



Building your data and analytics strategy

The tools every data professional needs to
build a world-class analytics organization



Building your data and analytics strategy

The tools every data professional needs to build a world-class analytics organization.

What's on the chief data and analytics officer's agenda? Defining and driving the data and analytics strategy for the entire organization. Ensuring information reliability. Empowering data-driven decisions across all lines of business. Wringing every last bit of value out of the data. And that's just Monday.

The challenges are many, but so are the opportunities. This e-book is full of resources to help you launch successful data analytics projects, improve data prep and go beyond conventional data governance. Read on to help your organization become truly data-driven with best practices from TDWI, see what an open approach to analytics did for Cox Automotive and Cleveland Clinic, and find out how the latest advances in AI are revolutionizing operations at Volvo Trucks and Mack Trucks.



3

5 ways to become data-driven



14

IoT data with AI reduces downtime, helps truckers keep on trucking



7

10 questions to kick off your data analytics projects



18

How to improve data prep for analytics: TDWI shares best practices



11

Data governance: The case for self-validation



21

Keeping an open mind about open analytics



5 ways to become data-driven

Most organizations believe that data and analytics provide insights, but few describe themselves as truly data-driven.

When it comes to being data-driven, organizations run the gamut with maturity levels. Most believe that data and analytics provide insights. But only one-third of respondents to a TDWI survey¹ said they were truly data-driven, meaning they analyze data to drive decisions and actions.

Successful data-driven businesses foster a collaborative, goal-oriented culture. Leaders believe in data and are governance-oriented. The technology side of the business ensures sound data quality and puts analytics into operation. The data management strategy spans the full analytics life cycle. Data is accessible and usable by multiple people - data engineers and data scientists, business analysts and less-technical business users.

TDWI analyst Fern Halper conducted research of analytics and data professionals across industries and identified the following five best practices for becoming a data-driven organization.

1. Build relationships to support collaboration

If IT and business teams don't collaborate, the organization can't operate in a data-driven way - so eliminating barriers between groups is crucial. Achieving this can improve market performance and innovation; but collaboration is challenging. Business decision makers often don't think IT understands the importance of fast results, and conversely, IT doesn't think the business understands data management priorities. Office politics come into play.

But having clearly defined roles and responsibilities with shared goals across departments encourages teamwork. These roles should include: IT/architec-

ture, business and others who manage various tasks on the business and IT sides (from business sponsors to DevOPs).

2. Make data accessible and trustworthy

Making data accessible - and ensuring its quality - are key to breaking down barriers and becoming data-driven. Whether it's a data engineer assembling and transforming data for analysis or a data scientist building a model, everyone benefits from trustworthy data that's unified and built around a common vocabulary.

As organizations analyze new forms of data - text, sensor, image and streaming - they'll need to do so across multiple platforms like data warehouses, Hadoop, streaming platforms and data lakes. Such systems may reside on-site or in the cloud. TDWI recommends several best practices to help:

- Establish a data integration and pipeline environment with tools that provide federated access and join data across sources. It helps to have point-and-click interfaces for building workflows, and tools that support ETL, ELT and advanced specifications like conditional logic or parallel jobs.
- Manage, reuse and govern metadata - that is, the data about your data. This includes size, author, database column structure, security and more.
- Provide reusable data quality tools with built-in analytics capabilities that can profile data for accuracy, completeness and ambiguity.

3. Provide tools to help the business work with data

From marketing and finance to operations and HR, business teams need self-service tools to speed and simplify data preparation and analytics tasks. Such tools may include built-in, advanced techniques like machine learning, and many work across the analytics life cycle - from data collection and profiling to monitoring analytical models in production. These "smart" tools feature three capabilities:

- Automation helps during model building and model management processes. Data preparation tools often use machine learning and natural language processing to understand semantics and accelerate data matching.
- Reusability pulls from what has already been created for data management and analytics. For example, a source-to-target data pipeline workflow can be saved and embedded into an analytics workflow to create a predictive model.
- Explainability helps business users understand the output when, for example, they've built a predictive model using an automated tool. Tools that explain what they've done are ideal for a data-driven company.

