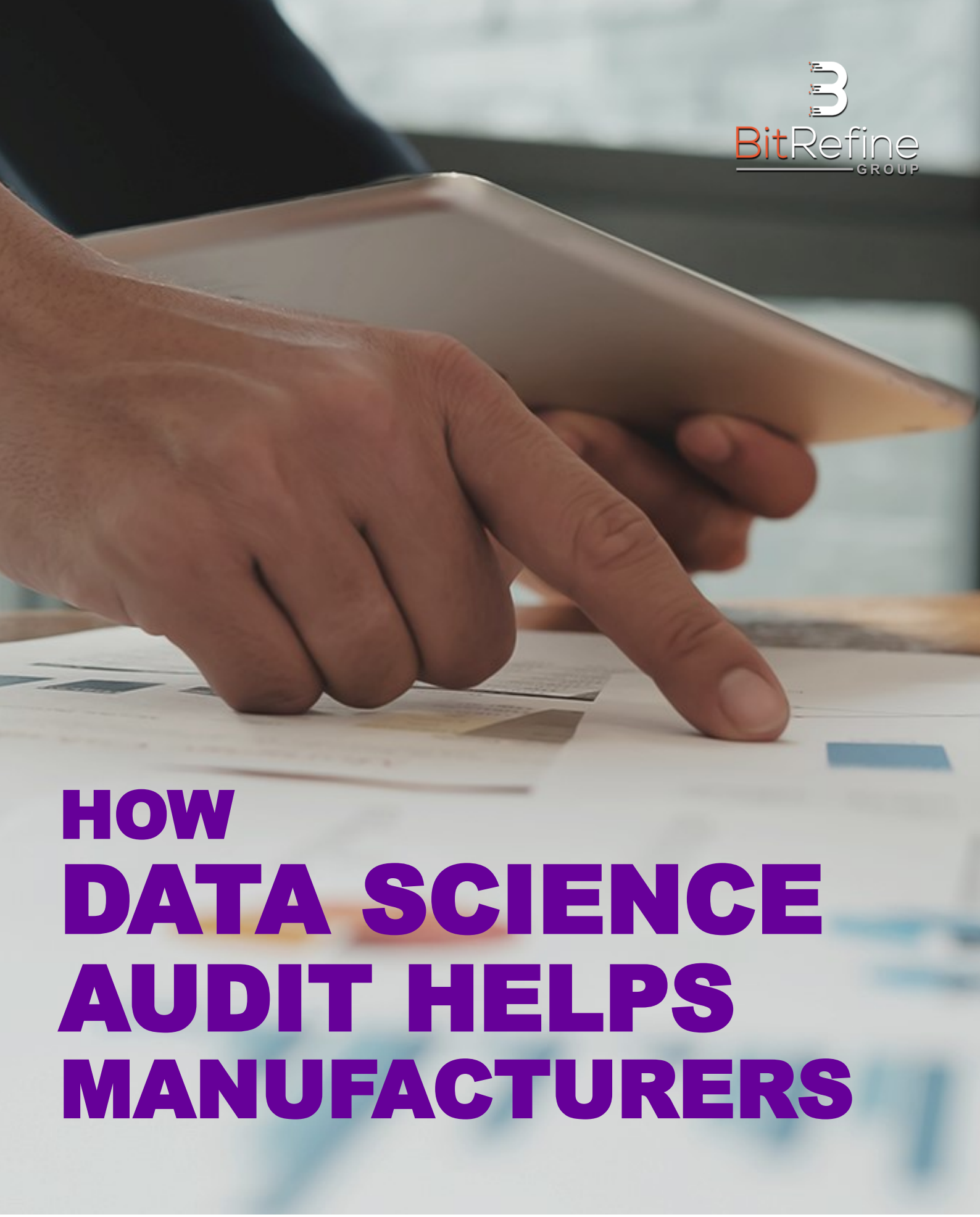


BitRefine GROUP

HOW DATA SCIENCE AUDIT HELPS MANUFACTURERS

Today, implementing data science isn't an option; it's mandatory. The question is whether you're capable of implementing data science across your organization and maximizing the benefits.

