

Dear John Doe,

Thank you for providing us the three data sets from Sprocket Central Pty LTD. There were a lot of problems with these data sets but I have cleaned them up.

But still I mentioning some of the flaws of your data set so that you can take care of them in future.

### **1. Customer Demographic Dataset**

- **Accuracy**
  - DOB was inaccurate
- **Completeness**
  - Ages were containing missing values
  - Job titles were missing
  - Job Industry category also contains missing values
- **Consistency**
  - Gender was inconsistent
- **Currency**
  - Removed the Deceased customers
- **Relevancy**
  - Default column deleted

### **2. Customer Address Dataset**

- **Consistency**
  - State column was inconsistent

### **3. Transactions Dataset**

- **Accuracy**
  - Profit is missing
- **Completeness**
  - Online order and brand columns contains missing values
- **Relevancy**
  - Customers that cancelled orders were removed
- **Validity**
  - Product sold date format was not up to the mark

*Recommendations:*

- 1. Outliers in the DOB are acceptable at your end as they can be removed by us easily but we cannot do anything of Ages missing, so make sure that you get age of every customer, as from ages we can find a lot of valuable interpretations and we also need at least Job industry category column to be filled by each and every customer.*
- 2. We don't need deceased customers, so try not to include deceased customers in your dataset.*
- 3. Do not include the columns that are of no need (like default column).*
- 4. Try to give the more detailed data set like adding the profit column.*
- 5. Major problem is with the inconsistent dataset seen in state column in Customer address dataset and Gender column in Customer Demographic dataset, so try to get only one key word for one category.*
- 6. Do not include the customer that had cancelled orders.*

*NOTE: Completeness problem were in the other columns too, but I only mentioned that were important to me or that were important to analyse data.*

**Thank You,**

**Sameer Sharma**