4. Consider a cohesive platform that supports collaboration and analytics

As organizations mature analytically, it's important for their platform to support multiple roles in a common interface with a unified data infrastructure. This strengthens collaboration and makes it easier for people to do their jobs. For example, a business analyst can use a discussion space to collaborate with a data scientist while building a predictive model, and during testing. The data scientist can use a notebook environment to test and validate the model as it's versioned and metadata is captured. The data scientist can then notify the DevOps team when the model is ready for production - and they can use the platform's tools to continually monitor the model.

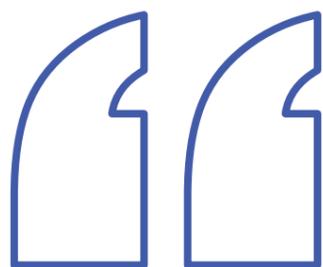
5. Use modern governance technologies and practices

Governance - that is, rules and policies that prescribe how organizations protect and manage their data and analytics - is critical in learning to trust data and become data-driven. But TDWI research indicates that one-third of organizations don't govern their data at all. Instead, many focus on security and privacy rules. Their research also indicates that fewer than 20 percent of organizations do any type of analytics governance, which includes vetting and monitoring models in production.

Decisions based on poor data - or models that have degraded - can have a negative effect on the business. As more people across an organization access data and build models, and as new types of data and technologies emerge (big data, cloud, stream mining), data governance practices need to evolve. TDWI recommends three features of governance software that can strengthen your data and analytics governance:

- **Data catalogs, glossaries and dictionaries.** These tools often include sophisticated tagging and automated procedures for building and keeping catalogs up to date - as well as discovering metadata from existing data sets.
- **Data lineage.** Data lineage combined with metadata helps organizations understand where data originated and track how it was changed and transformed.
- **Model management.** Ongoing model tracking is crucial for analytics governance. Many tools automate model monitoring, schedule updates to keep models current and send alerts when a model is degrading.

In the future, organizations may move beyond traditional governance council models to new approaches like agile governance, embedded governance or crowdsourced governance. But involving both IT and business stakeholders in the decision-making process - including data owners, data stewards and others - will always be key to robust governance at data-driven organizations.



As organizations mature analytically, it's important for the platform to support multiple roles in a common interface with a unified data infrastructure. This strengthens collaboration and makes it easier for people to do their jobs.

Five Data Management and Analytics Best Practices for Becoming Data-Driven

In a survey, TDWI found that one-third of organizations don't govern their data - and fewer than 20 percent do any type of analytics governance. Governance is just one discipline that's essential for becoming data-driven. Learn more in this checklist report from TDWI.

Download free checklist report now 



10 questions to kick off your data analytics projects

There's no single blueprint for beginning a data analytics project, but these 10 questions will help guide you to success

By Phil Simon, author, speaker and noted technology expert

There's no single blueprint for beginning a data analytics project – never mind ensuring a successful one.

However, I have found that the following questions help individuals and organizations frame their data analytics projects in instructive ways. Put differently, think of these questions as more of a guide than a comprehensive how-to list.

1. Is this your organization's first attempt at a data analytics project?

When it comes to data analytics projects, culture matters.

Consider Netflix, Google and Amazon. All things being equal, organizations like these have successfully completed data analytics projects. Even better, they have built analytics into their cultures and become data-driven businesses. As a result, they will do better than neophytes. Fortunately, first-timers are not destined for failure. They should just temper their expectations.

2. What business problem do you think you're trying to solve?

This might seem obvious, but plenty of folks fail to ask it before jumping in. Note here how I qualified the first question with "do you think." Sometimes the root cause of a problem isn't what we believe it to be; in other words, it's often not what we at first think.

In any case, you don't need to solve the entire problem all at once by trying to boil the ocean. In fact, you shouldn't take this approach. Project methodologies (like agile) allow organizations to take an iterative approach and embrace the power of small batches.

