Dear John Doe,

We appreciate your provision of the three datasets from Sprocket Central Pty Ltd. The summary table presented below outlines significant quality concerns identified across the three datasets. Should you have any inquiries regarding the outlined issues, please don't hesitate to reach out.

**Summary Table**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Accuracy** | **Completeness** | **Consistency** | **Currency** | **Relevancy** | **Validity** |
| **Customer Demographic** | * DOB: Inaccurate * Age: Missing | * Job Title: blanks * Customer ID: Incomplete | * Gender: Inconsistency | * Deceased Customers: Filter Out | * Default Column: Delete |  |
| **Customer**  **Address** |  | * Customer ID: Incomplete | * States: Inconsistency |  |  |  |
| **Transactions** | * Profit: Missing | * Customer ID: Incomplete * Online Order: Blanks * Brand: Blanks |  |  | * Cancelled Status Order: Filter Out | * List Price: Format * Product Sold Date: Format |

Here are detailed explanations of the data quality concerns we've identified, along with the employed strategies for addressing them. Additionally, included are suggestions and clarifications to prevent recurrent data quality problems. Implementing the ensuing recommendations will enhance the precision of the data, subsequently bolstering its effectiveness in guiding the future business decisions of Sprocket Central Pty Ltd.

**Accuracy Issues**

* **DOB was inaccurate for "Customer Demographic" and missing an age\_column; missing a profit column for "Transactions"**

***Mitigation:*** *Remove data anomalies in the DOB (Date of Birth) field.*

***Recommendation:*** *Introduce an* ***"age\_column"*** *to enhance data clarity and facilitate error detection. Establish a* ***"profit\_column"*** *within the "Transactions" category to verify the precision of sales records.*

Incorporating new columns for age and profit would simplify the process of error detection. The **profit\_column** will be particularly beneficial for forthcoming financial analyses.

**Completeness**

* **Additional customer\_ids were inconsistent among "Customer Demographic," "Customer Address,” and “Transactions"**

***Mitigation:*** *Apply a filter to include only* ***customer\_ids*** *ranging from 1 to 3500.*

***Recommendation:*** *Ensure consistency among the tables by using data from the same time*

*frame. For our model, exclusively utilize* ***customer\_ids*** *falling between 1 and 3500 due*

*to their comprehensive data availability.*

The data inconsistency across various spreadsheets introduces the possibility of biased

analysis due to incomplete information. This irregularity pertains to data completeness,

and to avert future instances, it is advised to cross-verify spreadsheets and synchronize

data.

* To address the issue, empty fields in the **"job\_title"** category within the **"Customer Demographic"** section should be filled. Similarly, in the **"Transactions"** section, empty spaces in the **"online\_order"** and **"brand\_column"** need attention.

***Mitigation:*** *Exclude any entries with missing values in the* ***"job\_title," "online\_order,"*** *and* ***"brand\_column"*** *fields.*

*Recommendation: Streamline the* ***"job\_title"*** *into an alternate category like* ***"industry\_industry"*** *or introduce predefined choices for* ***"job\_title."*** *Additionally, implement dropdown choices for* ***"online\_order"*** *and* ***"brand\_column."***

Blanks are treated as incomplete data and can skew further analysis results. The addition of dropdown options will allow to have more complete data and will result in more accurate analysis.

**Consistency**

* **Inconsistency in gender for “Customer Demographic” and “Customer Address” respectively**

***Mitigation:*** *Filter all* ***‘M’*** *under the category of* ***‘Male,’*** *and filter all* ***‘Femal’*** *and* ***‘F’*** *under* ***“Female”*** *for gender. Filter all* ***‘New South Wales’*** *to ‘****NSW’*** *and* ***‘Victoria’*** *to* ***‘VIC’*** *for states.*

***Recommendation:*** *Create dropdown options for* ***‘Male,’ ‘Femal,’*** *and* ***‘U’*** *in gender. Create dropdown options for all state abbreviations.*

Dropdown selections reduce the need for manual input and decrease the likelihood of human errors. They enable greater uniformity in the use of terminology. While implementing options for gender identity, it's important to be cautious due to the sensitivity of the topic.

**Currency**

* **People that are ‘Y’ in deceased\_indicator are not current customers for “Customer Demographic”**

***Mitigation:*** *Filter out customers checked* ***‘Y’*** *in* ***deceased\_indicator.***

***Recommendation:*** *Can be difficult to check for deceased customers, but once this information is received one should update data accordingly.*

‘Cancelled’ order\_status is irrelevant information for future analysis, as it can skew data – for example, a total number of customers per annum be an overestimate.

**Validity**

* **Format of list\_price, product\_sale\_date for “Transactions”**

***Mitigation:*** *Convert the* ***product\_sale\_date*** *into a concise date format and represent* ***list\_price*** *in currency format.*

***Recommendation:*** *Arrange columns in a way that predefined formats, like currency and decimal settings, are readily available for new data entry.*

Utilizing acceptable values will contribute to the clarity of data interpretation. Converting it to a price format and maintaining consistent placement of either 2 or 3 decimals will enhance readability. This modification will have a favourable effect on the speed and accuracy of business decision-related analyses.

This encapsulates the data quality issues unearthed during the preliminary analysis stage, accompanied by recommended mitigation strategies designed to enhance future analytical endeavours, benefiting both our internal and external analytical teams.

Should you seek further clarification on mitigation or the identified data quality matters, please do not hesitate to communicate with us.

Kind Regards,

Kagan Timur