Dear Client,

Thank you for making available Sprocket Central Pty Ltd datasets. The table below contain the summary and highlights key quality issues that was discovered within the datasets from table provided.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table category | Accuracy | Completeness | Consistency | Currency | Relevancy | Validity |
| Customer Demographic | 1.DOB: Inaccurate  2.Age: Missing | 1.Job title: Blanks  2.Tenure: blanks | 1.Gender: Inconsistency | 1.Deceased indicator: Deceased customer filtered out | 1.Default column: Deleted |  |
| Customer Address |  |  | States Column: inconsistency with the attribute |  |  |  |
| Transactions | Profit: Missing | 1.Order status: blank  2.Brand: blanks |  | 1.Cancelled order status: filtered out |  | 1.list price: correct format  2. product first sold date: correct format |

Below are detail descriptions of data quality issues discovered and also mitigation approached deployed in order to avoid future data quality issues. The recommendations below will also provide an improved data accuracy which will be used to influence business decision of Sprocket Central Pty Ltd for future.

**Accuracy Issues:**

1. **Customer Demographic**
2. DOB was inaccurate
3. Column for Age not included
4. Transactions
5. Column to calculate Profit not included

Mitigation: Filter out outlier in DOB

Recommendation: Inserted an Age Column, which allow for easy comprehension of data. Create a column for profit in Transactions to check accuracy of sales

Creating these additional columns will allow for easy identification of errors while the profile added will allow for future analysis.

**Completeness**

Blanks in job title and tenure for Customer Demographic, online order and brand for transactions.

Blanks are treated as incomplete data so therefore they were removed.

**Consistency**

Mitigation: for gender in Customer Demographic, Filter all M, Femal, and F under the right category Male and Female. Filter all new south Wales to NSW and Victoria to VIC for state in Customer Address.

Recommendation: create a fixed option like a dropdown for gender to maintain consistency because such option reduce errors and eliminate manual entry.

**Currency**

Mitigation: Filter out customers that are checked “**Y”** because they are deceased.

Recommendation: Unchecking **Y** and leaving **“N“** only customers that are alive simplify the data and increase accuracy.

**Relevancy**

Mitigation: deleted the default column. Filter out “cancelled” in order status

Recommendation: deleting the default column makes it easy to understand and comprehend the data

**Validity**

Mitigation: Format both product sale date and list price to their appropriate format, short date for product sale date and currency for list price.

Please feel free to reach out if you have any questions as regarding the mitigation and the data quality issues.

Kind regards.

It is important to keep in mind the business context when presenting your findings:

* What are the trends in the underlying data?
* Which customer segment has the highest customer value?
* What do you propose should be Sprocket Central Pty Ltd ’s marketing and growth strategy?
* What additional external datasets may be useful to obtain greater insights into customer preferences and propensity to purchase the products?
* Total sales/ profit by customer title, b. wealth segment / age group (where profit will be average)
* Customer distribution based on customer Title by customer id and wealth segment (Rank) %.
* Bike purchase in last 3 years by gender and profit
* Wealth segment by title and profit
* Profit(average) by customer title per month
* Profit by state
* Customer title distribution by state
* List of valuable(profit) by FLname,state

Specifically, your presentation should specify who Sprocket Central Pty Ltd’s marketing team should be targeting out of the new 1000 customer list as well as the broader market segment to reach out to.