3. What types and sources of data are available to you?

Most if not all organizations store vast amounts of enterprise data. Looking at internal databases and data sources makes sense. Don't make the mistake of believing, though, that the discussion ends there.

External data sources in the form of open data sets (such as data.gov) continue to proliferate. There are easy methods for retrieving data from the web and getting it back in a usable format – scraping, for example. This tactic can work well in academic environments, but scraping could be a sign of data immaturity for businesses. It's always best to get your hands on the original data source when possible.

Caveat: Just because the organization stores it doesn't mean you'll be able to easily access it. Pernicious internal politics stifle many an analytics endeavor.

4. What types and sources of data are you allowed to use?

With all the hubbub over privacy and security these days, foolish is the soul who fails to ask this question. As some retail executives have learned in recent years, a company can abide by the law completely and still make people feel decidedly icky about the privacy of their purchases. Or, consider a health care organization – it may not technically violate the Health Insurance Portability and Accountability Act of 1996 (HIPAA), yet it could still raise privacy concerns. Another example is the GDPR. Adhering to this regulation means that organizations won't necessarily be able to use personal data they previously could use – at least not in the same way.

5. What is the quality of your organization's data?

Common mistakes here include assuming your data is complete, accurate and unique (read: nonduplicate). During my consulting career, I could count on one hand the number of times a client handed me a "perfect" data set. While it's important to cleanse your data, you don't need pristine data just to get started. As Voltaire said, "Perfect is the enemy of good."

6. What tools are available to extract, clean, analyze and present the data?

This isn't the 1990s, so please don't tell me that your analytic efforts are limited to spreadsheets.

Sure, Microsoft Excel works with structured data - if the data set isn't all that big. Make no mistake, though: Everyone's favorite spreadsheet program suffers from plenty of limitations, in areas like:

- Handling semistructured and unstructured data.
- Dealing with size restrictions.
- Tracking changes/version control.
- Ensuring governance.
- Providing security.

For now, suffice it to say that if you're trying to analyze large, complex data sets, there are many tools well worth exploring. The same holds true for visualization. Never before have we seen such an array of powerful, affordable and user-friendly tools designed to present data in interesting ways. For instance, SAS® Visual Analytics, SAS Visual Data Mining and Machine Learning, and several open source tools are just some applications and frameworks that make dataviz powerful and, dare I say, cool.

Caveat 1: While software vendors often ape each other's features, don't assume that each application can do everything that the others can.

Caveat 2: With open source software, remember that "free" software could be compared to a "free" puppy. To be direct: Even with open source software, expect to spend some time and effort on training and education.

7. Do your employees possess the right skills to work on the data analytics project?

The database administrator may well be a whiz at SQL. That doesn't mean, though, that she can easily analyze gigabytes of unstructured data. Many of my students need to learn new programs over the course of the semester, and the same holds true for employees. In fact, organizations often find that they need to:



Phil Simon
Author, speaker and
technology expert

What will an individual, group, department or organization do with keen new insights from your data analytics projects? Will the result be real action? Or will a report just sit in someone's inbox?

- Provide training for existing employees.
- Hire new employees.
- Contract consultants.
- Post the project on sites such as Kaggle.
- All of the above.

Don't assume that your employees can pick up new applications and frameworks 15 minutes at a time every other week. They can't.

8. What will be done with the results of your analysis?

In *Analytics: The Agile Way*, I penned a case study about how one company's recruiting head honcho asked me to analyze applicant data in 1999. The company routinely spent millions of dollars recruiting MBAs at Ivy League schools only to see them leave within two years. Rutgers MBAs, for their part, stayed much longer and performed much better.

Despite my findings, the company continued to press on. It refused to stop going to Harvard, Cornell, etc. because of vanity. In his own words, the head of recruiting just "liked" going to these schools, data be damned.



Food for thought: What will an individual, group, department or organization do with keen new insights from your data analytics projects? Will the result be real action? Or will a report just sit in someone's inbox?

9. What types of resistance can you expect?

You might think that people always and willingly embrace the results of data-oriented analysis. And you'd be spectacularly wrong.