Purpose of this document

The idea of employing data science for business growth is becoming increasingly common. In today's world, large datasets are easily available. Every process has undergone 'datafication' i.e., been converted into data. Corporate executives must formulate governance plans and strategies along with create accompanying policies and procedures to enable the path to transformation into a data-driven business.

However, many manufacturers recognize that they have to clear hurdles to adopt machine-learning technology in most areas. Organizations find it challenging to see the value of the data and reveal business opportunities that their data could provide.

This paper demonstrates how an audit of data assets along with business processes can help your company identify the areas of greatest opportunity in the adoption of data-driven solutions. It summarizes key components of strategic audit of data assets and provides details about the steps and methodologies required.

Why consider data science, machine learning, and artificial intelligence?

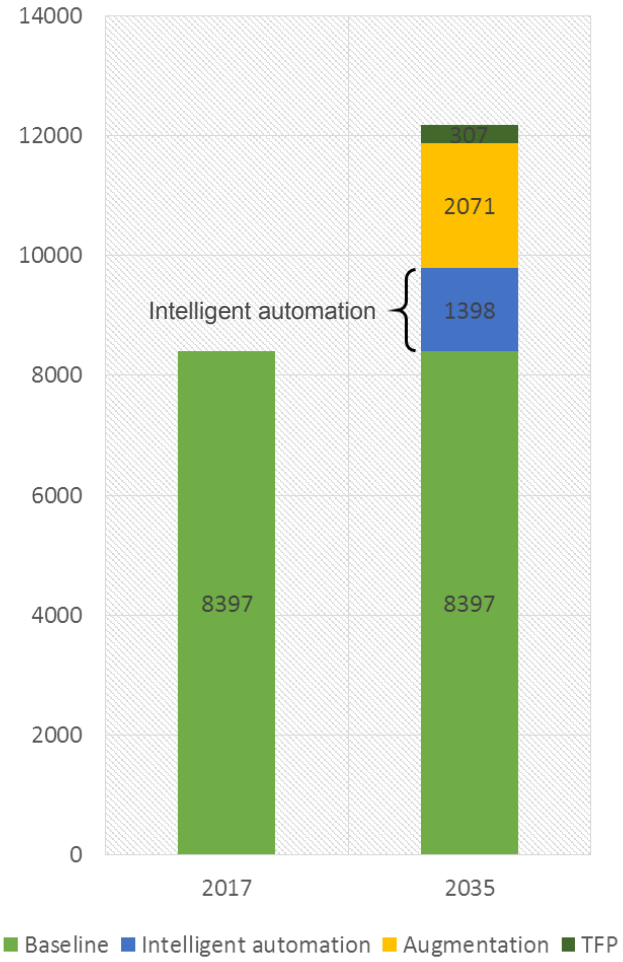
Industry trends

For the last five years at least, machine learning has been an IT buzzword. Applications and uses for machine learning are evolving in various industries. A recent report by Gartner stated that Data Science, AI, and Machine Learning are recurring themes among 2018 tech trends and they will continue to lead the technology marathon.

In manufacturing, the ability to apply a data-driven approach to enhance diverse processes will drive the payoff for digital transformation over the next decade. Machine-learning algorithms will help companies gain value from Big Data obtained from various resources such as sensors, logs, or the internet of things (IoT).

There is no doubt that, by 2025, most manufacturing industries will integrate ML into their processes to boost profit (or savings). At a global scale, the industrial sector generates revenues over $25 trillion. ML applications have the potential to save 5% (approx. $1 trillion) or more without heavy investing.

Other research by Accenture and Frontier Economics shows that, with AI, manufacturing can generate an additional $3.8 trillion in gross value added by 2035. There are three main channels estimated: intelligent automation, augmentation, and total factor productivity. Intelligent automation is expected to drive profits of 1.3 trillion for the manufacturing sector.



