

To

Sprocket Central Pty Ltd

Thank you for reaching out to us. After analysing the three datasets you provided for various data quality issues, here is what we found, we have provided you with both types of errors and mitigation that will help you to maintain your data quality and hence enhance your business growth.

## ***DATASET TABS***

### **❖ Transactions --**

Accuracy: I have used **Data Validation** function in excel to validate the data added in different columns. I have found inaccurate data in various columns. Each column has a particular format for entering the data, I have validated the columns corresponding to there format. Product\_id is 0 in some cells.

Completeness: To find no. of blank cells in the dataset, I used function COUNTBLAN function and found there are 1542 blank cells in Transactions.

Validity: product\_first\_sold\_date\_ is in number format instead of date format.

### **❖ Customer Demographic---**

Accuracy: DOB format is inaccurate.

Completeness: There are total 1045 blanks found.

Consistency: Gender data has inconsistency found. Gender is represented by various abbreviations like (F, Female, U, Femal, M)

Validity: default column contains random values which are not allowed.

### **❖ Customer Address –**

Accuracy – Cells contain inaccurate data. In customer\_id column, in row no. 5 and 10, this has affected the total count of data.

Consistency - state data has inconsistency found.

[New South Wales, QLD, VIC, NSW, Victoria]

Below we have provided detail description of findings and mitigation methods and suggestions,

1. Accuracy (Data Validation Function can be used)
  - a. Date of birth data is not accurate there are some outliers
    - i. **Mitigation: remove outliers**
    - ii. **Suggestions: create new feature age, age feature will help to easily understand errors**
2. Completeness (We can use COUNTBLANK() function in excel.
  - a. There are some missing data
    - i. **Mitigation: apply missing value treatments**

- b. customer\_id is incomplete among customer\_demographic and customer\_address
    - i. **Mitigation: use only records which are available on both sheets and filter out all other incomplete data or update spreadsheets with correct data**
- 3. Consistency
  - a. Inconsistency gender data in customer\_demographics
  - b. Inconstancy state data in customer\_address
    - i. **Mitigation: replace data to remove inconsistency**
- 4. Validity
  - a. There are some records with product\_id = 0 also check for any random values in a column.
    - i. **Mitigation: check database whether there is 0 product\_id and random characters in the column, create an alert in such case using create alert tab in data validation to prevent such blunders.**