Dear John Doe,

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

**SUMMARY TABLE:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Accuracy** | **Completeness** | **Consistency** | **Currency** | **Relevancy** | **Validity** |
| **Customer**  **Demographic** | +DOB:  Inaccurate  +Age  Missing | +Job title:  Blank  +Customer ID:  Incomplete | Gender:  Inconsistency | Deceased  Customers: delete | Default  Column:  delete |  |
| **Customer Address** |  | +Customer ID:  Incomplete | State:  Inconsistency |  |  |  |
| **Transactions** | +Profit:  Missing | +Customer ID:  Incomplete  +Online Orders:  blanks  Brand: blanks |  |  | Cancelled  Status order:  Filter out | List price: format  Product sold date: format |

Below are further descriptions of data quality issues discovered and methods of mitigation used. Furthermore, recommendations and explanations have been included to avoid further data quality issues in the future. Following recommendations will enhance 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:** Filter out outliers in DOB

**-Recommendations:**

* Create an age column for more comprehensive datasets, which may facilitates checking for errors.
* Create a profit column in “Transactions” to check accuracy of sales

Creating these two columns will allow easier identification of errors. Moreover, the profit column will assist monetary analysis in the future.

**Completeness Issues:**

* **Additional customer ids were inconsistent among “Customer Demographic”, “Customer Address” and “Transactions”**
* Mitigation: Filter 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 can be used as they have complete data

The data receive may not be in a sync across all spreadsheets, with incomplete data, the analysis results may be skewed.

* **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.

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

**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 category of “Female” for gender. Filter all “New South Wales” to “NSW” and “Victoria” to “VIC” for states
* Recommendation: Create dropdown options for “U”, “Femal” and “Male” in gender. Create dropdown options for all state abbreviations:

Dropdown options will minimize manual entry and human error, which allows the increase in consistency of terminology. However, gender identity can be a controversial topic that requires consideration in creating options.

**Currency:**

* **People that are “Y” in decease indicator are not current customer for “Customer Demographic**”
* Mitigation: Filter out all customers checked “Y” in decease indicator.
* Recommendation: It 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, which makes removing them from data will increase currency of data and will result in more accurate data analysis in the future.

**Relevancy:**

* **Format of list price, product sale date for “Transactions”:**
  + Mitigation: Format product sale date to short date format, format list price to currency format.
  + Recommendation: Set up columns so that format such as price and decimals are already in place when entering new data.

All available data will make data become more comprehensive. By formatting into price and allowing for either 2 or 3 decimals placed consistently will increase readability.

That summaries all data quality issues discovered through the first stage of data quality analysis. The mitigation strategies suggested are simple and effective ways of improving data quality for further analysis. These recommendations will not only improve the analysis output that one can perform within the company but also enhance the level of analysis that can be executed by KMPG and other hired analysis teams.

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

Best regards,

Hang Dinh