Value of data

The reasons the message that data is an important business asset is repeated over and over are obvious. Today, thanks to big data and IoT more and more people are paying attention to data's value as an asset.

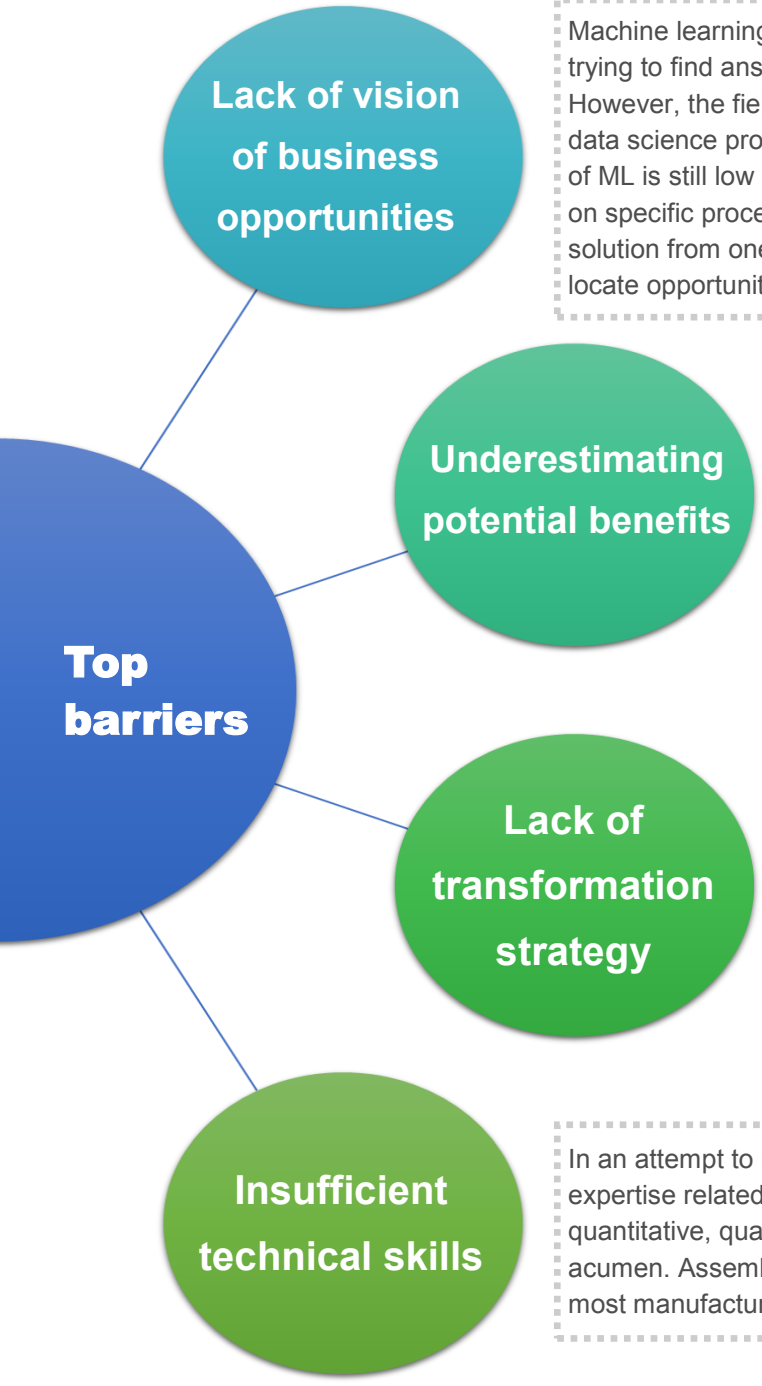
Most industrial manufacturing companies have highly data-intensive processes and generate enormous amounts of digital information. Data lakes store all the data with ease, but only data science, machine learning, and AI provide the opportunity to unlock a rich spectrum of real, measurable benefits.

