Hello,

Thank you for providing us with the three datasets from Sprocket Central Pty Ltd. The Summary table below highlights the issues that were discovered within the 3 datasets. Please let us know any queries surrounding the issues presented.

**Summary Table**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table Name | **Accuracy** | **Completeness** | **Consistency** | **Relevancy** | **Uniqueness** | **Validity** |
| Customer Demographic | DOB: Inaccurate  Age: Missing | DOB: Blanks  Job title: Blanks  Tenue: Blanks  Last Name: Blanks | Gender: Inconsistency  Job Title: Inconsistency | Default Column: Deleted |  | DOB: Age greater than 100 |
| New Customer List |  |  | Postcodes & Property Valuation: Values have different Datatype | Hidden Sheets | Property Valuation: Duplicate Values |  |
| Customer Address |  |  | State: Inconsistency |  |  |  |
| Transaction Data | Profit: Missing | Product Line, Product Class, Product size, Standard Cost: Blanks |  |  |  | Date: Values not Sorted |

Below are in depth description od data quality issues with mitigation methods used. Furthermore, recommendations have been provided to avoid the reoccurrence

of data quality issues and improve the accuracy of the underlying data used to drive business

decisions.

**Accuracy**

Customer Demographics table have **inaccurate DOB** and **Age missing.** Transaction Data has **Profit missing**.

*Mitigation: Filter out outliers in DOB column*

*Recommendation: Create Profit column to check accuracy in sales and create an Age column as well. These both columns can further assist us in analysis.*

**Completeness**

Lot of **Blank values** were seen in Customer Demographics and Transaction sheet.

*Mitigation: If small no. of records is missing then, we can filter out those records.*

*Recommendation: Filter out records*

**Consistency**

Customer Demographic, New Customer List and Customer Address have inconsistent data.

*Mitigation: In Gender, State and Job Title column values were replaced.*

*For example- Victoria being represented as “V”, “Vic” and “Victoria”. All values were replaced to Vic.*

*Recommendation: Enforce a drop-down list for the user entering the data rather than a free text field. Ensure that fact tables in the given database have constraints on data types.*

**Relevancy**

Default column in Customer Demographic has irrelevant data consist of symbols.

*Mitigation: Deleted Default column*

*Recommendation: Check for incomprehensible columns in data, delete or format them to make it comprehensible*

**Uniqueness**

Property valuation column in New Customer List has some Duplicate values.

*Mitigation: Replaced duplicate values*

**Validity**

DOB column in Customer Demographic has an outlier, age was greater than 100 years. Date values in Transaction sheet was not according to business standards.

*Mitigation: Filtered out the outlier and set date to short date format.*

*Recommendation: Setup columns so that formats are already in place when entering a value.*

**More Recommendations to avoid Data Issues**

**Electronic Forms**

We can create electronic forms for customer surveys. It makes it harder to skip a critical question on an electronic survey than on a paper-based one. And you also save a ton of time by avoiding digitizing your collections.

**Create data quality reports & dashboards**

Off the back of your data profiling set, measure data quality KPIs based on the data quality dimensions relevant to your data set and needs. Build a data quality dashboard to continuously monitor the trend of quality of the data entering your reporting and analytics environment.

**Create an issues log**

Create, maintain and report the outcomes of a data quality issues log. This enables you to document issues, create preventative measures and confirm the effectiveness of data cleansing processes.