standardized naming conventions, formatting rules, and data validation checks. * Educate Staff on Data Quality: It's essential to educate your staff on the importance of data quality and provide them with the training they need to understand how to input data correctly and consistently. * Invest in Data Management Tools/ Automate each process: There are a variety of data management tools available that can help streamline the data entry process, automate data validation checks, and detect and remove duplicates.   We believe that addressing these data quality issues will have a positive impact on the accuracy and completeness of your data, leading to better decision-making and overall operational efficiency.  Please let us know if you have any questions or concerns, and we will be happy to discuss this matter with you in more detail.  Thank you for your time and attention. Best regards Mubashir Bello |
| --- | --- |