**Subject**: Data Analysis - Data Cleaning Update

Dear Manager,

I hope this email finds you well. I am writing to provide an update on the data cleaning process for the four datasets from Sprocket Central Ltd that were recently provided to us.

The summary table below highlights key data quality issues we have discovered in the data cleaning process. Please let us know if you have any queries concerning the issues presented.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **Accuracy** | **Completeness** | **Consistency** | **Timeliness** | **Relevancy** | **Uniqueness** | **Validity** |
| (Dataset 1)  Customer Demographics | Erroneous: DOB | Missing: last name, job title, job industry, tenure | Format: gender | Out of date: deceased customer | Corrupted:  default column deleted |  | Unreliable:  age |
| (Dataset 2)  Customer Address |  |  | Format: state |  |  |  |  |
| (Dataset 3)  Transactions | Missing:  profits (cost - price) | Missing:  customer ID (3501-4999), online order, brand, product line/class/size, standard cost, first sold date | Format:  list price, standard cost |  | Filtered: cancelled orders |  | Format:  product sold date |
| (Dataset 4)  New Customer |  | Missing: last name, DOB, job title, job industry |  |  | Irrelevant:  unnamed columns | Duplications:  customers |  |

As outlined in the summary table, key data quality issues were discovered during the data cleaning process. These issues included inaccuracies in the date of birth, inconsistencies in customer IDs, missing values in various fields, inconsistencies in gender and state formats, and irrelevant or corrupted data.

To mitigate these issues, various strategies were implemented, such as filtering out outliers, filling in missing values, and standardizing formats. Additionally, recommendations were provided to improve the accuracy of the data sources in the future, such as creating an age column, ensuring tables are up to date, using drop-down options for input fields, and updating information on deceased customers.

Please let me know if you have any queries concerning the issues presented. I am happy to provide more in-depth explanations and answer any questions you may have.

Overall, the data cleaning process aimed to improve the accuracy and consistency of the data to inform future business decisions. I believe that by addressing these data quality issues, we will be able to provide more reliable and accurate insights to support your decision-making process.

Thank you for your time and consideration.

Best regards,

Uday Damerla

Data Consultant.