

Dear Sir/Madam,

We appreciate your confidence in us and also for furnishing us with the three datasets for analysis. Upon assessing the quality issues with the data presented, we identified some errors. This is presented in the table below for easy identification.

	TRANSACTION	CUSTOMER DEMOGRAPHIC	CUSTOMER ADDRESS
ACCURACY	1.Profit is not calculated  2. Some Product ID's are zero which is incorrect	1. Some DOB are Inaccurate 2. Age column is missing.	
COMPLETENESS	1.Customer Id is incomplete (3501-5033) 2.Online order has blanks in some spaces. 3.Brand has blanks 4. Standard cost has blanks	1.Job title has blanks 2. Customer Id is incomplete 3.Tenure has blanks	1.Customer Id is incomplete
CONSISTENCY		1.Gender is inconsistent.	1. State is inconsistent
CURRENCY		1.Deceased customers contains the dead.	
RELEVANCY	1.Order Status cancelled not relevant	1.Default Column not relevant	
VALIDITY	1.The List price did not contain any currency sign. 2.Product sold date is in the wrong format.		

Below are the explanations of the data quality issues discussed above. Recommendations and explanations have also been listed to assist in the corrections.

#### **ACCURACY ISSUES**

There is supposed to be a profit column under the **transactions** dataset. Some of the **product Id** are zero which is not supposed to be. Some **DOB** are also not correct for **Customer Demographic** and there is no age column. Which can make things clearer and easier.

**Recommendations-** Create a profit column in transactions for easy monetary analysis and correct the **product Id's** which are zero. Creating an age column will help check errors for the DOB.

### COMPLETENESS

**Customer Id's** throughout the three datasets are inconsistent. **Customer address** has **customer id** values up to 4000 while **transactions** goes further to 5000.

**Recommendations-** The **customer id** datasets presented must be equal to each other throughout the 3 datasets. This will give a more accurate analysis. Be sure to check the **customer id's** within the 3 datasets. The blanks in the **online order, brand** and **standard cost** in the transaction dataset need to be filled. The blanks in the **Job title** and **tenure** should also be filled under the **customer demographic**.

### CONSISTENCY

The words used for the **gender** under the **customer demographic** is inconsistent (female, male, u, m and f). For the **state** under the **customer address**, sometimes the short contraction is used while others use the full names.

**Recommendation-** under the **gender** column, either use the full words like female and male or choose the contraction which is M or F. They should not be mixed, its either the contraction form is used or the full words are used.

Under the **state** column too, either the full names of the states are used or the contraction form is used throughout.

### CURRENCY

Under the **deceased** column, the Y' must be removed because they are dead and are not current customers.

**Recommendation:** once the status of customers are updated that they dead, their information needs to be removed to give an accurate information.

### RELEVANCY

There is no comprehension in the **default** column in the **customer demographic** and needs to be removed. The cancelled **order status** under the transaction also needs to be deleted as it does not help analysis in anyway.

**Recommendation:** The cancelled **order status** needs to be deleted and the entire **default column** needs to be deleted.

### VALIDITY

Turn the **product\_first\_sold\_date** to a real date and add the currency sign to the list price. This will assist in easy handling of the data.

**Recommendation-** Add the currency sign to the list price and make a real date out of the **product\_first\_sold\_date**.

To conclude, the recommendations suggested will assist in improving the data quality issues identified. This will increase the quality of the analysis which will give the maximum benefit to your company.

Do not hesitate to contact us should you have questions on the data quality issues identified or recommendations.

Sincerely,  
Barbara Zen.