

Data Quality Assessment Report.

Dear Client,

We have performed the Data Quality Assessment on the three datasets of Sprocket Central Pty Ltd. The findings below highlight some data quality issues on the three datasets and how they can be mitigated.

		Data Quality Assessment Dimensions						
		Accuracy	Completeness	Consistency	Currency	Relevancy	Validity	Uniqueness
Dataset	Column							
Customer Demographic	DOM	Inaccurate	Contain blanks					
	Default					irrelevant		
	Gender			inconsistency				
	Job title		blanks					
	Last name		blanks					
	Tenure		blanks					
Transactions	profit_first		Blanks				Invalid	
	_sold_date						Format	
	List Price		Blanks				Invalid format	
	product_line product_size Brand,online order		Blanks Blanks					
	standard_cost		Blanks				invalid	
Customer Address	State			Inconsistency				

Description of Data Quality Issues and Recommendations.

CustomerDemographic dataset

Accuracy

The “**DOM**” column is inaccurate.

Mitigation: Replace the value.

Recommendation: Create a date column to check the ages. The date of birth shouldn't be 1843-12-21, which is 179 years old when converted to age. Thus, the date needs to be checked or the value to be replaced to 1943-12-2.

Relevancy.

There is a problem while retrieving a column named “**default**”, hence irrelevant.

Mitigation: Remove the column

Recommendation: The query should be checked in order to load the column appropriately from the table, or it be deleted.

Consistency.

The “**gender**” column contains inconsistent values of gender.

Mitigation: Replace Values

Recommendation: It should be “Female”, “Male” and “U”. Replace the values of M to Male, and Replace ‘F’ and ‘Femal’ to Female.

Completeness

The columns “Job_title, last_name,tenure,DOM cotain blanks thus incomplete.

Mitigation: Replace blanks or filter out.

Recommendation: Replace the blank values with “null” or delete them.

Transactions Dataset

Validity

Profit_first_sold_date, list_price, and standard_cost are not in the correct format.

Mitigation: Change format type

Recommendation:

Change the type of the “profit_first_sold_date” column to date

Change the type of list_price, standard_cost to currency

Completeness

The columns online_order, brand, product_line, product_size, list_price, standard_cost, profit_first_sold_date has blanks hence incomplete.

Mitigation: Replace values or delete them.

Recommendation: change the values to “null”

Customer Address Dataset

Consistency

The “State” column contains inconsistent values.

Mitigation: Replace values.

Recommendation: Replace the values of New South Wales to NSW, Victoria to VIC, and Queensland to QLD.

Conclusion.

Data quality assessment dimensions are used to evaluate the quality of data. Poor data quality can have significant negative impacts on organizations. This can lead to inaccurate decision-making, increased costs, decreased productivity, and legal issues thus it is important to perform data quality assessment and invest in strategies to improve it.

Kind regards,

Andrew Peter

Petremosh85@gmail.com