Data Quality

Idea In Short

Data is the foundation of the information hierarchy. Over the data foundation is information, which is data placed in context. Knowledge is derived from actionable information. Knowledge, when properly applied, becomes wisdom. Poor data quality results in bad information quality, and this propagates up the hierarchy, leading to poor business decisions.

Data quality describes the degree to which data fits the purpose it was intended for. Data is considered high quality when it accurately and consistently represents real-world scenarios. When data fits into an intended purpose and represents real-world constructs, it is considered to be of high quality. However, this can be contradictory in nature as well.

Business value drivers

For example, consider the master data for a customer using a product the company sells. The master data may be suffice to invoice the customer, but a lack of accurate details, such as address or telephone number may not help the customer service department. This can hamper customer engagement, issue remediation or lead to other business problems.

Ideally, the master data record should serve multiple purposes. For this, a real-world alignment is necessary where data fits its intended purpose and can be used for other business objectives as well. This should be done without a disproportionate need of resources to collect data. In other words, a balance of the two aspects of data quality definition is required. Working on correcting low quality data is time consuming, takes Herculean effort, and needs an ideal mix of people, better processes, and technologies.

Organizations that are investing in creating quality data are able to leverage data to make better business decisions.

Improved decision-making

Today, the market is consumer-centric. With high quality data, businesses will be able to facilitate better decisions.

Improved Collaborations

When the many departments of an organization have constant access to the same data of high quality, the result is far better, more effective communication. This makes it easier for all team members to remain aligned in terms of priorities, the messaging that goes out, as well as the branding. This comes together to ensure better results.

Improved customer engagement

With good quality data, companies are able to better assess customer interests and requirements. This helps an organization grow by creating better products that are driven from customer needs. Campaigns created can then be driven based on consumer desires and direct feedback from data, not just educated guesses.

IT value drivers

When data in an organization is of high data, the IT department can not only consolidate the underlying data infrastructure, but also scale up on demand, while ensuring that the data remains trustworthy and reliable. IT can automate infrastructure operations, so they can scale the underlying infrastructure up or down without necessarily incurring costs or additional manpower. Another huge benefit of high data quality is regulatory compliance. By ensuring that the organizational data is compliant with the set quality guidelines, IT can demonstrate compliance with the required data regulations (e.g., GDPR) and security requirements.

Data Quality Dimensions

A core element of Data Management, Data Quality takes a holistic view of all the data assets in an organization, combining these elements – often called the Dimensions of Data Quality – to provide a snapshot of the quality of data the organization holds.

Completeness

Are there gaps in the data and if so, where? Some gaps are worse than others and what is considered a gap depends on the process where the data is used. For example, if the billing department requires both phone number and email address, then no record missing one or the other can be considered complete. You can also measure completeness for any particular column. Profiling your data will uncover these gaps.

Consistency

Ensure that all iterations of any piece of data is the same across any report, analyses results, or spreadsheets that are being made and used. Look for inconsistencies as these can lead to bad quality data going forward. Good software should help remove or identify inconsistencies.

Validity

Are the postcode records you hold in a valid format? How confident are you that the email and postal address records you hold in your database are capable of receiving? Validity checks verify that the conforms to a particular format, data type, and range of values.

Since data-driven automation is so important nowadays, data has to be valid to be accepted by processes and systems that expect it.

Timeliness

Is new information entering your CRM every day in real-time or are you manually importing it? How often is the data refreshed? Timeliness is a crucial dimension because of the increasing need for up-to-date data.

Similar to other dimensions, timeliness is user-defined. One kind of data needs to be available on a quarterly basis for financial reporting. Other data must not be older than 5 minutes for real-time analytics.

Uniqueness

Do you have the same customer recorded twice in your data set or data catalog? Uniqueness measures how much duplicate data there is in a given data set, either within any particular column or as whole records. For example, in the orders table, each order should

have just one row. If, on the other hand, you encounter two records with the same order id, you have a duplicate. How did it get there? Someone could have mistyped the order number. This brings us to the next dimension: accuracy.

Accuracy

Perhaps the most important dimension, accuracy refers to the number of errors in the data. In other words, it measures to what extent recorded data represents the truth. Accuracy is tricky because data might be valid, timely, unique, complete, but inaccurate.

100% accuracy is an aspirational goal for many data managers, and once achieved, the principles of data governance can be combined with DQ to ensure the data does not degrade and become inaccurate ever again. Consistency

Do you have conflicting information about the same customer in two different systems? That means the data is inconsistent, which might lead to inconsistent reporting and poor customer service.

Format

Ensuring that data entry formats are consistent has to be the cornerstone of inputting data. Create a singular format and stick to it, for even the tiniest of details such as manufacturing year. American or English date format? All capitalization?

Integrity

This is a criteria that ascertains whether a dataset is in compliance with the rules and standards that are set by the organization. Missing values can disrupt the efficacy of data.

Data Quality Assessment

Given the fact that organizations stand to lose considerably if business processes are based on bad quality data, it becomes imperative that owners and managers understand how data quality can be assessed. This task includes setting up metrics and processes that assess data quality. Companies will need to work on making their data rank highly for both objective as well as subjective assessments. For businesses to improve data quality, they must:

- Assess deeply both objective and subjective data quality metrics
- Analyze results and ascertain the causes for any discrepancies
- Work on ways to improve

Subjective Data Quality Assessments

With subjective assessments, organizations are measuring how stakeholders, analysts, collectors, and other parties perceive the quality of data. If any one of the stakeholders makes a decision based on the data they receive, but finds that it is inaccurate or incomplete, then their decision will be affected. This has to be taken into consideration when looking to find loopholes in the quality of data.

Objective Data Quality Assessments

Objective data quality assessments look at measurable indications, which are recorded within a dataset and then evaluated from two perspectives:

- Its performance within a specific task
- From a viewpoint that it is a metrics-based dataset that can be used independently

To set these metrics for assessment of objective data, organizations can work on principles to develop key performance indicators [KPIs] that match their specific needs. These are known as functional forms. There are three ways in which functional forms are measured for quality:

- **Simple ratio:** Here the total number of desired outcomes is measured with the total possible outcomes. The range generally lies between 0 and 1, with 1 being the most preferred outcome. Both completeness as well as consistency can be measured with this ratio. The catch here is that both the two dimensions can be measured in several different ways and organizations will need to have set criteria in place for the best measures to come out.
- **Minimum or maximum:** Created to handle multiple data quality variables, this functional form has minimum as a conservative number and maximum at a more liberal number. The variables, like the accurate level of data, are marked by minimum. Aspects such as timeliness or/and accessibility are represented by maximum.
- **Weighted average:** Used as an alternative to minimum, these can be used when an

organization is trying to investigate and understand the value that each variable brings to the equation.

Once an organization has evaluated all objective and subjective data quality metrics, they can move on to taking measures that will help streamline their processes. Taking time to look at processes and make objective decisions is a waste of time unless the actions taken are effective and consistently carried out.

Improving Data Quality

For any organization, improving data quality is about the right mix of qualified people, intelligent processes and accurate technologies. When working on improving the quality of data, the main task is to work on enhancing the range of data quality dimensions. Ensuring that these dimensions are clearly adhered to will give organizations data sets that are accurate, high quality, and indispensable for quality decision-making. All this, combined with a proactive top-level management can help improve data quality substantially.

Summary

With high data quality comes trust and understanding in the data. While this might be something that is already present within the team that has produced and validated the data, this is not always the case for everyone within an organization. By implementing a data quality infrastructure, organizations are able to spread confidence throughout the company, allowing everyone to feel secure in the insights that are being derived from the data.