Dear Sir,

Thank you for providing me with the datasets from Sprocket Central Pty Ltd. After analyzing the data, I have identified several key data quality issues, which I would like to discuss with you. Please feel free to reach out if you have any questions or concerns about the issues presented below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sheet Name** | **Accuracy** | **Completeness** | **Consistency** | **Currency** | **Relevancy** | **Validity** |
| **Customer Demographic** | DOB: inaccurate | Blanks: job title, dob | Gender inconsistency | Deceased customers. | Default column misc. |  |
| **Customer Address** |  |  | States: inconsistent |  |  |  |
| **Transactions** | Profit: missing | Blanks: online orders, brands |  | Order\_status: cancelled. |  | List\_price format,  product\_sold\_date format |
| **New Customers list** | DOB: inaccurate |  | Gender: inconsistent |  | Rank column with formula included | past\_3\_year\_bike\_related\_  purchases, postcode,  property\_valuation.  Value column: appropriate  decimal places |

Here is a detailed description of the data quality issues discovered and the mitigation methods used. Implementing these recommendations will improve the accuracy of the data, thereby influencing future business decisions made by Sprocket Central Pvt. Ltd.

* **Accuracy:**
* The DOB was found to be inaccurate in the Customer Demographic and New Customers List Sheets.
* The Profit column was missing in the Transactions Sheet.
* Solution: Filtering out the outliers and invalid DOBs.
* Recommendation: Create an additional profit column (list\_price – standard\_cost) to check data accuracy and identify errors more easily. This will also assist in future financial analysis.
* **Completeness:**
* There are blanks in the job\_title and dob columns in the Customer Demographics Sheet.
* There are blanks in the online\_order and brands columns in the Transactions Sheet.
* Solution: Filtering out the blanks in the above-mentioned columns.
* Recommendation: Provide drop-down options for the job\_title, online\_order, and brand columns to ensure completeness and accuracy of the data.
* **Consistency:**
* Inconsistency in gender values for Customer Demographics and New Customers List.
* Inconsistency in state names in the Customer Address Sheet.
* Solution: Find and replace inconsistencies in gender (e.g., M for Male, F and Femal for Female). Find and replace New South Wales with NWS and Victoria with VIC.
* Recommendation: Implement drop-down menus at the client side to minimize human error and increase data consistency.
* **Currency:**
* Deceased Customers (denoted by 'Y' in the Deceased column) are not relevant in the Customer Demographic Sheet.
* Solution: Filter out the 'Y' values for deceased customers.
* Recommendation: Regularly update the deceased column whenever new information is received to keep the data up-to-date. This will greatly impact future analysis.
* **Relevancy:**
* There is an irrelevant default column in the Customer Demographic Sheet.
* There are cancelled order\_status values in the Transactions Sheet.
* Solution: Remove the default column in the Customer Demographic Sheet. Filter out the cancelled order\_status values in the Transactions Sheet.
* **Validity:**
* The format of the List\_price and product\_sold\_date columns in the Transactions Sheet needs adjustment.
* The format of past\_3\_year\_bike\_related\_purchases, postcode, property\_valuation, and value columns in the New Customers List also requires attention.
* Solution: Format the List\_price as currency and product\_sold\_date as a date in the Transactions Sheet. Format past\_3\_year\_bike\_related\_purchases, postcode, and property\_valuation as numbers. Format the value column with 3 decimal places.
* Recommendation: Set appropriate data types for input fields in the client-side form code to avoid time consumption during the data cleaning process.

The above quality checks summarize all the data quality issues discovered in Task 1 of the data quality assessment. By implementing the provided solutions, we can improve the level of analysis performed by the KPMG team, resulting in greater efficiency and improved output.

Please feel free to reach out if you have any questions regarding the solutions provided or any other data quality issues identified.

Kind regards,

Aditi Sharma