Dear Client,

Thank you for providing us with the Sprocket Central Pty Ltd datasets. In the table below are highlighted the key issues within the datasets.

**Summary Table**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Accuracy** | **Completeness** | **Consistency** | **Currency** | **Relevancy** | **Validity** |
| **Customer Demographic** | DOB: Inaccuracies | Job title:  Blanks | Gender:  Inconsistency | Deceased customers:  Not included | Default:  Column deleted |  |
| **Customer Address** |  | Customer id:  incomplete | States:  Inconsistency |  |  |  |
| **Transactions** |  | Online Order: Blanks  Brand: Blanks |  |  | Canceled status order:  Filtered out | List price: Reformatted  Product sold date: Reformatted |
| **New Customer List** |  | DOB: Blanks |  |  |  | Past 3 years bike related purchases:  Reformatted  Postcode:  Reformatted  Property value:  Reformatted |

Below is a more in-depth look at the issues discovered. Furthermore, there are included methods of mitigation and recommendations for future improvements to increase accuracy of the data.

**Accuracy Issues**

**Accuracy issues with *DOB.***

Filtered out outlier in *DOB*.

Recommendation: Add a “**profit”** column to “**Transactions’** to improve financial analysis.

**Completeness**

* ***Customer ids* were inconsistent among ‘’Customer Demographic”, “Customer Address”, and “Transactions.”**

Filtered all *customer ids* from 1 to 3500 to ensure we work with the same values.

Recommendation: Ensure tables are up to date to avoid discrepancies among spreadsheets.

* **Blanks in *job title, online order and brand column*.**

Filtered out the ‘blanks’.

Blanks can have undesirable effects on analysis such as skew, therefore, for simplification and completeness we remove them from the pool.

**Consistency**

**Inconsistencies in ‘gender’ columns and ‘states’.**

All ‘M’ turned into ‘Male’ and all ‘F’ into ‘Female’. Additionally, removing typos such as Female. Also filtering states ‘New South Wales’ to ‘NSW’ and ‘Victoria’ to ‘VIC’.

Recommendation: Make the selection of gender and possibly state selectable from a type of selection like a dropdown menu to minimize human error.

**Currency**

**People indicated as ‘Y’ for deceased are not current customers for “Customer Demographic”.**

Deceased clients filtered out.

**Relevancy**

**Incomprehensible data in *default* for “Customer Demographic” and *order status* for “Transactions”.**

Deleted the incomprehensible columns as they do not provide relevant information.

Recommendation: Check for incomprehensible data and check the formatting as it might be the source of the issue.

Also, ‘Cancelled’ for orders is irrelevant information for future analysis.

**Validity**

**Issues with the formatting for a number of columns.**

‘List price’, ‘Product sold’, ‘Past 3 years bike related purchases’, ‘Postcode’ and ‘Property value’ reformatted to fit the data types ‘number’, ‘date’, ‘currency’ accordingly to the data.

Recommendation: Make the input of such data be registered in formats such as price and make decimals available where relevant to improve the speed and accuracy of analysis.

That is the summary of all the quality issues discovered and the actions taken to remove them in order to improve the analysis. With the addition to some recommendations that could have a beneficial impact to the speed and accuracy of future analysis performance.

Please let us know if the are any queries regarding this first step of the analysis.

Yours sincerely,

Kaloyan Popdonchev