



POV

Data Quality: A strategic imperative for modern enterprises



I. Introduction

High-quality data is a valuable asset that enables businesses to make informed decisions and drive effective strategies. It ensures that information is reliable and trustworthy, reducing the risk of errors and misleading conclusions. Accurate, complete, and consistent data leads to increased agility, faster innovation, reduced costs, as well as better compliance and regulatory risk mitigation.

As data volumes continue to grow exponentially, driving data quality is a complex challenge for businesses across industries and can have severe financial implications.



Data quality issues cost companies **\$ 9.7 MN** annually.

Source: Gartner



Poor-quality data costs the US economy around **\$3.1 TR** per year.

Source: HBR



40% of organizations identified data quality as the primary obstacle to achieving their business objectives.

Source: Gartner



75% of businesses that improved data quality in 2021 exceeded their annual objectives in some way.

Source: Experian



II. Impediments to high-quality data

Enterprises manage large volumes of data that is constantly generated from a variety of internal sources such as ERP systems and external sources including third-party vendors, IoT sensors, social media, market research and publicly available data. This can lead to data being stored in different formats and structures, making it difficult to integrate and analyze. Additionally, data can be prone to errors, such as duplicate records or missing data, which can further complicate the data management process.



Data Challenges

Data silos and lack of integration

Data inaccuracy and inconsistency

Lack of data standardization

Data duplication and redundancy

Limited visibility into org-wide data quality

Impact

Missed opportunities and fragmented view of information

Suboptimal outcomes due to lack of trust in data

Compliance and regulatory risks leading to legal and financial consequences

Higher costs due to errors and rework, straining business resources

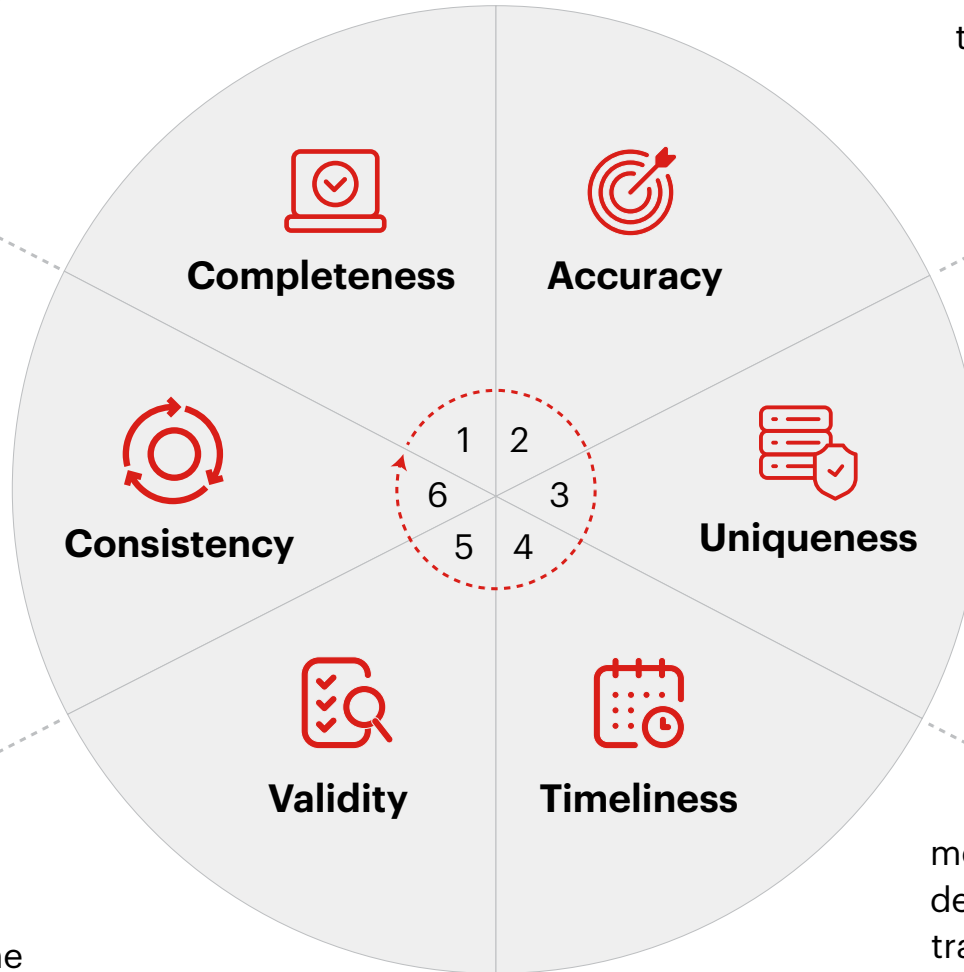
Ineffective data utilization and collaboration across departments

III. 6 Dimensions to measure data quality

All required records and values should be available with no missing information
E.g.: Customer ID should not have any null or blank values

Items of data taken from multiple sources should not (in an ideal world) conflict with one another. E.g.: For marketing campaigns, metrics like clicks, impressions and campaign budget should be consistent across different measurement platforms

