

Data Quality & the Path to Stronger Decision-Making

Why data quality is so important

In this age of big data, it's never been more important for businesses and institutions—of any industry or size—to ensure that the insights powering decision-making are based on an accurate understanding of their data.

Often, data resides in multiple locations, in various qualitative states, and depending how well it is assessed, cleaned, and applied (or scored) can have a range of impacts on a business's bottom line or a customer's level of satisfaction.

Most would agree that better data leads to better business decisions. But the pathway to better data is not so clearly defined, and it's harder than one might expect to measure data quality and ensure good data standards.

84%

of CEOs are concerned about data quality when making data-driven decisions

Source: KPMG's 2016 Global CEO Outlook



A robust analytics platform for data preparation, analysis, and access

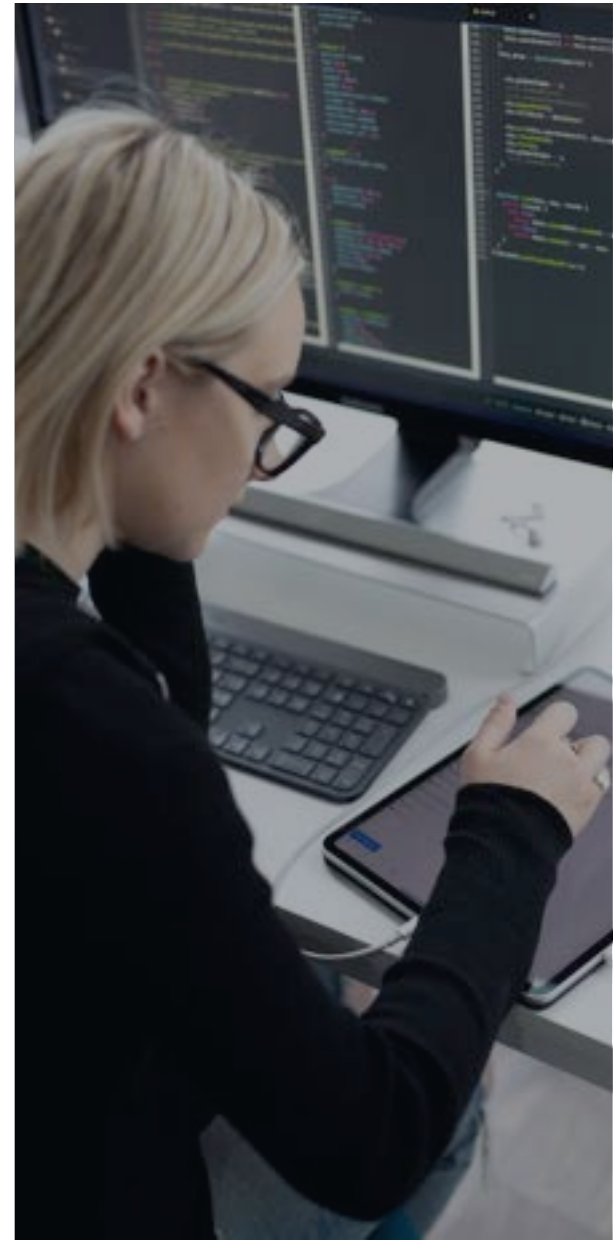
Download a free trial of the Rapid Insight platform

How do we define and measure data quality?

It's easy to think of data as static measurements of some action taken or fact about your customer—a descriptive element, from the past, and irrefutable. But indeed, data lives, and data changes. Data is written and rewritten, and often enough it is written incorrectly—the result of data-entry error, i.e. human fallibility.

Data is, in some sense, human, and it's made complicated by the many ways we treat it. Thinking of our relationship to data in those terms, it helps to approach the definition and measurement of data quality—how good or bad it is—by asking the right questions of it.

Because your data will give you answers.



A robust analytics platform for data preparation, analysis, and access
Download a free trial of the Rapid Insight platform

Questions we must ask of our data

If only it were as simple as asking, is this data good or bad? Data quality is defined and measured by several conditions, each of which acts as a pillar for responsible assessment and improvement of that data.

Is it complete?

Assess what percentage of each data record is complete. Have fields been left blank or otherwise deleted?

Is it accurate?

Has information been entered erroneously, or is there any reason to doubt the reliability and trustworthiness of the data source?

Is it unique?

Ensure that a record hasn't been mistakenly duplicated, e.g. that persons named Dave Smith and David Smith are not, in fact, the same person.

Is it orderly?

Data should adhere to agreed-upon documentation format, naming conventions, and other structure requirements to ensure its validity.

Is it consistent?

Data in multiple locations has a tendency to produce discrepancies. Matching records in different databases should do just that: match.

Is it timely?

Have changes to the data been made within a common, reasonable time frame, such that all records are, at any point of measurement, reliably up-to-date?

7 Essential Data Prep Tasks

Insightful reports are only as good as the quality of data that feeds them. Efficiently sourcing and then preparing the data are critical first steps.

Sourcing and analyzing your data may be complicated enough to consider seeking help from an outside firm. But the process of self-service data analytics is becoming simpler and simpler. Complex coding languages or dense database management procedures are less necessary with new tools available for building reports and reaping insights.