Case in point: Major League Baseball (MLB) umpires get close ball and strike calls wrong more often than you'd think. Why wouldn't they want to improve their performance when presented with objective data? It turns out that many don't. In some cases, human nature makes people want to reject data and analytics that contrast with their world views. Years ago, before the subscription model became wildly popular, some Blockbuster executives didn't want to believe that more convenient ways to watch movies existed.

Caveat: Ignore the power of internal resistance at your own peril.

10. What are the costs of inaction?

Sure, this is a high-level query and the answers depend on myriad factors. For instance, a pharma company with years of patent protection will respond differently than a startup with a novel idea and competitors nipping at its heels. Interesting subquestions here include:

- Do the data analytics projects merely confirm what we already know?
- Do the numbers show anything conclusive?
- Could we be capturing false positives and false negatives?

Think about these questions before undertaking data analytics projects

Don't take the queries above as gospel. By and large, though, experience proves that asking these questions frames the problem well and sets the organization up for success - or at least minimizes the chance of a disaster.



Data governance: The case for self-validation

Why you should move beyond
a conventional approach to
data governance

Most organizations understand the importance of data governance in concept. But they may not realize all the multifaceted, positive impacts of applying good governance practices to data across the organization. For example, ensuring that your sales and marketing analytics relies on measurably trustworthy customer data can lead to increased revenue and shorter sales cycles. And having a solid governance program to ensure your enterprise data meets regulatory requirements could help you avoid penalties.

Companies that start data governance programs are motivated by a variety of factors, internal and external. Regardless of the reasons, two common themes underlie most data governance activities: the desire for high-quality customer information, and the need to adhere to requirements for protecting and securing that data.

What's the best way to ensure you have accurate customer data that meets stringent requirements for privacy and security?

For obvious reasons, companies exert significant effort using tools and third-party data sets to enforce the consistency and accuracy of customer data. But there will always be situations in which the managed data set cannot be adequately synchronized and made consistent with "real-world" data. Even

strictly defined and enforced internal data policies can't prevent inaccuracies from creeping into the environment.

Why you should move beyond a conventional approach to data governance

When it comes to customer data, the most accurate sources for validation are the customers themselves! In essence, every customer owns his or her information, and is the most reliable authority for ensuring its quality, consistency and currency. So why not develop policies and methods that empower the actual owners to be accountable for their data?

Doing this means extending the concept of data governance to the customers and defining data policies that engage them to take an active role in overseeing their own data quality. The starting point for this process fits within the data governance framework - define the policies for customer data validation.

A good template for formulating those policies can be adapted from existing regulations regarding data protection. This approach will assure customers that your organization is serious about protecting their data's security and integrity, and it will encourage them to actively participate in that effort.

Examples of customer data engagement policies

- **Data protection** defines the levels of protection the organization will use to protect the customer's data, as well as what responsibilities the organization will assume in the event of a breach. The protection will be enforced in relation to the customer's selected preferences (which presumes that customers have reviewed and approved their profiles).
- **Data access control and security** define the protocols used to control access to customer data and the criteria for authenticating users and authorizing them for particular uses.
- **Data use** describes the ways the organization will use customer data.

- **Customer opt-in** describes the customers' options for setting up the ways the organization can use their data.
- **Customer data review** asserts that customers have the right to review their data profiles and to verify the integrity, consistency and currency of their data. The policy also specifies the time frame in which customers are expected to do this.
- **Customer data update** describes how customers can alert the organization to changes in their data profiles. It allows customers to ensure their data's validity, integrity, consistency and currency.
- **Right-to-use** defines the organization's right to use the data as described in the data use policy (and based on the customer's selected profile options). This policy may also set a time frame associated with the right-to-use based on the elapsed

time since the customer's last date of profile verification.

The goal of such policies is to establish an agreement between the customer and the organization that basically says the organization will protect the customer's data and only use it in ways the customer has authorized - in return for the customer ensuring the data's accuracy and specifying preferences for its use. This model empowers customers to take ownership of their data profile and assume responsibility for its quality.

