

How Good Data Quality Achieves Better Results and Boosts Productivity



When it comes to data, the colloquial expression "garbage in, garbage out" represents the recognition that poor-quality data entry leads to unreliable and erroneous data output. The information collected needs to be extremely accurate to produce high-quality results. Data collection is increasing in importance for many businesses and organizations, and it is important to have a plan in place to monitor data quality. According to Gartner, poor data quality costs organizations an average of \$12.9 million per year. Understanding how to monitor and improve data quality is the first step in producing good-quality data while also boosting revenue and productivity.

Four Features of Data Quality

Data quality is a multidimensional construct, consisting of four critically important, overarching features that are central to data-related research and the work of data scientists.

1. Data Accuracy. Defined as error-free records that can be used as a reliable source of information and is applicable to real-world scenarios for data-related research. In the realm of data management, data accuracy is the first and most critical component of any data quality frame-

work.

2. Data Validity. Validity refers to the consistency and reliability of data values based on a defined set of rules or within a defined domain. Essentially, validity determines how accurately a given method measures what it is intended to measure and helps to ensure accurate, data-driven results.

3. Completeness. The degree of comprehensiveness or wholeness of the data being utilized for research and analytical purposes. Having all the necessary data without any gaps or missing relevant information is vital for achieving better and more accurate results.

4. Availability. The process of ensuring that an organization's data is available for use not only within the organization itself but also to partners and end users when and where they need it. Accessibility helps partners and organizations make data-informed decisions in a timely manner.

Why Implement a Data Quality Framework

Organizations and stakeholders alike are affected by poor-quality data output because it is ultimately a waste of time, money and resources. According to Data Qual-

ity Statistics of 2022, there are some notable findings for industry leaders to consider, including:

- Only 16% of companies characterize the data they are using as "very good."
- 54% of businesses cite data quality and completeness as their largest marketing data management challenge.
- Approximately 47% of new data collected by businesses has one or more critical errors.
- Data quality issues cost the U.S. economy alone an estimated \$3.1 trillion per year.

To help address these concerns within the big data industry, developing an effective data quality framework is critical. Otherwise known as a data quality life cycle, a data quality framework is a process that monitors the current state of data quality and ensures that the data quality is maintained above defined thresholds. A data quality framework generally consists of four stages:

1. **Assess.** The first stage involves identifying the meaning and metrics of data quality as it pertains to an organization's intended use and assessing the current data sets against the defined metrics. Data quality assessment can also help identify any gaps in the data that may ultimately affect the quality of data output.
2. **Design.** The next stage is to design the business rules that will ensure conformance with the data model and targets defined in the assessment phase. A data model is essentially a blueprint to help facilitate a deeper understanding of the different types of stored data and the relationship between them among partners and end users.
3. **Execute.** The execution phase is when system performance is analyzed. The defined data quality improvements should be implemented in both existing and new incoming data.
4. **Monitor.** The final stage is when data quality is continually monitored to ensure it is maintained above the defined threshold. Data quality management tools are a critical component of ensuring these thresholds are met while also helping to create a culture of continuous improvement.

The Benefits of Good Data Quality

Some advantages of good data quality include increased productivity and innovation. The pathway for achieving better results is being able to identify data quality issues and having a plan in place to address them as well as how to respond to new problems that may arise in the future. The true value of data is determined by how useful it is in terms of analysis and real-world applications.

A clear set of goals and objectives in relation to data quality is a critical component to help big data industry leaders successfully navigate the modern, global landscape. After all, many industry experts assert that data has surpassed oil as being the world's most valuable resource. The basis for this assertion is the value of the knowledge and insight that can be extracted from the collected data. Therefore, for data to be considered valuable, it must be proven to be accurate.

Also, statistics show that poor data quality is a major hindrance for organizations that are seeking to embrace data integration and adopt IT solutions. Data quality is an integral part of digital transformation by helping to achieve better results, boost productivity and enhance innovation. The digital transformation of big data helps with data accessibility among organizations as well as streamlining the process of data collection through the automation of certain tasks.

A Necessary Investment

When it comes to data quality, it is important for data scientists, management teams and stakeholders to be aware of the importance of high-quality data to prevent a "garbage in, garbage out" scenario that is harmful for both business owners and consumers. Investing in the necessary resources and personnel is critical to developing a data culture that leverages technology and drives innovation while striving for consistent data performance improvement. Implementing a set of best practices for improving data quality is one of the most effective ways for organizations to address and treat data quality issues both at present and in the future.