

What challenges are there in data quality for analytics ?



Explore the top data quality challenges impacting analytics, from inaccurate and incomplete data to integration issues and human error. Learn how leading organizations overcome these obstacles to unlock reliable insights and drive smarter decisions.

2025

Table of Contents

Contents

Table of Contents	2
summary	4
Types of Data Quality Challenges	5
Common Data Quality Issues	5
Inaccurate Data.....	5
Incomplete Data.....	5
Duplicate Data	5
Outdated Data	5
Irrelevant Data	6
Challenges in Data Management.....	6
Inconsistent Data Definitions.....	6
Aging Data	6
Technical Challenges	7
Data Integration Issues.....	7
Legacy Systems	7
Inconsistent Data Standards.....	7
Data Validation and Cleansing.....	7
Scalability and Performance	8
Data Governance	8
Human Factors	9
Human Error.....	9
Resistance to Data-Driven Culture	9
Lack of Skilled Personnel	9
Methodological Challenges	9
Data Collection and Assessment	10
Challenges in Data Quality	10
Data Cleaning and Validation	10
Continuous Monitoring and Improvement	10
Case Studies.....	12
Walmart: Optimizing Inventory Management	12

Public Health England: Unreported COVID-19 Cases	12
Target: Predicting Customer Behavior	12
Verizon: Reducing Network Outages	12
Bank of America: Combatting Fraud	12
Amazon: Personalizing the Shopping Experience	13
References	14

summary

Data quality challenges are critical issues that organizations face when leveraging analytics for informed decision-making. These challenges can severely undermine the effectiveness of data-driven insights, leading to misguided strategies and potential financial losses. As businesses increasingly rely on data for operational and strategic initiatives, understanding and addressing these quality challenges has become paramount. Gartner estimates that poor data quality can cost organizations an average of \$12.9 million annually due to inaccuracies alone, highlighting the significant impact on both operational efficiency and profitability.[\[1\]](#)

Common data quality issues include inaccuracies, incomplete records, duplicates, outdated information, and irrelevant data. Inaccurate data can stem from human error or faulty data entry processes, while incomplete records may lead to flawed analyses. Duplicates inflate metrics, and outdated information can result in compliance issues. Additionally, irrelevant data complicates analysis efforts, making it essential for organizations to implement rigorous data management practices to ensure data integrity and relevance.[\[2\]](#)[\[3\]](#)[\[4\]](#)

Organizations also face broader challenges, including human factors such as resistance to data-driven cultures and a lack of skilled personnel. Human error in data entry processes remains a prevalent issue, particularly in sectors like healthcare where manual recording is common.[\[5\]](#) Additionally, a shortage of skilled analysts trained in contemporary data tools can hinder organizations' abilities to maintain high data quality standards, further complicating analytics efforts.[\[6\]](#)[\[7\]](#)

Technical challenges, such as data integration across multiple sources and the reliance on legacy systems, further exacerbate data quality issues. Inconsistent data definitions and standards can lead to discrepancies and confusion, while inadequate data validation and cleansing processes may allow errors to persist within datasets. As organizations scale and accumulate more data, the need for effective governance frameworks and continuous monitoring becomes vital to ensure data quality is maintained throughout its lifecycle.[\[8\]](#)[\[9\]](#)[\[10\]](#) By addressing these challenges proactively, organizations can enhance their analytical capabilities and drive better business outcomes.

Types of Data Quality Challenges

Data quality challenges can significantly impact the effectiveness of analytics and decision-making processes in organizations. Understanding these challenges is crucial for establishing robust data quality management practices.

Common Data Quality Issues

Several prevalent data quality issues can arise, including:

Inaccurate Data

Inaccurate data refers to information that contains errors, such as incorrect customer addresses, misspelled names, or erroneous numerical entries. Gartner estimates that inaccurate data costs organizations an average of \$12.9 million annually[1]. This type of error can lead to flawed analyses and misguided decisions. To mitigate inaccuracies, organizations should consider automating data entry and implementing robust data quality monitoring solutions to identify and rectify errors[2].

