



# The Continuously Improving ROI of Trusted Data

How successful firms grow and thrive with trusted, clean, reusable data – and the automation that ensures 95 percent data accuracy or better.

---



# What You'll Find In This Report

## A CHALLENGING QUESTION

ROI sounds simple: will we get more out of this business investment than it costs?

The reality is anything but. Calculations of return on investment (ROI) usually focus on gains in productivity and efficiency – will we save money; can we raise output for the same amount of input; will revenue increase? However, making changes in complex organizations has implications that are not obvious.

Business models may need to be revised to take advantage of what an investment has made possible.

New product opportunities or better ways to compete can emerge.

These effects can be hard to envisage and even harder to calculate, but it would be a fatal error to dismiss them as secondary. Not only can they fundamentally alter ROI assessments, but failing to appreciate their impact can have lasting repercussions. The ROI of business investments can change an industry's winners and losers, or even fundamentally alter an industry itself.

## THE ROI OF DATA IN THE MORTGAGE INDUSTRY

Data is an important driver of business costs and opportunities in many industries, not the least of which is financial services and especially mortgage lending. The mortgage industry has accepted high data costs as a fact of its business model and the means of minimizing risk in both origination and secondary markets.

Data is so pervasive and indispensable in lending that calculating the ROI of data that is the source of lending intelligence is an exceptionally difficult challenge. However, the potential impact of new data processing technologies, including automation and artificial intelligence (AI), has presented mortgage industry leaders with an ROI question they cannot avoid.

This paper considers the ROI of data in business, first generally and then in the context of the mortgage industry. It considers the role, impact and value of trusted data in contemporary business, before exploring three areas of ROI that may have a significant impact in the mortgage industry. It has been researched and written by some of the most seasoned analysts in the fields of business technology, ROI and disruptive analytics technologies including AI.

We hope that the questions and conclusions in this paper will be helpful as you evaluate the need, effects and timing of investments that will change how your mortgage business handles data.

# Table of Contents

<b>The ROI of Data</b>	<b>04</b>
Characterizing Data	04
The Data Flywheel Effect	06
How Clean Data Create Trust, Value, And ROI	07
Improving Trust in Data via Self-Improving Automation	08
Intelligent Automation of Clean Data	09
<b>The ROI of Trusted Data in the Mortgage Industry</b>	<b>10</b>
Business Elasticity	10
Cost to Correct	13
Exponential Potential	13
<b>Summary</b>	<b>15</b>
Important Questions	16
<b>Writers</b>	<b>17</b>
<b>Learn More</b>	<b>18</b>

# The ROI of Data

In our work with financial services organizations, we find that clean, reliable, trusted data is one of the most effective investments to deliver both immediate and ongoing ROI, including margin improvement, competitive advantage and business growth. This paper examines how discovering, nurturing, maintaining, and exploiting clean, trusted data throughout the business enables this ROI.

Though it may seem outdated in our digitally-driven 21st century, physical documents remain a core source of essential data in many industries – the mortgage industry among them. Almost any business process that relies on physical

documents or, more typically, images of physical documents (e.g., scans, photos, faxes) must handle diverse streams of unstructured data alongside structured data available from an increasing array of digital sources.

In mortgage lending, unstructured and structured data types need to be combined and coordinated to serve multiple work streams. They must be managed in accordance with exacting and frequently changing regulations, and their availability and reliability must keep pace with changing business conditions. The highest cost and source of risk for most such firms come from errors in workflows that rely on often dirty data.