Manufacturers are leveraging data assets to improve three general areas—to increase yields, reduce costs, and improve competitiveness by integrating intelligence into the products themselves. The actual range of possible use cases is limited only by the level of understanding of company's business processes and available data. In terms of improving process efficiency, proper use of existing data assets can bring anywhere from two to twenty percent of cost savings.

What are the barriers to data-driven business?

Technology has made collecting and storing data relatively cheap, so companies have started collecting data without any clear idea of how to use it. They are collecting everything, simply because they can. And today manufacturers have the opportunity to use these vast digital resources. However, the question for most companies is still, "How can we use our data to gain competitive advantage?"

A company needs to start with a clear vision of the business process it is trying to improve. With an objective in mind, your company can align this vision with available data, infrastructure, and data science resources. However, for many manufacturing organizations this first step is the most challenging. It requires the company's executives to combine a deep understanding of internal business processes with solid knowledge of data-driven technology the and opportunities it brings. We identify four common barriers preventing organizations from taking advantage of data science:



Machine learning is being widely discussed among researchers who are trying to find answers to many old and new manufacturing questions. However, the field is extremely broad, and this even presents a challenge to data science professionals. Today the number of successful implementations of ML is still low in manufacturing. In addition, ML applications are all focused on specific processes, which means it is difficult to extend an existing solution from one area to another. All this makes it difficult for companies to locate opportunities for ML within their internal structures.

The AI domain has developed so quickly that companies haven't had enough time to form their own assessments of new opportunities. In fact, experts argue that AI-technology brings many more advantages, in a larger area and much faster than a well-established business can comprehend without agile transformation.

Adoption of data-driven solutions requires organizational change. A clear digital strategy helps companies implement their digital initiatives from the top down with leaders who drive the transformation. The latest surveys show that more than 80% of companies that don't have clear digital strategies consider adoption of data-science technologies challenging, whereas more than 50% of companies that have a digital strategy consider data-science initiatives straightforward.

In an attempt to make use of data, businesses need to start accumulating expertise related to data science. Data science requires a blend of quantitative, qualitative, technical, and leadership skills along with business acumen. Assembling a team with all of those skills takes time. Currently, for most manufacturers, this is the step that is yet to be undertaken.

How does data-science consulting help?

To respond effectively to today's challenges, organizations need to capitalize on new AI-based technologies, leveraging their data assets in new ways to improve efficiency and minimize costs. BitRefine Group helps organizations quickly identify how their data can fuel high performance. We combine our deep data-driven strategy and technology implementation experience with clients' domain expertise to plan the adoption of AI-based solutions, helping clients set a technology agenda and position their business for the future.

We team with IT and business leaders to help organizations make the shift and transform to a truly data-driven business. With our assistance, companies learn to rely not just on instinct, traditional experience, or the highest-paid person's opinion, but also fully use the wisdom of data. We help organizations change the basis for making decisions at every level of the company.