Incomplete Data

Incomplete data involves records that lack essential information, such as missing ZIP codes or demographic details. This can complicate daily operations and lead to flawed analyses. Organizations can address this issue by enforcing mandatory fields in data entry systems and using tools that flag incomplete records during imports[2][3].

Duplicate Data

Duplicate data occurs when the same record appears multiple times within a dataset, leading to inflated metrics and confusion. This issue can arise from various sources, such as merging datasets without proper checks. Organizations can implement data monitoring tools to identify duplicates and maintain a clean dataset[1].

Outdated Data

Outdated data refers to information that is no longer current or relevant. As data ages, its accuracy diminishes, which can lead to lost revenue or compliance issues. Organizations should adopt regular data maintenance routines and update procedures to ensure that their datasets remain relevant[3].

Irrelevant Data

Not all captured data is useful; irrelevant data does not contribute to the organization's objectives and can clutter analysis efforts. Regular culling of unnecessary data can help maintain a focus on actionable insights[1].

Challenges in Data Management

In addition to specific data quality issues, organizations face broader challenges in data management: **Human Error**

Human errors in data entry or processing can introduce significant inaccuracies. Implementing automated systems can reduce reliance on manual input, thereby minimizing the potential for such errors[4].

Inconsistent Data Definitions

Inconsistent definitions of data types across systems can complicate data processing and analysis. Without standardized definitions, different teams may use varying methodologies, leading to discrepancies and unreliable insights[5]. Establishing clear data standards and guidelines is essential to enhance data coherence and reliability[3].

Aging Data

Data decay refers to the degradation of data quality over time, rendering it unusable. Organizations should regularly assess and refresh their datasets to counteract this phenomenon, as data integrity is critical for effective analysis[2].

Technical Challenges

The field of data analytics faces numerous technical challenges that significantly impact data quality. These challenges can hinder the ability to obtain accurate insights and drive informed decision-making within organizations.

Data Integration Issues

One of the primary technical challenges is related to data integration. Organizations often utilize multiple data sources, which can lead to silos and inconsistencies in data representation. Different departments may apply varying definitions and standards for data, resulting in discrepancies that affect the accuracy of analytics[6]. Effective data integration solutions are essential to consolidate data from disparate sources and provide a unified view, but implementing these solutions can be complex and resource-intensive.

Legacy Systems

The existence of legacy software systems poses another significant challenge. Many organizations rely on outdated technology that may not be compatible with modern data processing tools. Communication issues and a lack of knowledge regarding the data maintained within these systems can complicate data extraction and integration efforts[8]. Furthermore, the failure to properly upgrade or replace legacy systems can result in persistent data quality problems.

Inconsistent Data Standards

Inconsistent data standards across the organization can lead to poor data quality. Without established definitions and protocols, data may be collected and stored in various formats, causing challenges in data analysis and reporting[6]. Organizations must enforce standardization practices to ensure that data remains consistent and reliable for analytical purposes.

Data Validation and Cleansing

Another technical challenge is the need for robust data validation and cleansing processes. As organizations accumulate large volumes of data, maintaining its accuracy becomes increasingly difficult. Data validation tools are necessary to identify and rectify errors, inconsistencies, and duplicates within datasets[9]. However, the implementation of these tools can be resource-intensive and may require specialized skills.

Scalability and Performance

As the volume of data increases, ensuring that data storage and management systems can scale effectively is crucial. Organizations must select solutions that can handle growing data demands without sacrificing performance[11]. The challenge lies in balancing the need for scalability with the ability to maintain high data quality standards.

Data Governance

Implementing effective data governance is vital to managing data quality throughout its lifecycle. Organizations need to establish clear policies, processes, and standards that govern data access, privacy, and security[10]. However, the complexity of these governance frameworks can pose challenges in compliance and enforcement, particularly in large organizations with diverse data needs.

Human Factors

Human factors play a significant role in data quality issues, impacting the effectiveness of analytics initiatives across various sectors. Key challenges include human error, resistance to data-driven cultures, and a lack of skilled personnel.