This dimension measures how data conforms to predefined business rules. E.g.: Customer IDs should have a value in all the required corresponding tables



This dimension measures data that attempts to model real-world objects or events. E.g.: Data type for date should be date and format should be DDMMYYYY

This ensures the avoidance of the same data being mapped differently in multiple locations. E.g.: Customer ID should be mapped to only 1 account name

The data's actual arrival time is measured against the predicted, or desired, arrival time. E.g. Customer transactions to be updated within 1 minute of their occurrence

Case study 1

Problem statement

- Incorrect data affecting various business areas and causing delays in addressing data errors
- Need for a centralized solution to monitor and manage data quality organization-wide
- Proactive measures required to ensure accurate and correct data, reducing instances of data quality problems

Solution

- Plug-and-play solution that can be extended to multiple source systems, capable of scanning large volumes of data with complex quality checks
- Users can select and view data health reports
- Implemented email alerts and diagnostic reports for prompt identification and resolution of data quality issues
- Easy integration with existing flows and pipelines

Impact

Scanned **1.5 MN** records in less than 150 secs per quality check

Diagnostic reports generated per table for consumption

Total of **9.15 MN** records were scanned and processed for UAT

Case study 2

Problem statement

- Siloed data across departments and regions hampers digital marketing team management
- Inefficient data quality management hinders informed decision-making
- Lack of a robust data cataloging system impacts decision-making efficiency

Solution

- A data cataloging solution using Collibra for a unified marketing data view and streamlined management
- Analyzed client's marketing data landscape, identified sources, and collaborated with them on metadata and taxonomy definition
- Set up Collibra, integrated with Snowflake on AWS, and established automated workflows and alerts for efficient data quality management

Impact

30% improvement in data quality

Increase in productivity and reduction in manual effort

Informed decision-making with access to granular-level data

IV. Delivering clean, correct and consistent data across the enterprise

Sigmoid has developed a full-fledged data quality solution for data-driven entities. The accelerator derives diagnostic reports based on the 6 data quality dimensions to gain insights into the overall health of the data and flag any anomalies within the data itself. This is a plug-and-play platform that enables comprehensive data validation and profiling. With the ability to connect all available data sources, including flat files and databases, the solution integrates seamlessly with existing workflows, requiring minimal changes to infrastructure.

Features

A one-stop data quality solution for data validation and data profiling

The accelerator does not store or host any data other than schema and its performance parameters

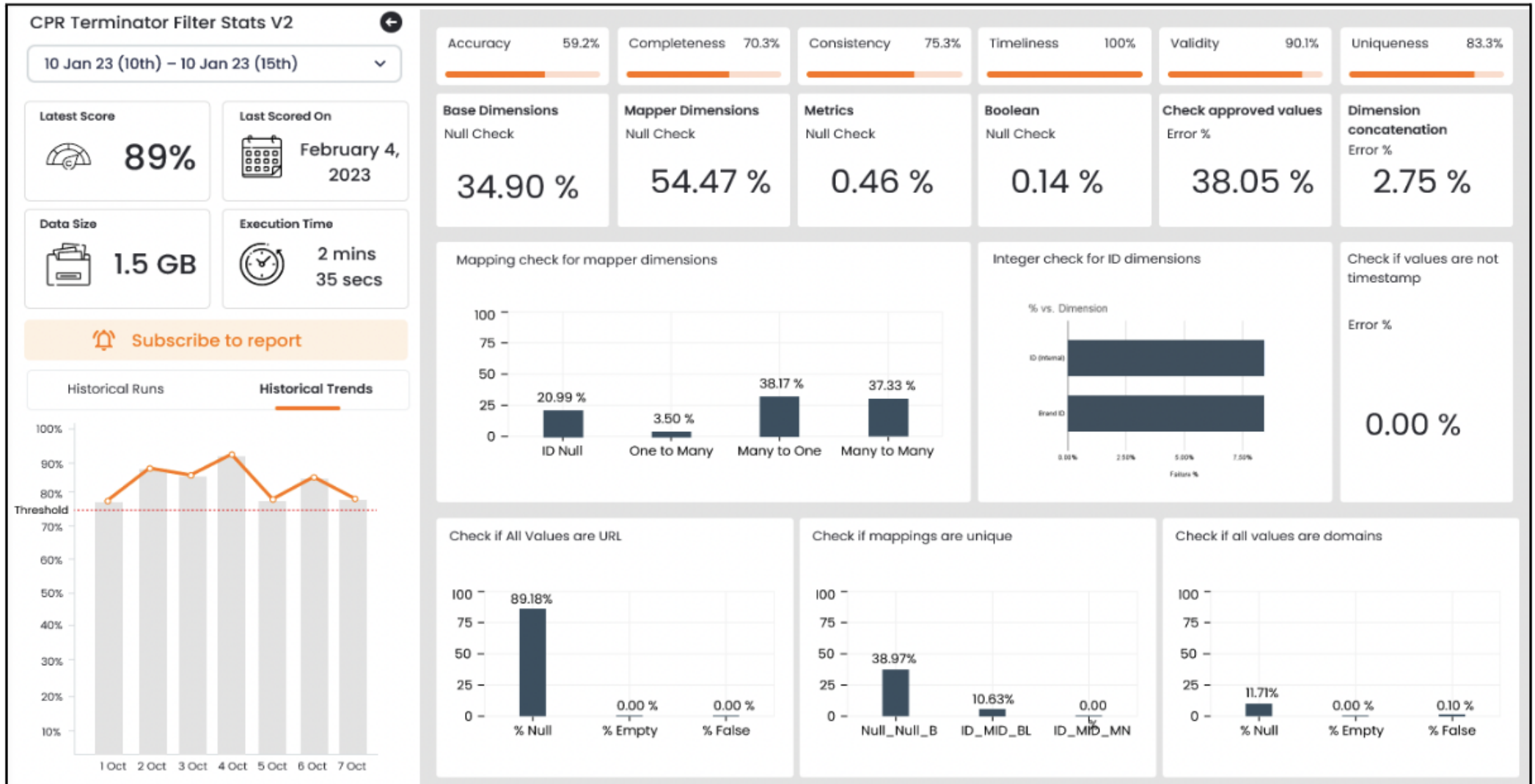
Data quality reports and metrics visualization, with alerts and notifications