Clearly articulating each party's responsibilities for data stewardship benefits both the organization and the customer by ensuring that customer data is high-quality and properly maintained. Better yet, recognize that the value goes beyond improved revenues or better compliance. Empowering customers to take control and ownership of their data just might be enough to motivate self-validation.

Clearly articulating each party's responsibilities for data stewardship benefits both the organization and the customer by ensuring that customer data is high-quality and properly maintained.



IoT data with AI reduces downtime, helps truckers keep on trucking

Volvo Trucks and Mack Trucks use sensor data and SAS® AI solutions to minimize unplanned downtime.

Every day, millions of trucks transport fuel, produce, electronics and other essentials across highways. From farms and restaurants to retailers and hospitals, nearly every part of the economy relies on the efficient movement of freight to function.

Unplanned downtime can exact a tremendous toll on any fleet operator and their customers who depend on timely deliveries. Operators can be out thousands of dollars a day when a truck with scheduled hauls unexpectedly breaks down. The impact on smaller regional owners can be even greater, because they're less likely than larger operators to have spare vehicles on hand.

Volvo Trucks and Mack Trucks, both subsidiaries of the Swedish manufacturer AB Volvo, have met this challenge with remote diagnostic and preventative maintenance services based on [Internet of Things](#) (IoT) technologies with analytics and [artificial intelligence](#) (AI) from SAS. With these solutions, Volvo Trucks and Mack Trucks can help their customers maximize a vehicle's time on the road and minimize the costs of service disruptions by servicing connected vehicles more efficiently, accurately and proactively.

Remote diagnostics as a service

Volvo Trucks and Mack Trucks use telematics to deliver unparalleled support services with the purchase of each truck. Volvo Trucks launched Remote Diagnostics with about 4,000 vehicles in 2012, with Mack Trucks offering a similar service called GuardDog Connect in 2014. Today, more than 175,000 trucks are supported with the always-on service that operates 24 hours a day, 365 days a year.

"One of the reasons customers buy Volvo Trucks is for uptime," says Conal Deedy, Director of Connected Vehicle Services for Volvo Trucks North America. "They have a job to do. It's important to keep the truck running to complete their mission - or ensure the least disturbance to the business if something happens on the road."

Volvo Trucks' service monitors data from each truck for fault codes triggered when something is amiss with one of the vehicle's major systems, such as the engine, aftertreatment or transmission. Thousands of sensors on each truck collect streaming IoT data in real time to provide context. This data includes where the event happened and what conditions were present during the fault, like altitude, ambient air temperature, truck gear, RPM level and torque load, to give the information context for diagnosis.

"We process a very large amount of data through the SAS Platform," Deedy says. "We quickly diagnose the fault and its severity with detailed information and a recommended action plan. Our agents in Mack's 24/7 Uptime Center explain the results to the customer and develop a plan for addressing it with



Conal Deedy
Director of Connected
Vehicle Services
Volvo Trucks
North America

Our engineers can now see issues before they impact customer operations and change the truck's design, so we have the best product on the road.

the least disturbance.” Agents may send detailed repair instructions to a local repair facility to help it complete the repair more efficiently and effectively. If the customer performs their own repairs, the detailed information can be sent directly to them. If an issue is software-related, the truck can be updated remotely - without disturbing operations - and quickly returned to its mission.

As the service has expanded, says Deedy, “SAS has not only allowed us to deliver diagnoses accurately and efficiently at scale, it also has allowed us to address more parts and failure modes than we could handle earlier.”

Similarly, Mack Trucks’ GuardDog Connect helps customers evaluate the severity of issues and manage repairs. The telematics-based service currently looks after more than 70,000 connected vehicles. “Our service lets us keep ahead of any issues on the vehicle before the driver has an in-cab experience,” explains David Pardue, Vice President of Connected Vehicle and Uptime Services for Mack Trucks.