## CHARACTERIZING DATA

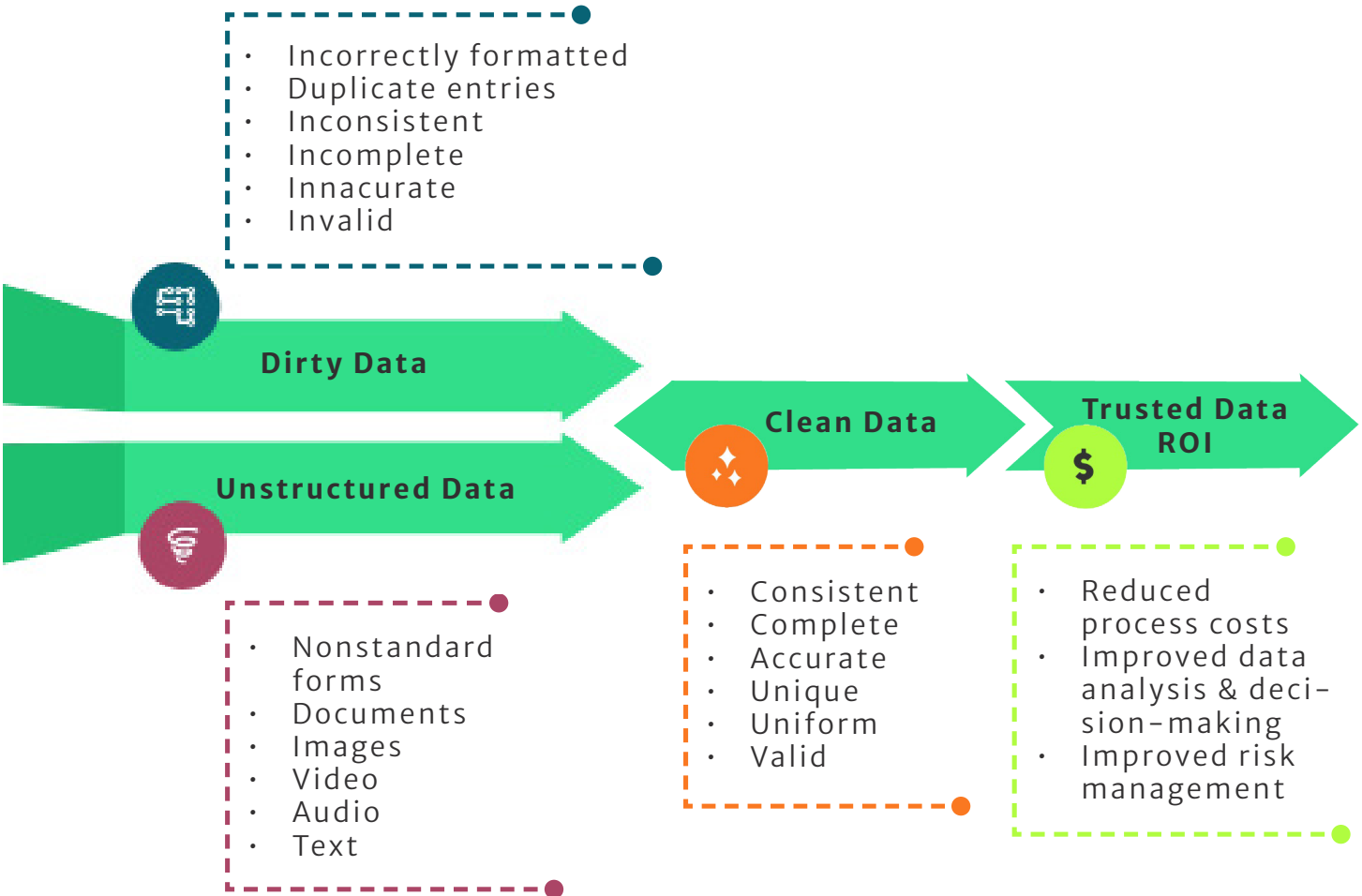
To err is human. When humans enter unstructured data into systems and processes, they can miss errors and sometimes add to them. Cassie Kozyrkov, Chief Decision Scientist at Google, has written about what she calls the Philadelphia problem that occurs when people living in the City of Brotherly Love enter their city’s name in a text field on a form but, well, they make mistakes! Here’s the list of spellings she uses to illustrate dirty data:

Phiadelphia	Philadaelphia	Philadephia	Philadelphia	Philadelphia
Philadelpha	Philadelphia	Philadelphia	Phladelphia	Philadelphia
Philadelpohia	Philadelphia	Philadalphia	Philadelphia	Philadelphioa
Philadelphia	Philadephia	Philly	Philadelhia	Philadriphia
Phialdelphia	Philadelphia	Philadelphia	Philadephi-	Filly
Philadelphai	Phila	Philidelfhia	aphia	Philadelp
Philadelpphia	Pihladelphia	Philadelphia	Philadlphia	Philadelphia
Phildadlphia	Philadephila	Philadelphia	Pholadelphia	Philaelphia
Phidelphia	Phildepphia	Philadelphia	Philadelhpia	Filadelfia
Philadelphi	Philad	Philadelaphia	Philadelphila	Philadelphia
Philadelpha	Philadelphia PA	Philadelphiap	Philadphia	Philla

**Figure 1**  
The Obscure Art of Data Design.  
Source: Towards Data Science, Cassie Kozyrkov, 2022.

Dirty data – structured and unstructured – can be found in native and imaged documents, email messages, written survey responses, transcripts of call center interactions, or even social media posts. Dirty data is incomplete, incorrect, invalid, unreadable, irrelevant, duplicated, or improperly formatted – and usually, combinations thereof.

Dirty data is bad for business, leading to increased costs and reduced productivity, but also introducing the potential to damage reputations, harm relationships, and ultimately limit business opportunities. In short, lousy data engenders distrust in the business. And as more business processes share more data, errors, costs, complexity, and distrust are compounded.



