**Subject** : Acknowledgment of Datasets Received and Data Quality Concerns

Dear Sir (Client),

I sincerely appreciate your prompt action in sending us the datasets (Transactions, Customer Address Customer Demographic, and New Customer List) from Sprocket Central Pty Ltd. We have received the files successfully, and I'd like to express my gratitude for your cooperation.

**Data Quality Dimesnsions and errors in respective sheets :**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Customer Demographics** | **Transactions** | **Customer Address** |
| **Accuracy** | Gender : 92 corrupted values  and outliers,  Missing age Column | Missing Profit Column |  |
| **Completeness** | Last name : 126 Blanks ,  DOB : 88 blanks,  Job\_title : 507 blanks  Tenure : 88 blanks | Online\_order : 361 blanks,  Brand, Product\_line, Product\_class,  Product\_size ,Standard\_cost, Product\_first\_sold\_date : 198 blanks |  |
| **Consistency** | Gender : Inconsistency |  | State : (VIC,NSW,QLD) 3822 rows with abbreviations for state |
| **Currency** | Deceased Indicator : 2 are dead |  |  |
| **Relevancy** | Default : irrelevant data | Product\_first\_solddate : Incorrect datatype |  |
| **Validity** | Job\_Industry\_category : 655 n/a | List\_Price : Currency missing, |  |
| **Uniquenes** |  |  |  |

**Description and Recommendations :**

**a) Customer Demographics :**

**1)Accuracy :**

92 corrupted values present in “Gender”column with some outliers. Remove the corrupted values.

Age column must be there for further analysis which was missing in the data. Generate age column from DOB using functions.

**2) Completeness :**

126, 88, 507 and 88 blanks in “last name, DOB, Job title and tenure” column respectively. Remove the blanks completely .

**3) Consistency :**

Gender can be either male or female . Need to remove the outliers and replace “M” by “Male”, “F and Femal” by “Female”

**4) Currency :**

Deceased Indicator implying 2 customers are dead. This data is of no use ,we can delete that.

**5) Relevancy :**

Column “ Default” contains irrelevant data which is of no use. Delete complete column.

**6) Valididty :**

“n/a” is present in “job industry category” 655 times. For insights generation this data is of no use , so we can delete all.

**b) Transactions :**

**1) Accuracy :**

Missing “profit” Column which is very necessary for analysis. Generate profit column via (list price – standard\_cost) .

**2) Completeness :**

361 blanks in “online order” and 198 blanks in “ Bramd, Product Line, Product Size, Product Class, Standard cost and Product first sold date” column

**3) Relevancy :**

Incorrect datatype in “product first sold date” column. Convert datatype to date .

**4) List Price :**

Currency missing in “List Price” , add dollar sign to column.

**C) Customer Address :**

**1) Consistency :**

Inconsistent data as State column contain different values for same states. Replace the abbreviated value to their full forms.

We value your expertise, and if you have any additional insights or different opinions regarding these data quality issues, please don't hesitate to contact us. Your input will be highly valuable as we work towards ensuring the accuracy and reliability of the data.

We greatly appreciate your cooperation and support in this endeavor.

Best regards,

**Nayan Srivastava**