GuardDog Connect remotely collects data from the vehicle in the form of fault codes and other parameter data and ranks them based on severity. If the fault requires immediate action, an agent contacts the customer and explains the situation in detail and the recommended action. If the truck requires service, the agent informs the repair facility of the issue, including parts needed, so technicians are ready when the vehicle arrives. Agents track the vehicle at the dealer to make sure it is back in service at the committed time. If the fault is less time-sensitive or does not involve a potential injury, agents inform the company’s decision maker so they plan the repair when it makes the most sense for the operation.

Proactively preventing problems

While these services help customers recover from problems faster, analytics also keeps problems from arising in the first place.

The company helps customers understand how the equipment should perform based on its specification and uses analytics to determine patterns

based on actual equipment usage. “This allows us to give a customer a more dynamic or optimized maintenance plan rather than a traditional calendar plan,” Pardue says.

Analytics is also applied to examine common traits of trucks in the field so improvements can be made in the design of the truck. The analysis identifies emerging issues across an engine type or model year much quicker with real-time streaming data and communicates these findings to the engineering group. “Our engineers can now see issues before they impact customer operations and change the truck’s design, so we have the best product on the road,” Deedy says.

A stronger analytics culture

Using SAS has enabled both Volvo Trucks and Mack Trucks to develop a stronger analytical culture. “Analytics has become part of our culture. We’re using analytics to rethink the way we do business,” Deedy says. “We use SAS



David Pardue
Vice President of
Connected Vehicle
and Uptime Services
Mack Trucks

With SAS, we’re working smarter - we’re seeing things that exist in our information that we couldn’t find before, so we can do things more efficiently and effectively, and drive better results for our customers.

Analytics to take our internal knowledge and leverage it fully to make a difference in our customers' success."

"With SAS, we're working smarter - we're seeing things that exist in our information that we couldn't find before, so we can do things more efficiently and effectively, and drive better results for our customers," Pardue says.

A service that customers can't live without

IoT technologies paired with SAS Analytics have delivered impressive results for Volvo Trucks. "For monitored faults, we've reduced diagnostic time 70 percent and repair time 25 percent when using SAS to process millions of records in real time, and point the agent to what needs to be done," Deedy says. "That's a huge savings for customers who want to have a truck serviced quickly and accurately."

Mack Trucks points to benefits for all stakeholders. Dealers experience a more efficient process. Mack Trucks can see how the vehicle and its key components are performing, leading to greater uptime. And greater uptime has made customers extremely happy. "We have a very high Net Promoter Score," Pardue says. "We're consistently getting very strong feedback like, 'this is a service we can't live without.'"

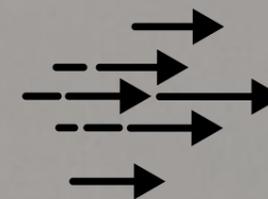
By using learning and automation capabilities that are integral to artificial intelligence, both companies' analytically driven services will keep pace with the changing needs of their customers, keeping them happy for a long time to come.

"Machine learning is an area we're putting a lot of emphasis on right now, using the SAS Platform," Deedy says. "We are uncovering hidden insights in our data and merging that with the truck knowledge from our engineering group. Together we are in a much better situation to understand exactly what the data is telling us. The future is extremely exciting, and the sky's the limit."

Volvo Trucks and Mack Trucks Facts and Figures



175,000
trucks supported with remote diagnostics



Millions
of records processed in real time



25%
reduction in repair time



How to improve data prep for analytics: TDWI shares best practices

Want to improve data prep for analytics? Then you need these best practices from TDWI

Could increasing self-service data preparation reduce the amount of time it takes to gain business insight? According to a survey from TDWI, a high percentage of organizations (81 percent) are banking on it. Nearly as many (76 percent) hope to increase data-driven decision making.

TDWI shares these survey results along with conclusions about how to improve data prep for analytics in a recent best practices report. Responses to the survey came from more than 400 professionals with different roles and levels of tool and platform expertise.