**Figure 2**  
Data -from Dirty to Clean to Increasing ROI.  
Source: Addressable Markets LLC, 2022.

## THE DATA FLYWHEEL EFFECT

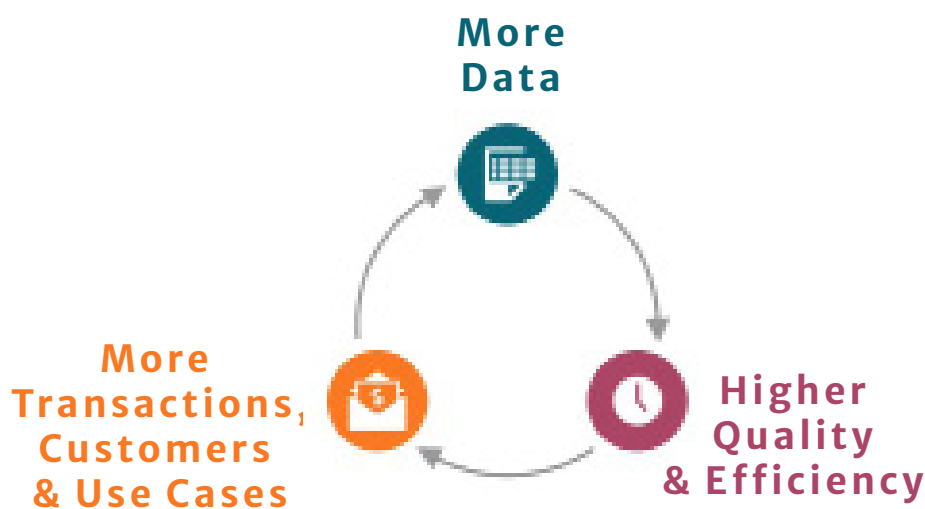
It's a fool's errand to assume all it takes to "solve the problem" of laborious business processes is to invest more in automation technology. Most all business processes rely on clean and readily available data; what can deliver the greatest ROI immediately and continuously? More technology is not (really) the answer. Better data is the answer, and technology can enable better data.

Data is one of the most critical business differentiators. High-quality data – rich, proprietary, accurate, and available – provides the building blocks to drive more efficient,

lower-cost processes that business technology can ultimately utilize.

High quality data does not necessarily mean higher costs. Cost of quality (COQ) is a methodology used to define and measure the resources used to maintain quality and prevent product problems, as opposed to the costs resulting from product failures (whether internal or external). COQ is sum of two factors: the cost of good quality and the cost of poor quality. This is more complex that it appears: the cost of improving quality may ultimately be lower than the status quo and improved quality will reduce product failures and their associated costs.

With data quality, there is a virtuous cycle – an accelerating flywheel effect – driven by discovering and exploiting more and better data. As illustrated in **Figure 3**, more great data can drive greater quality (risk reduction) and efficiency, that will attract more customers, add volume to transaction handling, which produces more data, which can further drive quality and efficiency, and so on. Dirty data drives up operational costs, creates liabilities, and kills business. It runs the flywheel backward.



**Figure 3**  
Data Flywheel: Valuing Data as an Asset.  
Source: © Laura Veldkamp, Columbia Business School,  
Published by Oxford University Press (2023) on behalf of  
the European Finance Association.

## HOW CLEAN DATA CREATE TRUST, VALUE & ROI

If data is to become a critical business differentiator and an enabler of continuously improving ROI, it must be reliable. Reliable data, at every stage of every process, enables firms not only to reduce operational costs and risk but also to sustain and leverage these improvements over years – and to grow more business with more profit in volatile markets.

To make data reliable – to create **trust** in the data – it must be cleaned. Clean data means data that is correct and consistent everywhere in the business. Data is made clean by removing or repairing the data problems listed in **Figure 2**. The earlier in the business process that data is made clean, the cheaper the process becomes, and the less risk the process generates.

The ROI of clean data begins immediately, and recurs and expands.

For example, an upfront improvement in data cleanliness from 80 percent<sup>1</sup> to 99 percent<sup>2</sup> complete and accurate (or higher with the best solutions) can save the business much more than the 19 percent improvement. The clean data itself becomes a multiplier applied to every process using that data.

As that spinning flywheel of data as an asset extends more deeply into the business and trust in data grows, staff become more efficient, confident, and able to accomplish more. Many of our CFO clients report staff efficiency and productivity improvements throughout their Finance organizations as their use of clean data reduces the need for staff involvement per process step. This frees hours of skilled labor for tasks that add value, such as sales and customer support. It also enables (and may pay for) retraining/reskilling to develop even more value from staff.

### What's the Problem with Automation?

Businesses that are forced to use dirty and unstructured data typically employ combinations of IT and humans to gather, enter, and clean it, typically at the first stage of a business process. Many firms have invested in automation to reduce the need for human labor in data cleansing and to speed processes along. Figure 2 represents a combination of human and automated IT reviewing and cleaning data.

Too often, we see some automation adding limited value because it works on only parts of the data or process. It may still require substantial and costly human intervention. And in too many cases, the automation simply speeds up the progress of bad data, pushing it further into processes and exponentially increasing the costs of using that data.

