

Dear Sprocket Central team,

Greeting, I am a new intern student from Analytics team. Thank you for trusting and giving us the opportunity to work on your dataset. Below are the table of summary statistic from the three datasets received.

	<b>Customer Demographic</b>	<b>Customer Address</b>	<b>Transaction Data</b>
<b>Accuracy</b>	<ul style="list-style-type: none"><li>• DOB: inaccurate</li><li>• Age: missing</li></ul>		<ul style="list-style-type: none"><li>• Profit: missing</li></ul>
<b>Completeness</b>	<ul style="list-style-type: none"><li>• Customer ID: incomplete</li><li>• Last name: blanks</li><li>• Job title, industry: blanks</li></ul>	<ul style="list-style-type: none"><li>• Customer ID: incomplete</li></ul>	<ul style="list-style-type: none"><li>• Customer ID: incomplete</li><li>• Online order: blanks</li><li>• Brand: blanks</li></ul>
<b>Consistency</b>	<ul style="list-style-type: none"><li>• Gender: inconsistency</li></ul>	<ul style="list-style-type: none"><li>• State: inconsistency</li></ul>	
<b>Currency</b>	<ul style="list-style-type: none"><li>• Deceased customer: some are deceased</li></ul>		
<b>Orderliness</b>		<ul style="list-style-type: none"><li>• Postcode: format general to number</li><li>• Property value: format to number</li></ul>	<ul style="list-style-type: none"><li>• List price: format number to currency</li><li>• Product first sold: format number to date</li></ul>
<b>Relevancy</b>	<ul style="list-style-type: none"><li>• Default column: hided</li></ul>		<ul style="list-style-type: none"><li>• Order status: Cancelled filtered out</li></ul>
<b>No. of Record</b>	4000	3999	3494
<b>Date Received</b>	2021-08-14	2021-08-14	2021-08-14

I have carefully looked through the dataset that you have provided for us, the data does demonstrate a good potion that will actually help us in the next analysing phases. However, with the dataset you have provided, there are a few quality problems within, also mitigations and a few recommendations are provided under each issue stated. Please let us know if the findings below are not aligned with your understanding.

**Accuracy:**

- i. **Inaccurate DOB in Customer Demographic table which result in unrealistic interpretation.**

*Mitigation:* Filter out the inaccurate DOB (outlier).

*Recommendation:* When collecting the meta-data, use a predetermined date format and range of date for more accurate selection. For this, dropdown-list or election on calendar would be a good tool.

ii. **Missing Age column for Customer Demographic and Profit column for Transactions**

*Recommendation:* Create the Age column for more comprehensible data and Profit column for easier check in income and profitable.

**Completeness:**

i. **Blank columns happened to be the common problem in various tables for all three datasets (such as job-title, last name, brand of the bicycle).**

*Mitigation:* Filter out the blanks if not necessary, otherwise, try to fill the blanked columns with based on relevant information that is distributed in the same dataset.

*Recommendation:* I assumed these are customer's preference not to provide the information, if not, this might because of the issue with data collector or else. What the data collector could do is to make sure all the questions they are asking to have the appropriate answer or make the question as required if that is in the type of survey or form. On the other hand, the automated data collection with autofill is also a good approach to eliminate missing value.

ii. **There are Customer IDs missing from all three datasets.**

*Mitigation:* Filter out customer ids that are missing from other tables, and left only those that are integrated in all tables. In this case we only keep customer ids from 1-3400.

*Recommendation:* Ensuring each table is up to date and collected with the same unique key values from the same period of time.

**Consistency:**

i. **There are inconsistent data within the gender row from Customer Demographic and the state name from Customer Address, they are not consistent which result in multiple values refer to the same attribute.**

*Mitigation:* These issue can be easily solved by setting a clear and consistent word to be referred to for each specific gender, using Replace too, Filter or VLOOKUP in excel.

*Recommendation:* When collecting such data, utilize the drop-down list instead of open text field for those questions with a few certain answers.

**Currency:**

i. **There are customers who is deceased but not removed from the dataset.**

*Mitigation:* Filter the outlier and keep only current or active customers.

*Recommendation:* It is recommended that the team could spend time cross-checking the table and remove inactive or outdated information, which would be good for more effective future analysis.

### **Orderliness:**

i. **Inconsistent Data type for the attribute.**

*Mitigation:* The postcode and property valuation do come with a little inconsistency within the format that should be altogether set as number. The value in product first sold date come with an incorrect format which should be converted into the readable date format instead of counted dates.

*Recommendation:* Make sure the tables have the right data type before or after entering the data.

### **Relevancy:**

i. **There is an inaccurate row of data that contain meaningless and uninterpretable meta-data, which is the default column.**

*Mitigation:* This should be removed carefully when reviewing the data initially as this row of data does not add any values to our analysis.

*Recommendation:*

ii. **From Transactions, the Cancelled order status should be removed.**

*Mitigation:* Filter out the Cancelled in order status to avoid overestimate for business purpose.

*Recommendation:* Set up the column to automatically filter out the cancelled order.

**The above issues are what I have figured out are the key to enhances our data analytics quality, and for this reason, we may wish to make correction or execute our mitigation strategies as this will ensure that we have the correct and meaningful data for our storytelling analysis and to help provide a clear, accurate and effective strategy recommendation for the growth of your business.**

**Kind Regards,  
Kimnay Oeng**