Human Error

One of the primary contributors to data quality issues is human error, particularly in data entry processes. In environments such as healthcare, where data is often manually recorded and transferred, inaccuracies can arise from simple mistakes during the input phase[5][12]. This reliance on human input leads to inconsistencies and inaccuracies, with critical fields potentially being overlooked, despite rigorous checks on more crucial data points, such as financial transactions[12].

Resistance to Data-Driven Culture

Analytics initiatives frequently encounter skepticism from employees and leadership regarding the role of data in decision-making[7]. This resistance can stem from a lack of understanding of data analytics, fears about job security, or a cultural preference for traditional methods over data-driven approaches. Consequently, organizations may struggle to implement effective data governance frameworks that promote accountability and a shared commitment to data integrity[13][14]. To overcome this resistance, it is essential to promote a learning culture, foster leadership support, and demonstrate the value of data-driven insights through pilot projects and small wins[7].

Lack of Skilled Personnel

A significant barrier to achieving high data quality is the shortage of skilled personnel trained in analytics tools and methodologies. Many organizations face challenges in finding staff proficient in critical technologies such as SQL, Python, and machine learning algorithms[7]. To address this skills gap, organizations can invest in upskilling their teams through professional development programs, leverage advanced AI tools that simplify data analysis, and strategically hire external experts to fill temporary knowledge gaps[7]. By enhancing the skill sets of their workforce, organizations can improve their data management processes and, ultimately, their data quality.

Methodological Challenges

Data quality assessment presents several methodological challenges that can significantly impact the accuracy and reliability of analytics outcomes. These

challenges arise from the complexities of data collection, processing, and validation, which necessitate a well-structured approach to ensure high-quality data is maintained throughout the analytics lifecycle.

Data Collection and Assessment

The initial challenge lies in the identification and selection of appropriate assessment indicators for various data quality dimensions. Each dimension requires tailored measurement tools and techniques that align with the specific conditions of the business environment. This results in differences in assessment times, costs, and the allocation of human resources, complicating the overall data quality evaluation process[15]. Moreover, determining clear goals for data collection is paramount. Organizations must strategically choose data based on operational, decision-making, and planning needs, further complicating the assessment process if these objectives are not clearly defined[15][16].

Challenges in Data Quality

Common issues in data quality include inconsistent formats, missing values, duplicate records, and concerns regarding data accuracy[7][17]. These inconsistencies can lead to significant analytical errors if not addressed proactively. Furthermore, the overwhelming volume of data complicates effective integration and management, posing additional challenges for organizations striving to maintain high-quality data standards[18][7].

Data Cleaning and Validation

The process of data cleaning, essential for improving data quality, introduces another layer of complexity. Different methods of data cleaning—manual implementation, specialized programming, and application-specific solutions—must be carefully considered to suit the data at hand[15]. Each method carries its own set of challenges in terms of effectiveness and applicability. Furthermore, organizations often face resistance to adopting a data-driven culture, which can hinder the commitment to implementing rigorous data quality standards and methodologies[18][7].

Continuous Monitoring and Improvement

Finally, the lack of a continuous monitoring framework can lead to the degradation of data quality over time. Effective data governance, which includes regular assessments and updates to data quality strategies, is vital to addressing emerging data quality challenges and ensuring that data remains reliable for decision-making[16][11]. Organizations must also focus on training staff to recognize and

address data quality issues actively, fostering a culture that prioritizes data integrity as a key component of successful analytics[16][19].

Addressing these methodological challenges requires a comprehensive approach that integrates data assessment, cleaning, governance, and ongoing monitoring to ensure that high-quality data is consistently used in analytics efforts.

Case Studies

Real-world case studies illustrate the challenges and successes organizations face in improving data quality for analytics. These examples highlight the importance of accurate data in driving decision-making and operational efficiency.