<sup>1</sup> Human data classification and extraction accuracy rates in the mortgage lending industry typically average 80 percent during the initial phase of the origination process. Rates fall below other human data entry tasks because there are more opportunities for errors to occur due to (a) the input data being unstructured and, in many cases, recorded in physical documents, plus (b) data within documents is not recorded consistently, plus (c) normal human data entry accuracy rates, plus (d) inelastic processing capacity that results in higher error rates when volumes are high.

<sup>2</sup> TRUE Data Intelligence achieves data classification and extraction accuracy rates of around 95 percent initially, rising above 99 percent once analytics engines in the product are trained with a mortgage lender's customer data.

## IMPROVING TRUST IN DATA VIA SELF-IMPROVING AUTOMATION

Automation should, by definition, reduce the costs of human intervention in many business processes. But we minimize or even negate ROI when we automate things that don't already work correctly.

What happens when we effectively automate the process to correct itself? Our client work indicates an increasing scope of examples where self-improving process automation reduces the need for staff intervention and

enables continuous quality improvement while increasing trust in the data itself.

Effective automation enables increasingly accurate – trusted – data at every step in as many processes as possible, from as many sources and data types as needed. Automate this intelligently, and we can develop ever-more-valuable data automatically with little to no added human cost.

### What We Really Mean by AI

Artificial intelligence (AI) is something of a misnomer: the machines aren't thinking as humans do. Not yet, at least!

The most common AI reality is machine learning – algorithms that process data in ways that imitate the way that humans learn and gradually improve in accuracy.

Machine learning is well-established: IBM pioneered it by “teaching” a mainframe to play checkers in the 1950s. It is widely used in automated systems programmed to identify what's acceptable and not acceptable, and to adapt system operations based on what they see, do and record.



## INTELLIGENT AUTOMATION OF CLEAN DATA

Machine learning (ML) enables an automated ability to convert unstructured data into structured data. The self-learning capability inherent in ML allows human intervention in business processes to be significantly reduced and, over time, almost entirely eliminated. This not only provides consistent levels of data cleanliness but also allows greater volumes and a more comprehensive range of data to be cleaned, including data that currently might not seem useful. And as data sources, types and regulations proliferate and diversify, ML allows us to accept, clean, use and trust more types of data more easily, more cheaply and more accurately in more areas of a business at little to no additional cost.

ML also enables significantly increased business intelligence and opportunities. Outside of a firm's own internal operations, intelligent automation of data processing and analytics enhances almost any firm's ability to perceive, identify and analyze the impact of potential market disruptions before these disruptions occur. This means that ahead of market upswings and downturns, businesses need to adapt staffing, operations and even business models. Intelligent automation with data analytics even helps position the firm as a source of market intelligence itself. Clean, trusted data can be a valued, saleable commodity in most markets.<sup>3</sup>

<sup>3</sup> Douglas B. Laney, author: Data Juice: 101 Stories of How Organizations Are Squeezing Value from Available Data Assets, and Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage.

# The ROI of Trusted Data in the Mortgage Industry

## PART 1: BUSINESS ELASTICITY

The mortgage industry is inherently cyclical, with volumes in home buying and associated mortgage lending following well-established annual patterns. When we look at loan originations over time, as in **Figure 4**,

the quarter-to-quarter variability in volume indicates a relentless challenge for mortgage professionals to build, operate, measure and adjust the processes to run the business at sustainable margins.

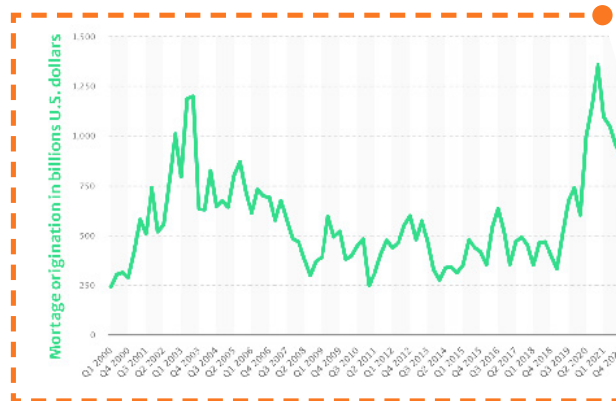


Figure 4: Total mortgage originations in the United States from 2000 through the fourth quarter of 2021. Source: Statista <https://www.statista.com/statistics/205937/us-mortgage-originations-since-1990/>

Each peak represents a burden of work for lending professionals, each valley a lack of operational efficiency. Traversing the two bring more problems: delays for borrowers, increased risk as operations teams struggle to maintain throughput, or changes to staffing or outsourcing levels in response to rising and falling demand. Adding or removing capacity creates the conditions for future pain as changes will inevitably need to be reversed as the cycle swings back.