Across the board, organizations want to improve data prep for analytics

More than a third (37 percent) of TDWI's survey participants indicated dissatisfaction with their ability to easily find relevant data and understand how to use it appropriately for business intelligence (BI) and analytics. More specifically, survey participants said the top data-related barrier to improving how data is prepared for user BI and analytics projects was difficulty in accessing and integrating data across system or application data silos. They cited a closely related factor around speed and agility of existing ETL and data integration processes. Other top issues noted were insufficient inbound data quality, and difficulty of integrating data preparation with BI and analytics tools.

How can self-service data preparation help?

Organizations want to harness the power of analytics to make better decisions, and they want to do it as quickly as possible. But analytics can only be as good as the data. Getting good results from your data means doing a

good job of preparing the data for analytics - that is, blending, integrating, cleansing, transforming, governing and defining the metadata of multiple data sources (including raw big data in Hadoop). This process has historically been slow, difficult and tedious.

Most organizations TDWI surveyed want to help business users and analysts do more on their own to serve their BI, data discovery and analytics needs without IT hand-holding. Many new, smarter self-service tools automate processes so users won't need to do as much manual work to find the right data, cleanse it and transform it. Through self-service data preparation, users can become less dependent on IT and data specialists - a boon to everyone, providing IT is willing to give up a certain amount of control. Ideally, business users and IT will work together to ensure their self-service data preparation processes make users more productive without increasing data chaos or duplicating work.

Data prep - the underlying challenges

You don't have to look far to see why "data chaos" is a pain point at many organizations. Spreadsheets are a big culprit - they're the most commonly used tools for data access, queries, reporting, analysis, presentation and sharing. Even at organizations where there's a data warehouse, spreadsheets are where many users continue to do their data preparation.

TDWI found that users tire of waiting for new data to become available in the data warehouse, and they don't want to wait for long IT processes to complete. Instead, they copy and paste data on their own into spreadsheets, and try to cleanse and prepare it there for personal or departmental use.

Spreadsheets are certainly an affordable way to view data, do calculations, create graphs and perform some data analysis. But spreadsheets are fraught with errors - many related to the manual, ad hoc nature of how they're used, the lack of documentation around them and the number of disparate spreadsheets most organizations have.

Best practices – how to improve data prep for analytics

Data variety and velocity are still on the rise, and improving self-service data preparation is a viable way to speed time to insight from all that data. Consider that newer big data sources require integrating semistructured and unstructured data – like customer behavior data, machine or sensor data, log file data, geolocation data and feeds from external sources. By giving business users better self-service data prep techniques, organizations can position themselves to quickly analyze diverse data sources and variables while spotting trends, patterns and correlations.

Some of the data preparation for analytics best practices that TDWI recommends include:

- **Make shortening the time to achieving business insight a data preparation improvement priority.** Put this at the top of your list. Apply new data preparation for analytics technologies and methods that trim delays in getting users from data to insight.
- **Focus on reducing how long data preparation takes to deliver valuable data.** Evaluate your current data preparation procedures to get rid of unnecessary routines. And develop

strategies that rely on automation and standardized processes for incorporating and integrating new data. Then, once it's prepared, register the data in a data catalog so it can be reused by others.

- **Use new technologies and methods to achieve higher levels of repeatability.** Say no to one-off data prep processes that will likely have to be redone each time there's a new requirement or new data. Evaluate how you can apply technologies and adjust processes so that you can reuse scripts, workflows and other elements for different situations. And adopt a collaboration framework that will encourage people to share repeatable methods, scripts and workflows.

Read all the details and recommendations from TDWI by downloading the best practices report [Improving Data Preparation for Business Analytics](#).





Keeping an open mind about open analytics

What if we stopped arguing over which analytics software is best, and decided instead to use them all? Cleveland Clinic and Cox Automotive did.

Today's data scientists come from many different backgrounds, and they bring a wide range of skills to the job. If they have access to a variety of analytics tools, along with a system to govern and deploy models consistently, they have more options for solving complex problems.

Cleveland Clinic and Cox Automotive are two organizations that have benefited from this realization. As a result, their data science programs are thriving - and so are their larger organizations.