Here are seven essential data preparation tasks that a typical user will encounter—and with the right tool can perform—when preparing to analyze a dataset:

Country	Winner
Argentina	Adolfo Perez Esquivel
Argentina	Carlos Saavedra Lamas
Austria	Alfred Hermann Fried
Austria	International Atomic Energy Agency
Austria-Hungary	Baroness Bertha Sophie Felicitas Von Sutt
Bangladesh	Grameen Bank
Bangladesh	Muhammad Yunus
Belgium	Auguste Marie Francois Beernaert
Belgium	Henri La Fontaine
Belgium	Christian Lous Lange
Belgium	Georges Henri Piro
Belgium	Institute Of International Law
Burma	Aung San Suu Kyi
Canada	Lester Bowles Pearson
Canada	The Pugwash Conferences On Science Ar
China	Liu Xiaobo
Costa Rica	Oscar Arias Sanchez
Cross-national	European Union
Cross-national	Organisation for the Prohibition of Chem
Denmark	Fredrik Bajfer



Country	# of Peace Prize Winners
Argentina	2
Austria	2
Austria-Hungary	1
Bangladesh	2
Belgium	5
Burma	1
Canada	2
China	1
Costa Rica	1
Cross-national	2
Denmark	1
East Timor	2
Egypt	2

1. Aggregate

Aggregating is sorting data and then expressing the data in summary form in terms of another field. For example, a dataset listing Nobel prize winners and the country from which they hail can be aggregated to show the total number of prize winners from each country.

7 Essential Data Prep Tasks

2. Filter

Filtering a dataset narrows it down to a specific group of records. Take this list of elements and their element type, which we've filtered down to just the metal elements.

Name	Type of Element
Actinium	Metal
Aluminum	Metal
Americium	Metal
Antimony	Semimetal
Argon	Non-Metal
Arsenic	Semimetal
Astatine	Semimetal
Barium	Metal
Berkelium	Metal
Beryllium	Metal
Bismuth	Metal
Bohrium	Metal
Boron	Semimetal
Bromine	Non-Metal
Cadmium	Metal
Calcium	Metal
Californium	Metal
Carbon	Non-Metal
Cerium	Metal
Cesium	Metal
Chlorine	Non-Metal



Name	Type of Element
Actinium	Metal
Aluminum	Metal
Americium	Metal
Barium	Metal
Berkelium	Metal
Beryllium	Metal
Bismuth	Metal
Bohrium	Metal
Cadmium	Metal
Calcium	Metal
Californium	Metal
Cerium	Metal
Cesium	Metal

3. Merge

When data is scattered in multiple datasets, merging allows you to combine the relevant parts of those datasets to create a single new file to work with.

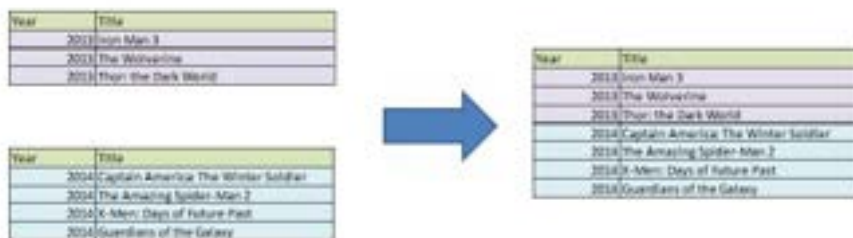
State	Capital
Alabama	Montgomery
Alaska	Juneau
Arizona	Phoenix
Arkansas	Little Rock
California	Sacramento
Colorado	Denver

State	Population
Alabama	4,779,736
Alaska	710,231
Arizona	6,392,017
Arkansas	2,915,918
California	37,253,956
Colorado	5,029,196



State	Capital	Population
Alabama	Montgomery	4,779,736
Alaska	Juneau	710,231
Arizona	Phoenix	6,392,017
Arkansas	Little Rock	2,915,918
California	Sacramento	37,253,956
Colorado	Denver	5,029,196

7 Essential Data Prep Tasks



4. Append

To append two datasets is to stack them to create on larger dataset. Usually when appending, the datasets contain the same (or very similar) fields.



5. Dedupe

To dedupe is to remove duplicates from a dataset. A list of emails from a newsletter might include the same name twice, and deduping will make sure there's only one entry for each person on the list.

7 Essential Data Prep Tasks

6. Transform

To transform a column is to perform an operation on a column that results in a different version of the inputted column. For example, a dataset with separate columns for first and last names can be combined to create a brand-new column, "Full Name."