Onboarding of complex custom rules tailored to business needs

Intuitive UI to allow user to map quality checks and visualize validations



Sample data quality diagnostics report



V. Impact of high-quality data for the modern business

Good data quality can also improve customer experience by providing accurate and complete customer data, leading to increased customer satisfaction, loyalty, and retention. Having access to up-to-date and relevant data can enhance business agility, enabling businesses to adapt to changing market conditions and stay ahead of their competitors.

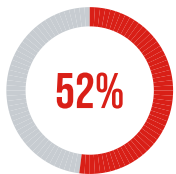
Improved

- Decision-making and revenues
- Customer experience
- Operational efficiency
- Forecasting and planning
- Risk management
- Business agility and innovation



Organizations that prioritize data quality see a 25% increase in revenue

Source: Experian



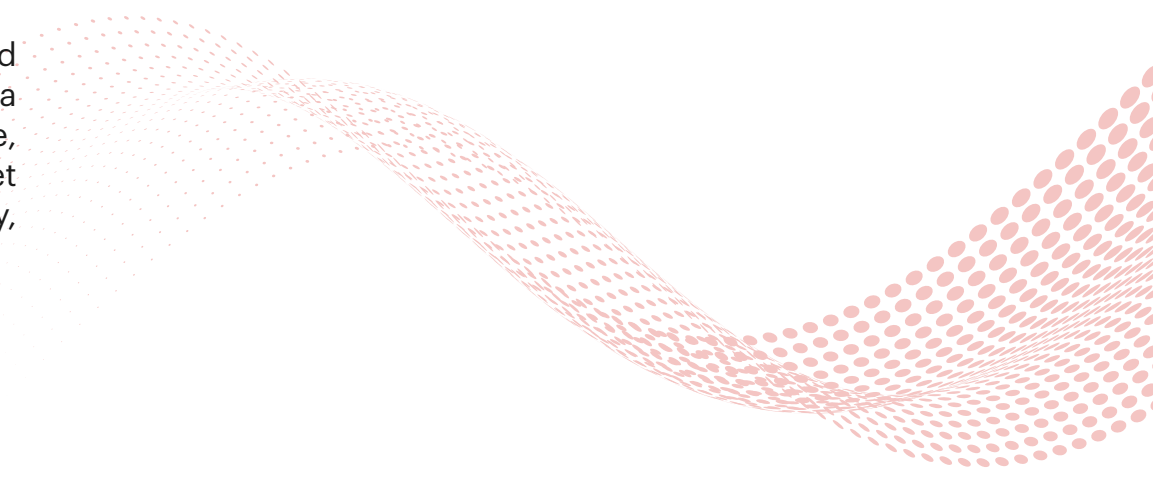
52% of organizations reported improved operational efficiency after implementing data quality initiatives

Source: TDWI



VI. Conclusion

High-quality data is vital for informed decision-making, innovation, and competitive advantage in today's data-driven world. Prioritizing data quality helps businesses improve decision-making, customer experience, and operational efficiency. Sigmoid has been enabling enterprises to meet the necessary standards for data accuracy, completeness, and reliability, supporting informed decision-making and achieving business objectives.



About Sigmoid

Sigmoid enables business transformation using data and analytics, leveraging real-time decisions through insights, by building modern data architectures using cloud and open source. Some of the world's largest data producers engage with Sigmoid to solve complex business problems. Sigmoid brings deep expertise in data engineering, MLOps, and deep learning.

Visit www.sigmoid.com to know more.

Let's Get Started!

Reach out to us at marketing@sigmoid.com to enable high quality data for your business.

Stay Connected