We reveal immediate opportunities for data-driven solutions

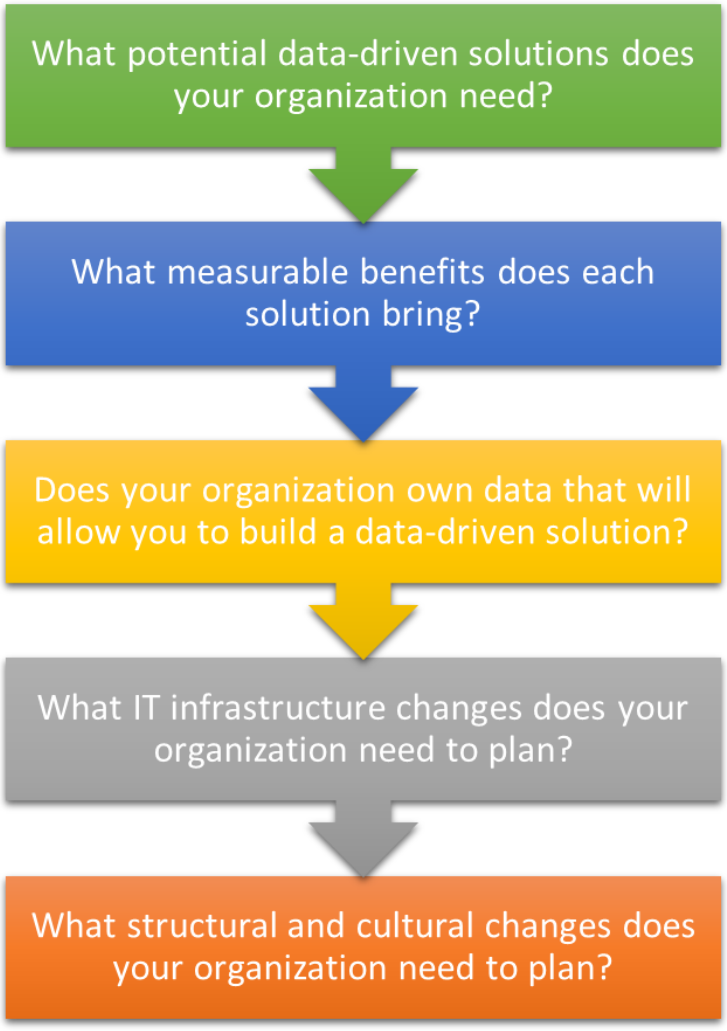
To identify the artificial intelligence and deep analytics opportunities for your company, we map your business processes against available data.

First, we team up with leaders of different departments and take an inventory of business processes to identify appropriate chances to invest in machine-learning capabilities. Next, we audit available data assets to understand if the data is suitable for machine-learning modeling and is capable of achieving defined business goals.

After the mapping is completed, the client gains a full understanding of the current state of data-driven automation in his company, including how it is currently used and its possible applications. Organizations that have involved us in their data-driven transformations typically have revealed dozens of fresh opportunities and reduced their operational costs by anywhere from five to 20 percent.

We show you the road for future development

Business and technology leaders must view data science, and cognitive technology not as a one-off project, but as an across-the-board functionality, propelling innovation throughout the enterprise. We help companies develop a roadmap for your business and answer the following questions:



How a data-science audit works

To leverage the potential of data, an organization must first be aware of the availability and quality of its digital assets. Second, the organization must evaluate whether the data will bring measurable benefits to any of its business processes. Conducting an internal or external audit will provide this information, raising awareness of the present situation and the issues involved in improving your overall digital strategy. An organization that is knowledgeable about its data's potential puts itself in a position to maximize its value through effective use. Here, we present the main steps in data-science audit methodology.

