

Dear XYZ,

Thankyou for providing us with the three datasets from Sprocket Central Pty Ltd.  
The summary table below highlights key quality issues that we discovered within the three datasets.Please let us know if you have any queries surrounding the issues presented.

## SUMMARY TABLE

	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
<b>Customer Demographic</b>	1.DOB: Inaccurate 2.Age: Missing	1.Job title: Blanks 2.Customer ID: Incomplete	1.Gender: inconsistency	1.Deceased Customers: Filter out	1.Default Column: Delete	
<b>Customer Address</b>		1.Customer ID: incomplete	1.States: inconsistency			
<b>Transactions</b>	1.Profit: Missing	1.Customer ID: Incomplete 2.Online Orders: Blanks 3.Brand: Blanks			1.Cancelled Status order: Filter out	1.List Price: Format 2.Product sold date: Format

Below are more in depth descriptions of data quality issues discovered and methods of mitigation used.Recommendations and explanations are also been included to avoid further Data quality issues in future.Following recommendations will improve accuracy of data used To influence business decisions of Sprocket Central Pty Ltd in the future.

## **ACCURACY ISSUES:**

- **DOB was inaccurate for “Customer Demographic” and missing an age column and missing a profit column for “Transactions”.**

### **MITIGATION:**

Filter outliers in DOB.

Create an age column for more comprehensible data and easier to check for errors. Create a profit column to check accuracy of sales.

## **COMPLETENESS:**

- **Additional Customer IDs were inconsistent among “Customer Demographic,” and “Customer Address” and “Transactions”.**

### **MITIGATION:**

Filter all customer ids from 1 to 3500.

The data received may not be in sync across all spreadsheets, with incomplete data the Analysis results may be skewed. This is ‘Completeness’ issue, to prevent future occurrences it is encouraged to cross check spreadsheets and sync data.

- **Blanks in job title for “Customer Demographic” and for online\_orders and brand\_column for “Transactions”.**

### **MITIGATION:**

Filter out blanks for job\_title, online\_orders and brand\_column.

Blanks are considered as incomplete data and can skew further analysis results.

## **CONSISTENCY:**

- **Inconsistency in gender for “Customer Demographic” and “Customer Address respectively”.**

### **MITIGATION:**

Filter all ‘M’ to Male and ‘F’, ‘Femal’ to ‘Female’ for gender.

Filter all ‘VIC’ to ‘Victoria’ and ‘NSW’ to ‘New South Wales’.

Create dropdown options for all state abbreviations .It minimize human error.

### **CURRENCY:**

- **People that are “Y” in deceased\_indicator are not current customers for “Customers demographic”**

#### **MITIGATION:**

Filter out all customers checked “Y” in deceased\_indicator.

Deceased customers are not current customers,removing those customers will increase the currency of data.

### **RELEVANCY:**

- **Lack of relevancy in default\_column for “Customer Demographic” and order\_status for “Transactions”**

#### **MITIGATION:**

Filter out “Cancelled” order status.Deleted metadata in default\_column.

“Cancelled” order\_status is irrelevant information for future analysis - as it can skew data.

### **VALIDITY:**

- **Format of list\_price and product\_sale\_date for “Transactions”**

#### **MITIGATION:**

Format product\_sale\_date to short date formats,format list\_price to currency.

Allowable values will make data to be interpreted more easily.Formatting into price and allowing for either 2 or 3 decimals placed consistently will increase readability.This will reflect positively on speed and accuracy of analysis for business decisions.