First Name	Last Name
George	Washington
Thomas	Jefferson
James	Madison
James	Monroe
John Quincy	Adams
Andrew	Jackson
Martin	Van Buren
William Henry	Harrison
John	Tyler



Full Name
George Washington
Thomas Jefferson
James Madison
James Monroe
John Quincy Adams
Andrew Jackson
Martin Van Buren
William Henry Harrison
John Tyler

7. Data Cleanse

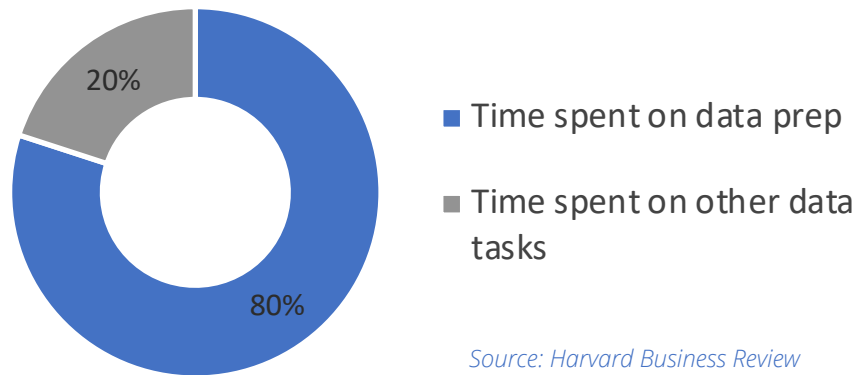
To cleanse a column is to clean up the values within that column, most commonly by replacing them. Say a column listing genders has entries that aren't uniform, e.g. "female," "fem," and "F" all representing "female." Cleansing makes the entries consistent within the column.

Gender
Male
M
Female
Fem
M
F
Male



Gender
Male
Male
Female
Female
Male
Female
Male

What if data prep was efficient, easy, and enjoyable?



Source: Harvard Business Review

With the mountains of data any business generates on a daily basis, often across disparate, disconnected systems and locations, the time that analysts must spend accessing, organizing, and cleaning that data—using legacy tools—is potentially costing a business as much in lost revenue as it is in resources.

But there's an easier way.

Welcome to the Rapid Insight platform, combining data prep, automated predictive modeling, and end-user exploration to deliver data-informed insights across an enterprise.

A data prep software solution for everyone



It all begins with data prep. With Construct, seamlessly connect to any data format, run processes to blend, cleanse, and aggregate that data for analysis and reporting.



Connect

Effortlessly integrate data from virtually any source, including Excel, Access, SQL, SAS, SPSS, MySQL, and more



Prepare

An easy, drag-and-drop workflow enables any level of user to build step-by-step analytic processes



Analyze

Build and schedule jobs to run automatically, dramatically boosting workload efficiency



Share

Share reports and datasets with your team using Rapid Insight's Bridge or other visualization tools like Tableau

Data quality comes to Dallas Baptist University

A CASE STUDY

“It took over 3.5 hours to prepare a report in the old system. It took 30 minutes in Construct. This represents a 568% improvement.”

Challenge:

At Dallas Baptist University (DBU), the situation was complex: scattered data requiring intensive labor to clean and integrate, multiple data systems with their own formatting limits, and a heavy reliance on manual work and outdated macros in Excel. Preparing and presenting that data required a considerable amount of expertise and time.

DBU owned a few software applications to approach the daunting data tasks, but they imposed hurdles of their own. As DBU's Director of Institutional Research said of these other data tools, “I became familiar with them, but they were seldom used. Each time I had to become reacquainted with the programs.”

Solution:

DBU discovered the solution to this cumbersome process in Rapid Insight's data prep solution, Construct. Construct uses graphic objects to represent different tasks to be applied to the data, which are easily dragged around, reordered, and interconnected. Regular data-cleansing routines, reporting, and other procedures have never been easier.

The success of DBU's new systems has translated into reports that change less frequently, leading to an improved understanding among those who use the reports to evaluate performance and make decisions. Construct has saved countless hours of work, and produced impressive, powerful results.



A robust analytics platform for data preparation, analysis, and access

Download a free trial of the Rapid Insight platform

Better data, better decision-making, better outcomes

Most businesses and institutions today agree that informed decision-making is based on an accurate understanding of the data that underpins their goals. But ensuring that the quality of your data is strong requires more than most of those businesses may realize.

To effectively measure the quality of your data, you need to know what questions to ask of it—in other words, what are the conditions by which “good” data is defined? These include accuracy, orderliness, consistency, and timeliness of the data.

Self-service data analytics solutions, such as those available with the Rapid Insight platform, allow you to easily perform what we’ve identified as the seven essential data prep tasks. Whether aggregating, merging, or deduping data sets, the right data prep tool can not only save you lots of time but can protect you—and your business interests—from inaccurate or misleading conclusions.

The Rapid Insight platform combines data prep, automated predictive modeling, and end-user exploration, enabling data-driven insights across an organization. It even makes data analysis fun.



A robust analytics platform for data preparation, analysis, and access
Download a free trial of the Rapid Insight platform



Get started with Rapid Insight today

Discover how we're empowering enterprises and institutions around the globe to prepare and leverage their data to better achieve their strategic goals.



Download your FREE trial:
rapidinsight.com/free-trial

Rapid Insight was founded with a mission of empowering people of all skill levels with the ability to build predictive models and perform advanced data analyses. Our groundbreaking technologies are simplifying everything that used to be complex in the world of business intelligence and predictive analytics. We love our customers and we love enabling them to turn their data into actionable information as quickly and easily as possible. Visit www.rapidinsight.com.