Planning the audit

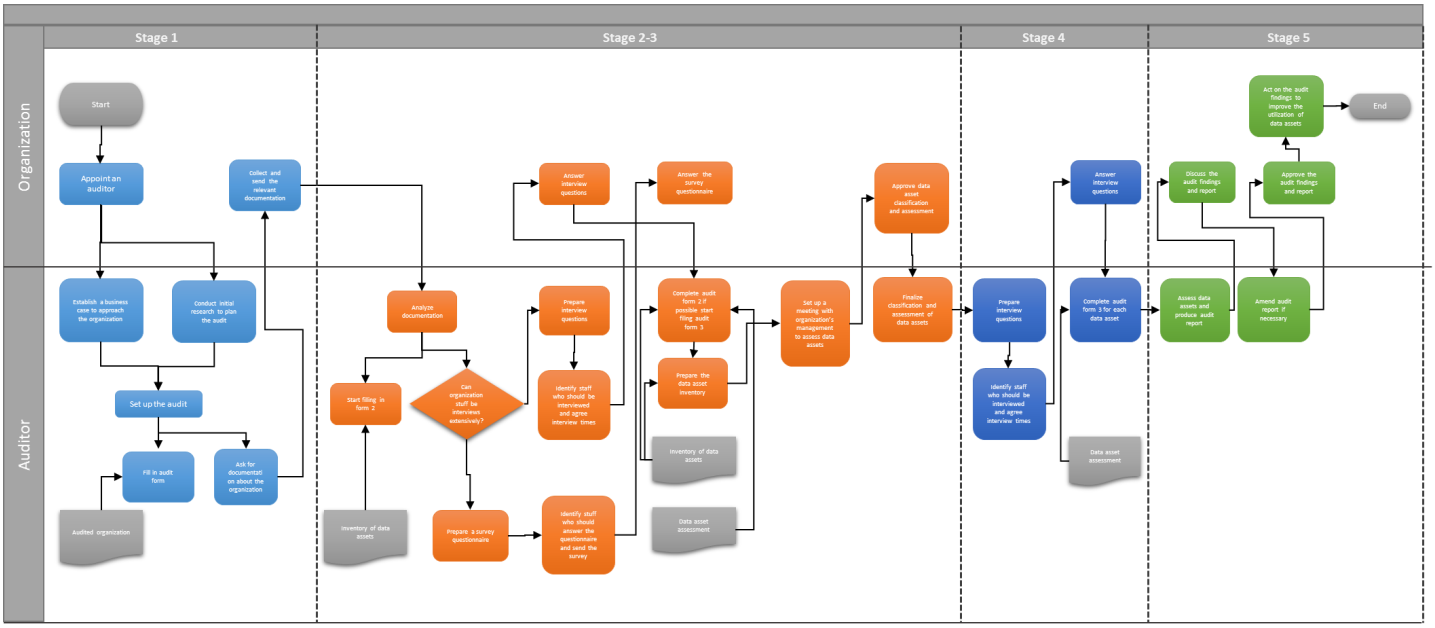
The key objective of the planning stage is to prepare as much as possible in advance so the time spent on-site is put to the best use. By conducting preliminary research, we also minimize demand on our client's staff as we become better informed about the context. In addition, scheduling interviews in advance ensures that staff will be prepared to contribute to the audit.

We begin by identifying the main individuals responsible for data assets and target business processes. This allows us to prepare an audit schedule, and set up interview times. We explore ways to gain access to target data assets. As gaining access to sensitive data may include setting up approvals, we request corresponding documentation at the earliest opportunity.

As soon as we have studied the available data and identified the key personnel we'll be working with, we prepare questionnaires and send them through to the departments. Finally, we schedule interviews to make sure staff are available and have the information requested ready to hand.

By the end, we have completed the following:

- **Obtain general approval for the audit to take place**
- **Define the scope of data and documentation to be accessed**
- **Conduct initial research into the target business process**
- **Define key personnel who will support the audit**
- **Send interview questions and requested documentation in advance**
- **Agree on dates for the audit to take place and set up interviews with key personnel**



Identifying and classifying data assets

The aim of this stage is to draw a complete picture of the existing data assets and to classify them according to type. First, we analyze documentation, understanding the context where the data assets are being created, used, and managed. This helps with identification and facilitates the interview process. Depending on availability of staff, we either follow up with interviews or conduct written surveys to locate available data.

As soon as we locate a data asset we explore a sample dataset to classify it and document its properties. The classification applied at this stage is general, containing a description, format, and distinguishing type: static vs dynamic and content e.g. texts, code, images, video, time series. Detailed analysis of the data will be conducted later at the data-examination stage.

By the end, we have completed the following :

- **Contact key personnel via either a questionnaire or an interview**
- **Identify all data assets**
- **Assign all data assets to a category to define the scope of the next stage**
- **Document the audit process and findings**

Identifying business challenges

The purpose of this stage is to identify the business processes that can potentially be improved by data-driven solutions. In some cases, it is likely that the organization has already recognized benefits of a data-driven application for a number of potential business processes. The auditor's goal is to document the existing demand and broaden this list with all possible applications to give the organization a clearer picture of what data transformation means for them.