Confusing the picture are multiple direct and indirect factors: economic growth, inflation, interest rates, availability of staffing, rates of pay, new regulations, social trends, and more. These may significantly

compound or moderate the effect on volumes, yet they are difficult to calculate or even know. The complexity challenges mortgage industry executives to identify what issues to focus on to manage sustainable, resilient, scalable and efficient operations.

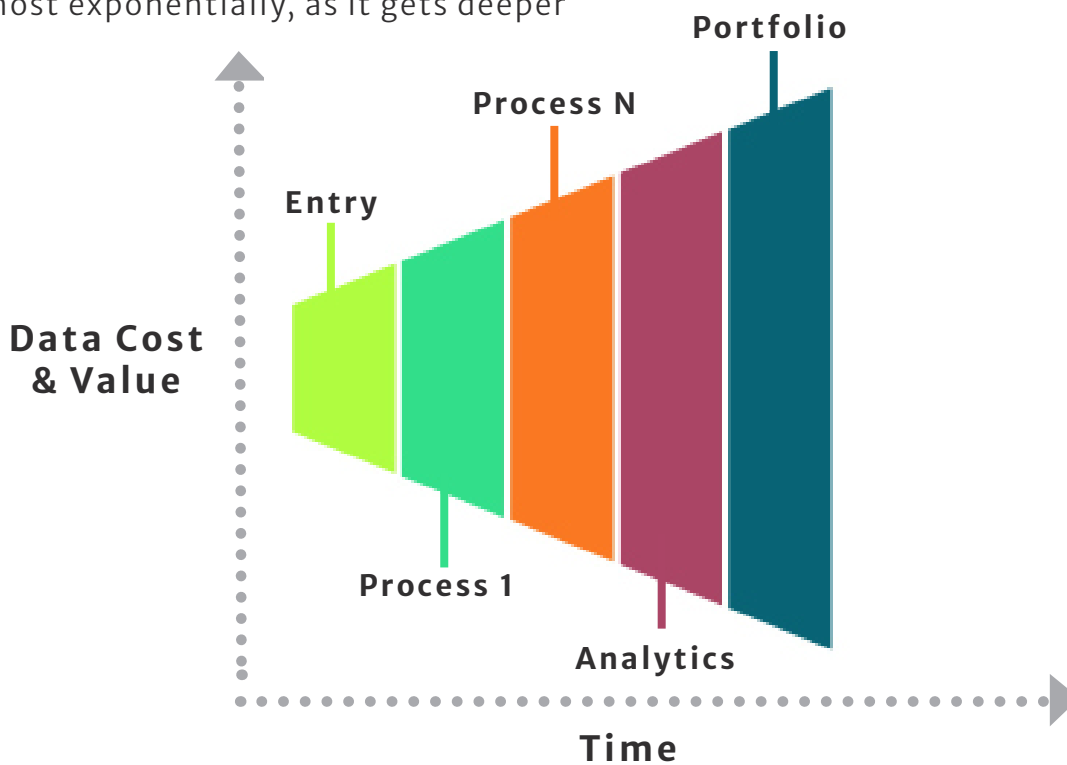
Mortgage lenders have long sought to improve elasticity in their business model, but productivity improvements are masked by the investment necessary to facilitate continual adaptation in the face of unpredictable variability. The answer is to uncouple the relationship between volumes and staff numbers, an outcome that can only be achieved by the intelligent automation of clean data.

## PART 2: COST TO CORRECT

“The sooner, the better” is a motto to live by when facing up to mistakes. It applies equally well to fixing data errors in business processes.

**Figure 5** illustrates how the relative cost of using bad data increases, almost exponentially, as it gets deeper

into business processes. Put another way, the further data progress through workstreams, the greater its value. Conversely, the cost impact of bad data on that workstream also increases at least at the same rate.



**Figure 5. Data Value & Cost Impact Over Time.**  
Source: International Institute of IT Economics, 2023.

Consider an incorrectly transcribed dollar value caused by a poorly scanned bank statement at the beginning of a loan application process. Caught early, the cost of correcting the error is a small amount of time, and the corrected data can be used reliably throughout this process and others. Caught late, that same data error may invalidate everything from the applicant’s qualification through the underwriting – or worse – undermining the value and credibility of the loan portfolio.

