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 concerning the issue presented.

**Summary Table**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Datasets** | **Accuracy** | **Completeness** | **Consistency** | **Currrency** | **Relevancy** | **Validity** |
| **Customer**  **Demographic** | * DOB:   Inaccurate   * age:   Missing | * job\_title:   blanks | * gender:   Inconsistency | * deceased\_indicator:   Filter out | * default column:   Delete |  |
| **Customer Address** |  |  | * state:   Inconsistency |  |  |  |
| **Transactions** | * profit:   Missing | * online\_order:   blanks   * brand:   blanks |  |  | * order\_status:   Cancelled filtered out | * list\_price: format * product\_first\_sold\_date:   format |

Below are more in-depth descriptions of data quality issues discovered and methods of mitigation used. Recommendations and explanations have also been included to avoid further data quality issues in the future. The following recommendations will improve the accuracy of data used to influence the business decisions of Sprocket Central Pty Ltd in the future.

**Accuracy**

* **DOB was inaccurate for “Customer Demographic” and missing an age column, missing a profit column for “Transactions”.**

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

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

**Completeness**

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

**Mitigation:** Filter out “blanks” for **job\_title, online\_order, 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.

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 Demograhic” 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 a dropdown option for ‘Male’, ‘Female’ and ‘U’ in **gender**. Create dropdown options for all **state** abbreviations.

Dropdown options minimize manual entry and human error. Allows for increase of consistency of terminology. Gender identity can be a sensitive topic, proceed with caution when creating options.

**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.

Deceased customers are not current customers, removing them from data will increase the currency of data and will result in more accurate estimates in future analysis.

**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**.

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

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

**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.

Allowable values will make data to be interpreted more easily. Formatting into price and allowing for either 2 or 3 decimals placed consistently will increase readability. This will reflect positively on speed and accuracy of analysis for business decisions.

That summarizes all 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 teams.

Please let us know if you have questions regarding mitigation or any data quality issues identified.

Kind regards,

Arun S