Dear client,

Thank you for providing us with the three datasets from Sprocket Central Pty limited. The summary table below highlights key Quality issues that we discovered within the three datasets. Please let us know if you have any further queries regarding issues presented.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Accuracy** | **Completeness** | **Consistency** | **Currency** | **Relevance** | **Validity** | **Uniqueness** |
| **Customer demographic** | **DOB:**  **Inaccurate**  **Age: missing** | **Job title: Blanks**  **Job Industry: Blanks** | **Gender.: Inconsistency.** |  | **Default column: delete** | **Deceased customer: Filter out.** |  |
| **Customer address** |  | **Last Name: Null Values** | **States.: Inconsistency.** |  |  |  |  |
| **Transactions** | **Profit: missing** | **Online order:**  **Blanks.**  **Brand:**  **Blanks.** |  |  |  | **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. Following recommendations will improve accuracy of data used to influence business decisions of Sprocket Central Pty Ltd in the future.

***Accuracy issues:***

* ***DOB was inaccurate for “customer demographic” and missing an age column missing a profit column for “transactions”***

***Mitigation:***

* Create an **age column**, allowing for more data and easier to check for errors. Create a **profit column** in **transactions** to check accuracy of sales.
* Creating additional columns for age and Profit will allow for easier identification of errors. The **Profit column will** assist in future monetary analysis.

***Completeness issues:***

* ***Few columns like Job title Followed by job industry in “customer demographic” Last name in customer Address and Online order Followed by brand in “transactions” have a lot of null values.***

***Mitigation:***

* Filter out or Provide dropdown options for Columns tables mentioned above.
* Blanks are treated as incomplete data and can skew further analysis results. The addition of dropdown options will allow to have more complex data and will result in more accurate analysis.

***Consistency issues:***

* ***Inconsistency in gender and states for “customer demographic” and “customer address”.***

***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**.

***Relevancy issues:***

* ***Lack of relevancy. In default column for “customer demographic”.***

***Mitigation:***

* Deleted metadata in default column.

***Validity issues:***

* ***People that are Y in Deceased Indicator are not current customers. For “customer demographic”***
* ***Format of product for sold date for “transactions”.***

***Mitigation1:***

* Filter out customers checked with deceased indicator.
* 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 currency of data and will result in more accurate estimates in future analysis.

***Mitigation 2:***

* Format product sold date to short date format.
* Set up columns so that formats such as price and decimals are already in place while entering new data.
* Allowable values will make data to be interpreted more easily. Formatting into price and allowing for either two or three decimals placed consistently will increase readability. This will reflect positivity on speed and accuracy of analysis for business decisions.

This summarizes 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 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 me know if you have questions regarding mitigation or any data quality issues identified.

Kind regards,

Manasa.