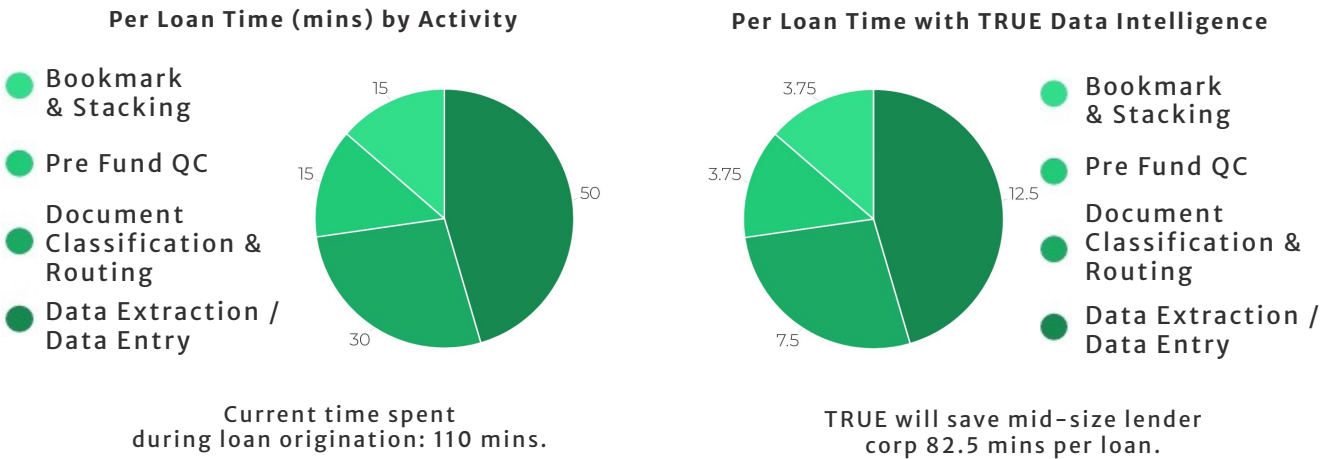
Simply multiplying the time spent at each step of the process by the hourly wages of all the people involved can provide a rough estimate of the initial cost of that bad data. If the entire process needs to be restarted, those

costs get incurred again. And that doesn’t account for business losses that likely occurred, such as the loss of time to perform other work, the potential loss of the applicant’s business, or – if such errors recur enough – the potential loss of current or future business from other applicants and partners.

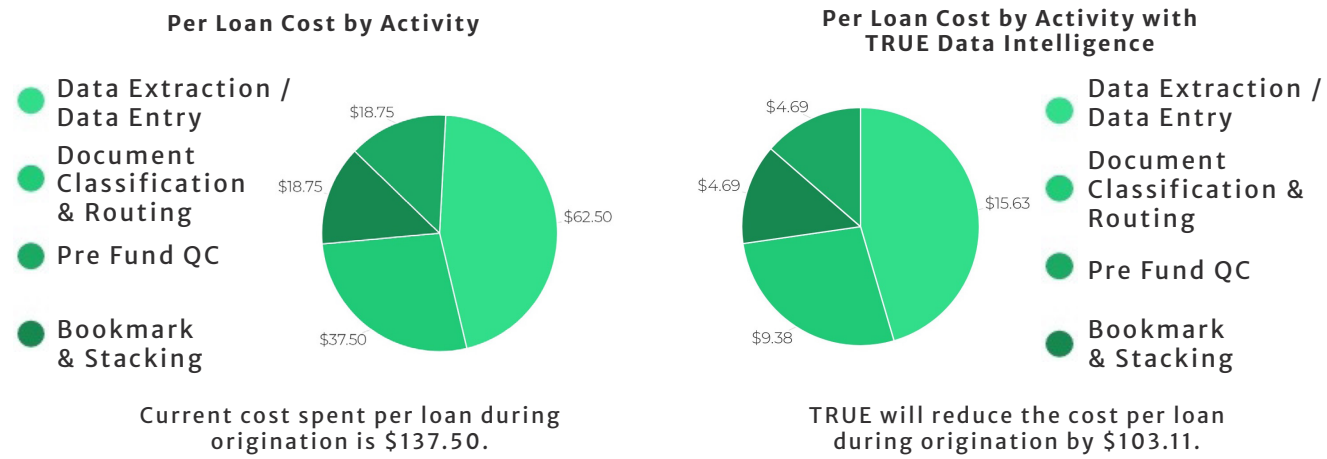
Intelligent automation using clean data significantly reduces the time and financial costs of errors by eliminating the need for human labor to transcribe data, and identify and correct errors. Applying automated clean data processes at multiple stages throughout the loan origination process also helps to prevent data errors from passing initial underwriting and seeping into the loan portfolio.

**Figure 6** shows the typical time and cost savings per loan from introducing automated clean data processes (based on a mid-sized mortgage lender closing 7,200 loans per year – find a suitable play to display full data inputs). **Figure 7** shows the annualized impact.