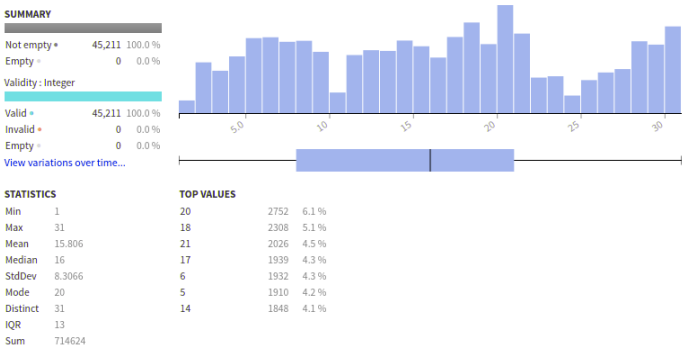
Effective communication between staff and the auditor is crucial here. We set up interviews with business unit leaders to listen to the client and ask questions to ascertain what the data-science challenges are and formulate challenges into coherent data-science problems. We provide the client with detailed advice, clarifying the ways a data-driven solution works and providing ideas for best practices along with examples of past implementations.

By the end, we have completed the following:

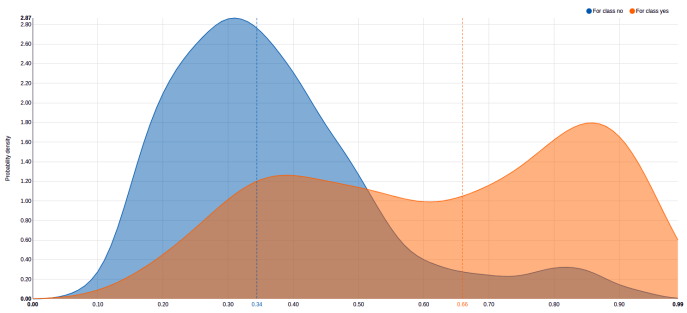
- **Interview leaders of business units**
- **Map suggested potential data-driven solutions**
- **Formulate data-science problems**

Exploring sample datasets

After we locate data assets and collect sample datasets we start exploring them using advanced statistical methods and draft machine-learning modeling. Here we attain an understanding of the given data. We examine both the individual variables and the relationships among them. Histograms and other graphical representations show the shape of data distribution, reflecting skewness and kurtosis. We examine the set for missing data and check options for assessing it. In the end, we estimate chances of building machine-learning models for target business goals. This ensures general evaluation for building a prediction model based on given variables without extensive feature engineering.



During exploration of data related to complex industrial processes, the data scientist may request additional information from staff who own domain expertise. We contact key personnel either via a questionnaire or via an interview to understand the physical meaning of the explored dataset's variables.



This stage will result in a complete understanding of the data's current properties and its possible use for solving the business goals we have identified during previous

steps. In addition, we provide details on how the data can or should be enriched to achieve more accurate target insights.

If the data doesn't present the expected insights and can't help us achieve a business goal, this is also documented along with alternative ways approaching the challenge. Data examination may also extend the list of initial business cases if the data scientist reveals additional patterns in the dataset and comes up with an idea of how to use them in a practical application.

By the end of this step, we have completed the following:

- **Understand properties of datasets**
- **Evaluate chance of using datasets for solving business tasks**
- **Provide recommendations on improving or acquiring additional data**
- **Document the process and findings**

Reporting the results

The final stage is concerned with merging information about the organization's data assets, results of data exploration, and actual business challenges into a complex map. We prepare an initial version of a report that contains a picture composed of all findings that will inform management of the breakdown of their assets according to their value for building data-driven solutions. In the final section of the report we provide recommendations and point the organization toward actions that will lead to an increase in business efficiency.

After the initial draft of the audit report has been completed we present it to management in preparation for a discussion of the findings and recommendations. This allows us to receive feedback from management and correct the final report in view of any suggested changes. At the same time, it gives the organization an opportunity to seek further guidance from the auditor on the best way forward.

