Data Valley

What is Data Quality ?

LEARN TOP 4 DIMENSIONS OF DATA QUALITY BY EXAMPLE

DIMENSION #1
Completness

DIMENSION #2
Accuracy

DIMENSION #3
Consistency

DIMENSION #4
Validity

Lets see examples of data quality dimension assuming that we have our data in a database table
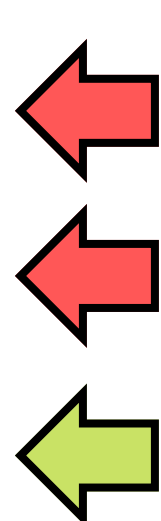
DIMENSION #1
Completness

We say that data has a completeness issue when we are missing data from a table or a column in a table

For example, if we have table to track sales transactions on a retail store, we would expect that customer information like customer ID should be complete and nothing missing, BUT if we have information like customer id, or order number is missing that means our data has a completeness issue

Date	Customer ID	Order Number	Product Name
22-10-2022		1000	Mobile
23-10-2022	21380		TV
26-10-2022	943850	29948	IPhone 14

In Row number 1, and Row number 2 clearly we have a missing data that shouldn't be missing as its very crucial to completely understand the information

In Row number 3, we have the complete information as we needed

NOTE

We are not saying here, that all the data in your table or in your source of data should be completed, for sure, there are some data that is not mandatory, we call the completeness issue when we are missing critical data

DIMENSION #2
Accuracy

We say that data is accurate and doesn't have Accuracy issues when we can rely on it to build insights or extract KPIs, in another words, when the data is verified and tested to check if its reflecting the real-world or not.

Applying the Accuracy check to our example of the sales transaction table, we will say we have an accuracy issue when for example:
- The phone numbers entered for the customer are not correct
- The transaction date is not correct, is not reflecting the real transaction date

Possible Reasons for Accuracy issue:
- Wrong Data Entry for the data
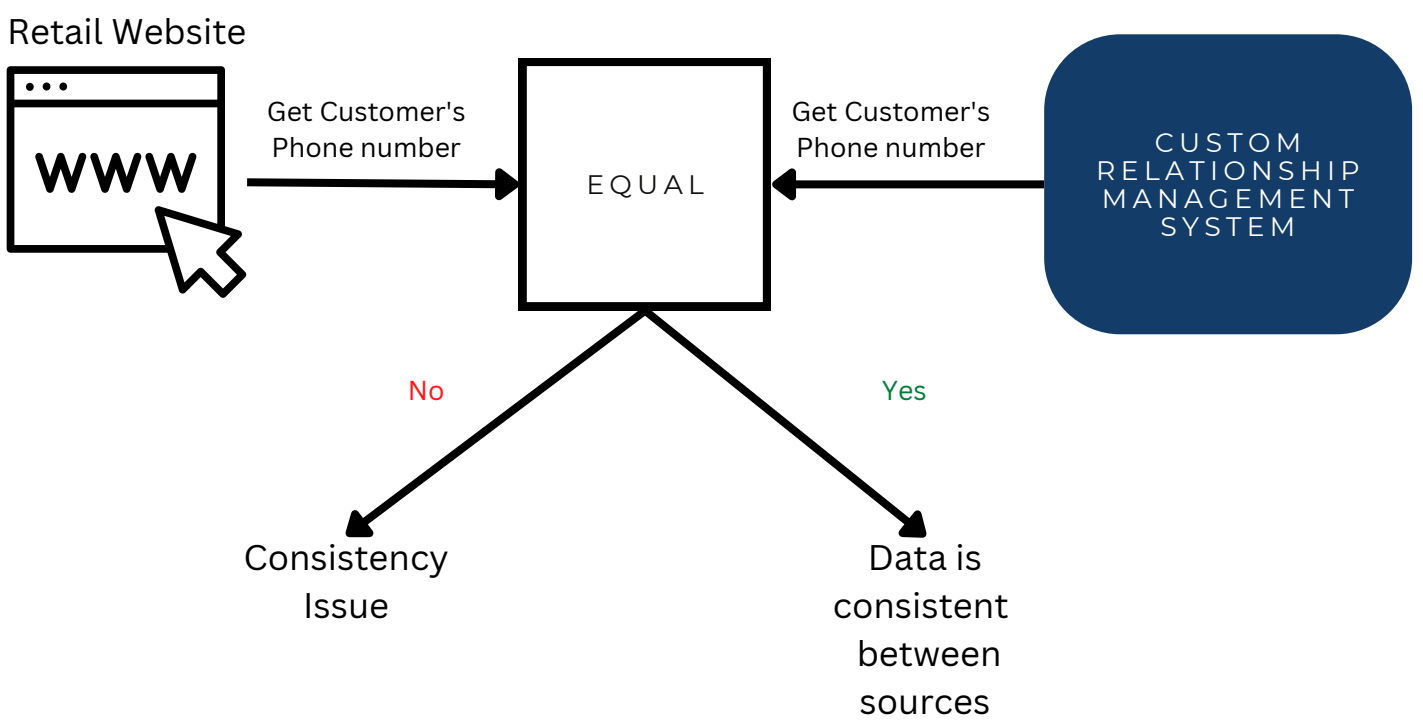- Wrong ETL Operations that lead to corruption of data in the tables

DIMENSION #3
Consistency

Consistent in data checks that the data is consistent and logical between different sources of data.

Testing Consistency is harder than Accuracy and Completeness as it will be a scenario-based not based on direct values, for example, We will have consistent issues in our data if we are getting the customer information from two sources and let's say names are not matching between two sources or phoe numbers is not matching.

Another example of consistency issue, in our case of a retail store, if a customer made a transaction for buying IPhone for example, and we found out that this transaction didn't reduce number of available IPhone in our stock then we have an issue

Retail Website

WWW

Get Customer's Phone number → **EQUAL** ← Get Customer's Phone number

CUSTOM RELATIONSHIP MANAGEMENT SYSTEM

No → Consistency Issue

Yes → Data is consistent between sources
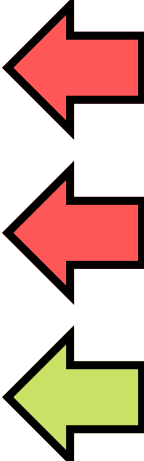
DIMENSION #4
Validity

Validity checks if the data we have complies with the standard values or not.

For example, if customer's phone number in our table doesn't comply with company standards or worldwide standards that means we have a Validity issue.

Lets say standard phone number should start with country code first with the plus.

Customer ID	Phone Number
100	00283948348
200	324093243
300	+20123739047

NOTE

Phone numbers used in this example is not real, this just to demonstrate the idea of Validity issue.

CHECK MORE CONTENT ON OUR WEBSITE