### Time Savings Per Loan

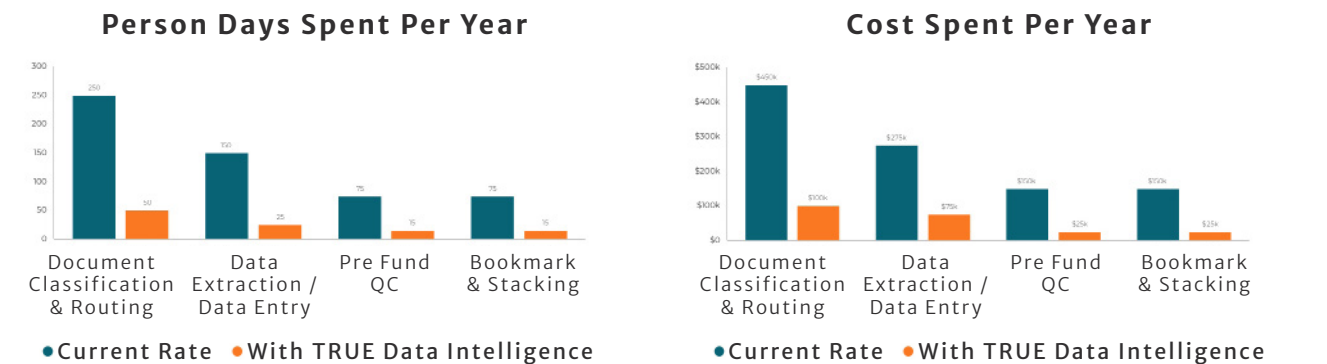


### Cost Savings Per Loan



**Figure 6: Time & cost savings per loan before and after introducing automated clean data processes.**  
Source: TRUE.

### Annualized Benefits for Mid-Sized Lending Corp



**Figure 7: Annualized time and cost savings before and after introducing automated clean data processes.**  
Source: TRUE.

## PART 3: EXPONENTIAL POTENTIAL

As we established earlier, high quality data provides the building blocks for better processes. When the data is proven to be clean, it becomes trusted data. The combination of data quality and trust enables a tremendous range of additional use cases – a powerful multiplier of clean data ROI for business. Trusted data ROI can grow exponentially. It begets powerful analytical capabilities, new product development, closer customer and supply chain relationships, and value of the product portfolio based on it.

Our work with clients suggests that it takes very little time for most organizations to recognize where and how they can expand their use of clean and trusted data, because they can more easily see and understand so much more about their business. It is not unusual to see clients devising and implementing additional process and portfolio improvements within weeks of implementing effective clean data solutions.

## CONTINUOUSLY IMPROVING ROI

The ROI of clean, trusted data combined with intelligent automation grows exponentially as it builds on itself and enables transformation of business processes, especially in data-intense and highly-volatile financial operations.

For example, it is not unusual for clients properly implementing AI/ML-enabled process management software to realize first-year operational cost reductions of up to 30 percent. These cost improvements stem mainly from increased efficiency based on combinations of unified, clean, trusted data and intelligent automation that identify process issues while reducing errors and the problems associated with those errors.

ROI grows from there. In highly-regulated industries like financial services, we see client firms able to leverage their improved business and market intelligence to manage regulatory change more efficiently and rapidly than competitors.

But again, there's more ROI to come. With intelligent automation, process efficiencies can continue to grow as markets change. We see clients with operational efficiencies improving as much as 5 percent to 10 percent annually for three or more years as their ability to use trusted data in more areas of the business expands and improves over time. The value and impact across the business is cumulative, as **Figure 8** explains.

Work Stage	Entry	Process 1	Process N	Analytics	Portfolio
Task	Quality Control	Operations Control	Operations Control	Decision Review	Audit
Impact	Redo	Repair	Repair	Validate	Mitigate
Operation	Test	Restart Rerun Process	Restart/ Rerun Multiple Processes	Resubmit Analysis	Damage Control
Cost/ Value Factor	Labor \$	Labor \$\$	Labor + Revenue \$\$\$	Opportunity \$\$\$\$	Reputation \$\$\$\$\$

**Figure 8: Cumulative Data and Impact.**  
 Source: International Institute of IT Economics, 2023.

**This extended ROI** also includes the ability to create new lines of business – including product innovations or entry into new markets – at the lowest possible cost and with more accurate planning. As an example: a US-based financial institution client firm used cloud-based, intelligent automation to clean and analyze transactional data to reduce data errors, handling costs, and process times, which improved its credit processing accuracy and efficiency while reducing customer and partner support costs.

**Trusted data** everywhere also enabled the client to explore previously unseen opportunities for new products, like credit and financing offerings tailored to individuals, affinity groups, and businesses by geography, profession, industry, and lifestyle. And it enabled new data products that were valuable to merchants and banks, to purchasing and spend management software and services providers, and to inventory and supply chain applications vendors and services providers.

# In Summary

- Clean data enables immediate operational cost reduction, and creates trust in data throughout the business.
- AI/ML-optimized, intelligent automation enables faster, repeatable, more accurate, and cheaper data cleansing at every step of every process.
- Clean data combined with intelligent automation enables exceptional data trust and value within and outside the business.
- Properly-implemented solutions that enable trusted data through intelligent automation can deliver first-year cost improvements of between 20 percent and 30 percent across the business, and ongoing improvements annually.
- The continuously improving ROI of trusted data includes: the ability to minimize and manage risk, enhanced market intelligence, smarter planning and decision making, more robust business models, new product innovations and markets, a better customer experience, and brand value.