Walmart: Optimizing Inventory Management

Walmart, one of the largest retailers globally, encountered significant challenges with inventory management due to inaccurate product data and inconsistent supplier information. To address these issues, Walmart invested in data quality initiatives that included data cleansing, validation, and enrichment processes. They also implemented a master data management (MDM) system to centralize and standardize product data, ultimately enhancing operational efficiency and customer satisfaction[20].

Public Health England: Unreported COVID-19 Cases

During the COVID-19 pandemic, Public Health England faced a major data quality issue when 15,841 positive cases went unreported between September 25 and October 2, 2020. This oversight occurred at a critical time when accurate data was essential for contact tracing efforts. The incident underscored the need for robust data governance and real-time monitoring systems to prevent such occurrences in the future[21].

Target: Predicting Customer Behavior

Target utilizes data analytics to predict customer behavior by collecting data on purchase history, browsing patterns, and social media activity. This data is crucial for sending targeted marketing messages to customers at the right time, significantly improving the effectiveness of their marketing strategies. By enhancing data quality, Target has successfully increased customer engagement and sales[22].

Verizon: Reducing Network Outages

Verizon has leveraged data analytics to improve its network performance by analyzing data on usage patterns, traffic flow, and outages. As a result of their initiatives to enhance data quality, Verizon reduced network outages by 50%, demonstrating how data-driven decision-making can lead to significant operational improvements[22].

Bank of America: Combatting Fraud

In the financial sector, Bank of America collects detailed transaction data to identify fraudulent activities. Through improved data quality measures, the bank has been

able to reduce fraud losses by 50%. This case highlights the critical nature of data accuracy and completeness in maintaining customer trust and financial security[22].

Amazon: Personalizing the Shopping Experience

Amazon exemplifies the effective use of data analytics for personalization by analyzing customer purchase history, browsing behavior, and search queries. The company's focus on data quality enables it to recommend products tailored to individual customers, thus enhancing the overall shopping experience and driving sales growth[23].

These case studies demonstrate that while organizations face various challenges related to data quality, strategic investments and initiatives can lead to substantial improvements in operational efficiency, customer satisfaction, and profitability.

References

- [1]: [10 Common Data Quality Issues \(And How to Solve Them\) - FirstEigen](#)
- [2]: [What the Factors that can affect Data Quality?](#)
- [3]: [Data Quality Problems? 8 Ways to Fix Them in 2025 - Atlan](#)
- [4]: [Common challenges in data quality management | Secoda](#)
- [5]: [How to Overcome Common Data Quality Issues - OWOX BI](#)
- [6]: [Why Data Quality and Governance Are Imperative for People](#)
- ...
- [7]: [The challenges and opportunities of continuous data quality ...](#)
- [8]: [5 Ways to Improve Data Quality - Plauti](#)
- [9]: [How to Improve Data Quality in 12 Actionable Steps? - Atlan](#)
- [10]: [How to Manage Data Quality in Healthcare - Gable.ai](#)
- [11]: [11 Lingerin Data Quality Issues - Information Week](#)
- [12]: [Common Challenges in Data Analytics & How to Solve Them - Emory](#)
- [13]: [9 Common Data Quality Issues to Fix in 2025 - Atlan](#)
- [14]: [Data Stewardship: Definition, Benefits & Key Roles Explained \(2025\)](#)
- [15]: [The Challenges of Data Quality and Data Quality Assessment in the ...](#)
- [16]: [How to Improve Data Quality: 12 Effective Strategies - FirstEigen](#) [17]: [What are the most challenging data quality issues you face frequently?](#)
- [18]: [Case Study: Navigating through Data Quality Challenges in Market ...](#)
- [19]: [Data Quality: A Key Success Factor in Analytics Projects](#)
- [20]: [Case Studies in Data Quality: Success Stories from Leading ...](#)
- [21]: [5 Examples Of Bad Data Quality: Samsung, Unity, Equifax, & More](#)
- [22]: [Data Analytics Case Studies & examples for various Industries - SciklQ](#)
- [23]: [Why do data quality? A case study - MIOsoft Corporation](#)