Dear John Doe,

Thank you for providing us with the three datasets from Sprocket Central Pty Ltd. The summary table below highlights key quality issues that we discovered within the three data sets. Please let us know if you have any queries surrounding the issues presented.

**Summary Table:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Accuracy** | **Completeness** | **Consistency** | **Currency** | **Relevancy** | **Validity** |
| **Customer Demographic** | 1. DOB: inaccurate  2. Age: missing | 1. Job title: blanks  2. Customer id: incomplete | 1. gender: inconsistency | 1. Deceased customers: filtered out | 1. Default column: delete |  |
| **Customer Addresses** |  | 1. Customer id: incomplete | 1. states: inconsistency |  |  |  |
| **Transaction** | 1. Profit: missing | 1. Customer id: incomplete  2. Online Order: blanks  3. Brand: blanks |  |  | 1. Cancelled status order: filter out | 1. List price: format  2. Product sold date: format |

**Accuracy Issues:**

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

*Mitigation: Filter out outliers in* ***DOB****.*

*Recommendation: create an* ***age\_column****, allowing for more comprehensible data and easier to check for errors. Create a* ***profit\_column*** *for* ***“Transactions”*** *to check accuracy of sales.*

**Completeness:**

* **Additional customer\_ids were incomplete among “Customer Demographic”, “Customer Address”, and “Transactions”.**

*Mitigation: Filter out all* ***customer\_ids from 1 to 3500.***

*Recommendation: Ensure tables are up to date (from the same time period). For our model, only* ***customer\_ids from 1 to 3500*** *will be used as they have complete data.*

* **Blanks in job\_title for “Customer Demographic”, in online\_order and brand\_column for “Transactions”**

*Mitigation: Filter out ‘blanks’ for* ***job\_title, online\_order, and brand\_column.***

*Recommendation: Simplify job\_title to another category such as* ***industry*** *or provide dropdown options for* ***job\_title****, Provide dropdown options for* ***online\_order*** *and* ***brand\_column****.*

**Consistency**

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

*Mitigation: Filter all ‘M’ under category of ‘Male, 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.*

**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.*

**Relevancy**

* **Lack of relevancy or comprehensibility in default\_column for “Customer Demographic” and order\_status for “Transactions”**

*Mitigation: Delete Metadata in default\_column. Filter out ‘Cancelled’ order\_status.*

*Recommendation: Check for incomprehensible Metadata and delete or format to make comprehensible.*

**Validity**

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

*Mitigation: Format* ***product\_sale\_date*** *to short date format, format* ***list\_price*** *to currency.*

*Recommendation: Set up columns so that formats such as price and decimals are already in place when entering new data.*