# Important Questions

- ? How does your firm liberate value from customer data?**  
Exploiting your data in ways that are largely similar to competitors does not create value that differentiates your products or the customer experience.
- ? What is the value to your business of unstructured or dirty data?**  
Mortgage lenders have to process unstructured data sources and the human effort required is costly and time-consuming, leading most firms to focus on what is necessary rather than what is possible.
- ? How much of your potentially available data do you use and how much do you trust?**  
Capturing richer data is one thing, but putting it to use requires data that is trusted and seen as a worthwhile resource for exploitation.
- ? What business improvements would result from a 20 percent improvement in data quality?**  
Gains in productivity and efficiency are important, but the easiest to envisage. Consider how better-quality data might feed a cycle of continuously improving ROI.
- ? What current or future key performance indicators (KPIs) would best represent a successful implementation?**  
The business case for clean data solutions depends on your ability to capture and monitor metrics that can reliably reveal business impact.
- ? How much of what you spend on data cleansing and analysis is a variable cost vs. a fixed cost – and what are those costs?**  
In most mortgage businesses, human intervention is required at multiple stages through loan origination and represents a sizeable fixed (or inelastically variable) cost. Automation can turn data processes into a responsive variable cost, reducing the pain of transitioning to cyclical volume changes.
- ? What have you learned from experience with mortgage process automation? How can the experience be improved?**  
Many mortgage lenders report a lack of ROI from their loan origination system (LOS) due to a lack of process automation in these platforms. Intelligent automation of clean data unlocks ROI from sunk LOS investment.
- ? Where can your firm gain value from intelligent (AI/ML-based) automation? And how can that automation best be implemented for your firm?**  
As intelligent automation takes hold in the mortgage industry, the firms that derive most value will be those who conceive of and implement differentiated ways of exploiting data.
- ? Which solution vendors offer these capabilities tailored to firms like yours?**  
Many mortgage lenders are reluctant to replace existing productivity solutions, despite underwhelming outcomes. Identify vendors that can integrate with your technology stack and unlock the ROI existing investments. The right partner will fit with existing goals, such as touchless automation, and be able to deliver change without major disruption.



## About the Authors

**Tom Austin** headed Gartner Research and its Maverick program, helping launch and build Gartner's AI research, and leading research on the business value and realities of AI, including how fusing AI with other tech enables new kinds of digital value. Tom played a key role in early identification of such major disruptive tech trends as social software, client-server and cloud architectures, and implementation and the impact of commoditization and consumerization of IT. As co-founder of The Analyst Syndicate, he has led a broad range of tech and business analysts in cooperatively developing leading-edge insights into disruptive and emergent IT. He is also the founder of The Austin Group, LLC (a personal consultancy to executives and enterprises).

**Bill Kirwin** developed the original IT total cost of ownership (TCO) models and associated research, consulting and benchmarking programs at Gartner. Working at the edge of technology and disruption, Bill is a thought leader and a trusted advisor to IT vendors, buyers and users. Bill has a comprehensive background in all stages of IT management combined with understanding the application of technology to business challenges. A founding member of the Value Selling and Realization Council, Bill is also currently the Managing Partner of the International Institute of IT Economics and Chief Research Officer for Sitarian Corporation.

**Bruce Guptil** has developed and led business IT cost and value research, advisory and consulting groups from early days of PCs through the latest digital transformations. In the early 2000s, his work on IT-as-a-service helped shape cloud investment and business strategies for IBM, SAP, Microsoft and more. In the 1990s, he led research for Gartner and others on online marketplaces, web commerce, and the ROI of new IT. Today as Chief Strategist of Addressable Markets LLC, Bruce analyzes and translates changing user business needs into IT developer and service provider offerings and market planning. Bruce also leads Provider Lens comparative research on Procurement services and platforms, and U.S. Public Sector IT, for global tech consulting firm Information Services Group (ISG).

**James Ollerenshaw** is a marketing strategist focused on innovative and disruptive B2B technology, especially in the fields of financial services, artificial intelligence, automation and robotics. He co-founded and chaired a working group for entrepreneurs for the Institute of Directors. He led the Technology Marketing Council EMEA, a professional group for marketing leaders of B2B tech businesses, for global analyst and consulting firm Forrester Research. James provides strategic advisory to tech businesses in the U.S. and U.K. and is Chairman of Curzon, a communications consultancy based in London, U.K.

## Dig **Deeper**

Learn more about the TRUE Platform and find out how the right lending intelligence solution can power the entire loan decision process with clean data.

**Visit [TRUE.ai](https://TRUE.ai)**