Once changes are made, the final version of the audit report is passed along for formal approval and becomes a working document that we use as guidance for planning data-driven transformation activities.

Finally, we have:

- **Collated and analyzed audit findings for inclusion in the final report**
- **Prepared a list of recommended changes**
- **Discussed the draft audit report with management**
- **Issued a final audit report**

Do we need process-specific domain expertise to do high-quality data-science modelling?

A rich dataset allows us to build a model without any knowledge of the target process. This means we don't need to understand the physics of a particular manufacturing process and be an engineer with years of experience to reflect the required rules and dependencies in the code. Machine learning doesn't rely on physics but purely on data.

However, we still do need to spend time immersed in the process details to a certain degree to achieve a more robust solution in the end. As soon as a data science project starts, the data scientist needs to figure out the proper way of treating all the variables that they receive within a dataset. In some cases, the meaning of the presented variables is obvious, such as, for example, price, time, or product amount. In some cases, we need to study the technology of a target process and secure for some level of communication with the client's technical staff to understand basics of the process routes and meaning of variables, such as readings from chromatographs, pressure sensors, or parameters of input materials.

What are the benefits of data science audit?

There's no question that companies can save money and increase efficiency by adopting artificial intelligence. However becoming a data-driven enterprise requires deep shifts. Companies will need to develop awareness of their data collections. They need to develop the motivation to work with data to improve products and practices. They need to deploy diverse data-driven applications and test their performance. Companies need to make data and machine-learning products available to a broader spectrum of employees. Data science audit is the first step down this road.

The information collected during data science audit is sufficient to provide a clear overall picture of organizational data assets, current opportunities, and paths for future development. Organizations armed with this information can begin to undertake activities to implement data-driven solutions.

Why BitRefine Group?

BitRefine provides data-science audit, consulting, and development services to organizations across more than ten industry sectors. We deliver measurable and lasting results that help to reinforce trust in data-driven approaches, inspire clients to make their most challenging business decisions with confidence, and lead the way toward digital transformation.

To this moment, thanks to continuous development and improvement, we have accumulated a wealth of knowledge that we pass on to our clients through our work. Our approach ensures that we can quickly deliver a team with the right experience and expertise to every client, anywhere in the world.

- **Management and technical expertise**

Our team brings considerable consulting expertise involving management and operations reviews, workflow, and process analyses, information technology and systems, project management and controls, quality assurance, and other aspects of business process improvement projects.

- **Extensive data science experience**

BitRefine is an active developer of machine-learning solutions that improve the efficiency of a wide range of processes. We use this strong working knowledge of challenges that organizations face to deliver the best possible results within our audit and consulting services.

- **Technology expertise**

Our technology stack is based on a number of frameworks and platforms that have proved great efficiency, and that have been commonly accepted by developers and backed by either tech giants, or nonprofit organizations.

- **Project management expertise**

Our team has a significant background and experience in managing projects. We keep our work well organized, creating realistic project plans, budgets, time and effort estimates to secure on-time delivery of results.

- **Efficient approach**

We have structured our overall approach in a straightforward, practical way to ensure that the environment we develop projects in will result in high-quality work results and deliverables.

We believe that BitRefine's team is unmatched in experience, balance, and dedication to our clients' business, and that our services is fully responsive to clients' needs

About BitRefine Group

BitRefine Group is a global professional services company that provides a broad range of solutions and services in data science, machine-learning, and computer vision. Combining extensive experience and specialized skills, BitRefine solves clients' toughest challenges, turning emerging technology into a real-world product. Among the applications that BitRefine develops are deep data analysis, visual information comprehension, robotics, automation, medical image processing, and behavior prediction – this is the short list. Working at the intersection of business and technology, BitRefine helps companies improve their performance, stay ahead of trends, and maintain a competitive advantage in the market. Visit us at www.bitrefine.group

BitRefine Group
December 2017

www.bitrefine.group