"We have employees who are trained in multiple languages and technologies. We want to enable people to access and use languages they're comfortable with but using a common approach," says Chris Donovan, Executive Director of Analytics for Cleveland Clinic.

Cleveland Clinic hopes to grow analytics maturity across the whole health care system, says Donovan. Instead of centralizing analytics skills and capabilities in one team, they're building a broad program across the enterprise. "Having a platform that enables that is critical for analytics to be successful."

[This focus on analytics has helped Cleveland Clinic transform](#), along with the industry, from a focus on palliative care to preventive care. Donovan explains: "How can we move away from just taking care of you when you show up as an individual patient in the ER or the doctor's office, to looking at a population of patients and thinking about how to prevent people from getting sick in the first place?"

Becoming 'code agnostic'

At Cox Automotive, [SAS® Viya®](#) is the glue holding the analytics organization together, says Shawn Hushman, former Vice President of Decision Sciences for Cox Automotive. "It removes the different political debates between the source systems, so we can focus on the modeling itself, the versioning of the models and the delivery of the models. The open platform allows everyone to use their code, leverage open source opportunities, and it opens everyone up to new code bases."

In particular, Hushman praises the ability to integrate disparate code, processes, and information into one hub that provides consistent delivery of information.

Hushman's team includes people around the world who code in Python, R and SAS. "We have people programming in multiple interfaces, and they're using different ways to collaborate on model development. They have different ways they like to publish and show the output and different ways to deploy the models. SAS allows us to stitch it all together."

The solution allows Cox Automotive to be "code agnostic," says Hushman. Instead of debating over a preferred code base, everyone can discuss the best solution for a problem together.

"We don't care about your code preferences," he says. "Let's look at frameworks that can bring real change to the organization, instead of battling over which package I'm going to use for modeling."

With that mindset, Cox Automotive has been able to make the most of its data as it transitions offline businesses like Autotrader and Kelley Blue Book to the online world.

“SAS has the ability to bring together the data scientist community like no other solution can,” says Hushman. “We think of model management as the center of our hub, because that’s where we can be agnostic and make sure everything connects.

“Our responsibility is to deliver results efficiently and make it seamless,” he says. “We want all the goodness that comes with the diversity of different algorithms, and then we bring alignment around our delivery.”

Opening analytics to executives

At Cleveland Clinic, giving more people access to analytics is also a priority. Beyond data scientists and programmers who are adept at writing code and doing advanced analytics, Donovan and other leaders at Cleveland Clinic want to make data easily available to executives and managers with drag and drop capabilities and simple interfaces.

“Our leaders may not know how to build a predictive model but they need to be able to use data to make better decisions,” says Donovan. “Not everyone is a data scientist. But everyone needs to be able to interact with data at their level.”

Donovan says Cleveland Clinic is redefining what an analyst is and working to create a common entry point for all levels of users. “Before, we had data everywhere and multiple tools, but we’re trying to invert that. Instead of taking the data to the people, it’s bringing the people to the data. We believe that if we create a world-class platform, that will draw the people to the platform – which will drive consistency, build communities of practice, and link people across the organization to find standard approaches.”

Opening your business strategy with analytics

Hushman also emphasizes the importance of analytics across the organization at Cox Automotive. “I prefer to view analytics as the heart and soul of our organization and the foundational element of everything we do. Analytics

is improving all our products, driving new products, and growing revenue across all our product suites. Analytics doesn’t drive our business; it is our business.”

Broadening the use of analytics to support all users is more than a technology tactic. It’s a business strategy. “Analytics is not just a capability that supports your core strategy,” says Donovan. “It has to be a core strategy of its own. You need to become an analytically mature organization, and you need to be world-class in that space or people will leapfrog you.”



Chris Donovan
Executive Director of Analytics
Cleveland Clinic

“Analytics is not just a capability that supports your core strategy. It has to be a core strategy of its own.”

Discover how you can orchestrate your data and analytics journey with a perfect balance of choice and control with the SAS Platform: sas.com/platform

Follow us:



To contact your local SAS office, please visit: sas.com/offices

