Dear Manager,

Sprocket Central Pty Ltd,

Thank you for providing the three datasets from Sprocket Central Pty Ltd. The summary table below highlights key quality issues that were discovered within the three datasets. Please let me know if you have any queries surrounding the issues found.

**Summary Table:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Accuracy** | **Completeness** | **Consistency** | **Currency** | **Relevancy** | **Validity** |
| **Customer Demographic** | * DOB: Inaccurate * Age Missing | * Job Title: Blank * Customer ID: Incomplete | * Gender: Inconsistency | * Deceased customers: Filter out | * Default Column: Delete |  |
| **Customer Address** |  | * Customer ID: Incomplete | * States: Inconsistency |  |  |  |
| **Transactions** | * Profile Missing | * Customer ID: Incomplete * Online order: Blanks * Brand: Blanks |  |  | * Cancelled Status Order: Filter out | * List price: Format * Product sold data: Format |

Below is more in-depth description of data quality issues found and methods of mitigation used.

**Accuracy issues:**

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

Mitigation: Filter out outlier in DOB

**Completeness**

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

Mitigation: Filter all customer\_ids from 1 to 3500

* Blanks in job\_title 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.

**Currency**

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

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

**Relevancy**

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

Mitigation: Deleted metadata in default\_column. Filter out “cancelled” order\_status.

**Validity**

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

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

That sums up all the data quality issues discovered through the first stage of the data quality analysis. The mitigation strategies suggested are simple and effective ways of improving data quality for future analysis. They will not only improve the analysis output that one can perform within the

company but will increase the level of analysis that can be performed by KPMG and other hired analysis and other hired analysis teams.

Please let me know if you’ve queries regarding mitigation or any data quality issues identified.

Kind regards,

Sunil Kumar