ANALYTICS FOR SPORT PERFORMANCE AND HEALTH:

A REALITY CHECK

KITMAN LABS

AUTHORS

Dr. Martin Buchheit,
Head of Performance Research
at Kitman Labs

Dr. Derek McHugh,
Head of Data Science
at Kitman Labs

brought to you by the Performance Intelligence Research Initiative (PIRI)

EXECUTIVE SUMMARY

The need for data and evidence to support the decision-making process at every level of elite sport has never been higher. Demand is surging across the industry for objective, shared intelligence about all aspects of players, which can include physical and mental health, on-field performance, tactical and technical progress, development trends, short- and long-term potential, financial value, and more. This holistic view of each player can be used to collaborate across departments to achieve the goals of the organization, an approach we call Performance Intelligence.

Yet the tools available to do this across the industry are highly variable in approach, promise, usability, efficacy, and impact. Ultimately, the right approach can have a significant impact on outcomes and any team not choosing the right path is placing itself at a disadvantage relative to its competition.

This paper offers details on the Kitman Labs approach to analytics, developed after years of research on data, analytical technical techniques, team environments, and organizational behavior. Through this approach, teams can improve processes, strengthen performance, and position themselves for greater success.

THE PROBLEM WITH DATA IN SPORTS

The global sports analytics market is expected to reach $4.6 billion by 2025. But for many of us, the potential of analytics has been vastly overpromised and under delivered.

Let's focus on the data first. Teams vary widely on the data volume spectrum with some teams recording loads of data from multiple sources—others, not so much. Data may include financial and operational data, game and opposition data, demographics, medical, wellness, nutrition, physical, internal and external load. It's typically stored in different formats on different systems, sometimes centrally, sometimes locally on individual laptops.

Deep silos of data have been one big inhibitor to success with analytics. But success it isn't just about aggregating data and verifying its accuracy; it's about building and continuously iterating models so that organizations can draw the right insights from the right data to know what matters most and why.

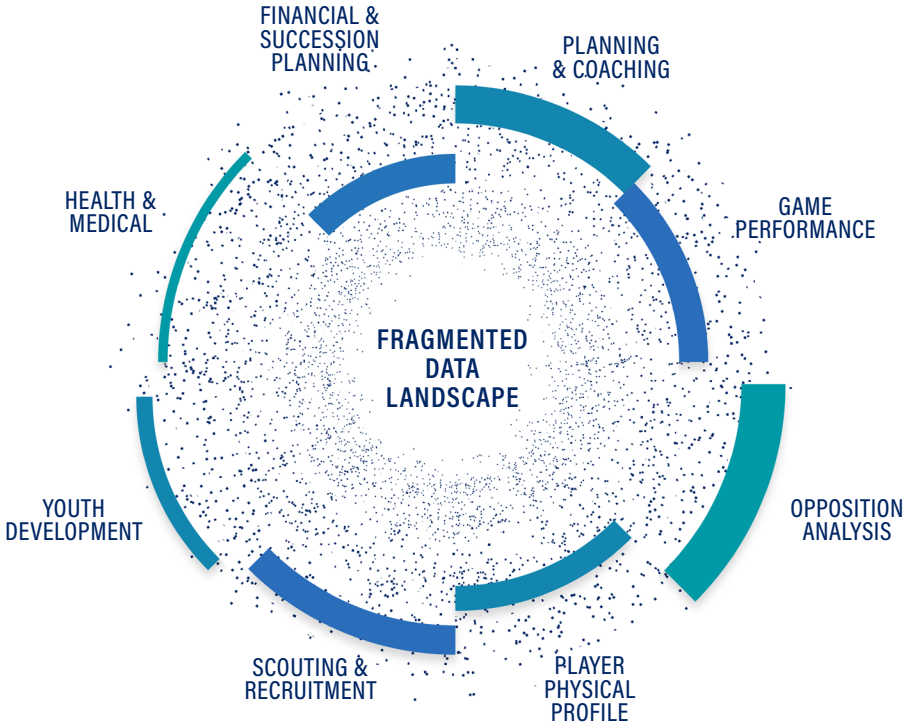
THE DATA CHALLENGE

"The key challenge for a club is to put the increasing amount of data that they have at their disposal in **the right context and to draw the right conclusions from the data."**

—

SciSports Chief Analytics Officer Jan Van Haaren

These factors have meant analytics have historically been difficult and time-consuming and have often not yielded reliable, high quality and actionable outputs. This has limited teams from harnessing their data to improve.



Value Tucked Away In Data Silos

A focus on collecting more and more data has led to deep data silos scattered across the organization, making it nearly impossible to capitalize on the value of data to learn and improve.

WHY APPROACH MATTERS

Analytics is a tool. It is a way to process the vast amount of data in a smart and efficient manner to provide objective inputs into the decision-making process. However, it will never be perfect, and we must be aware of that.

> "All models are wrong; **some, though,** are better than others"
>
> – **George Box,** Statistician

Given our imperfect world, choosing the right approach to performance analytics is critical. Data is fast becoming the common denominator for teams with a relentlessly high performing culture and the approach they choose can help create a distinct competitive advantage.

KITMAN LABS APPROACH: FIVE CORE PRINCIPLES

01 - OUTCOME-FOCUSED
02 - SCIENTIFIC
03 - UNBIASED
04 - MULTIVARIATE
05 - TRANSPARENCY

At Kitman Labs, our approach to analytics is based on the foundation built by our Performance Experts team and complemented by our experience as former players and practitioners. It is very much anchored in years of research and our experiences and has evolved over many years working with more than 700 sports teams helping solve their most pressing questions.

We believe sports most talented and passionate practitioners should be able to enhance their expertise and experience with objective insights to the greatest degree possible. We believe that an organization's aspirations to attain the highest standards possible and leverage the same quality of tooling that other industries now rely on should be a reality. To this end, we have deeply researched analytics over the last number of years to scientifically assess what works and what doesn't work. This research has yielded the five core principles we use in all of our analytics solutions to help organizations achieve the outcomes they want.

Through our Performance Intelligence Process, we work to understand the critical questions teams need to answer to perform better. We bring together our Performance Experts—data science, sports science, and performance science—along with the organization's own support staff, to tackle these questions holistically. Our approach employs five core principles, which we'll discuss briefly.

01 OUTCOMED-FOCUSED

Analytics isn't about math or algorithms, ***it's about results.***

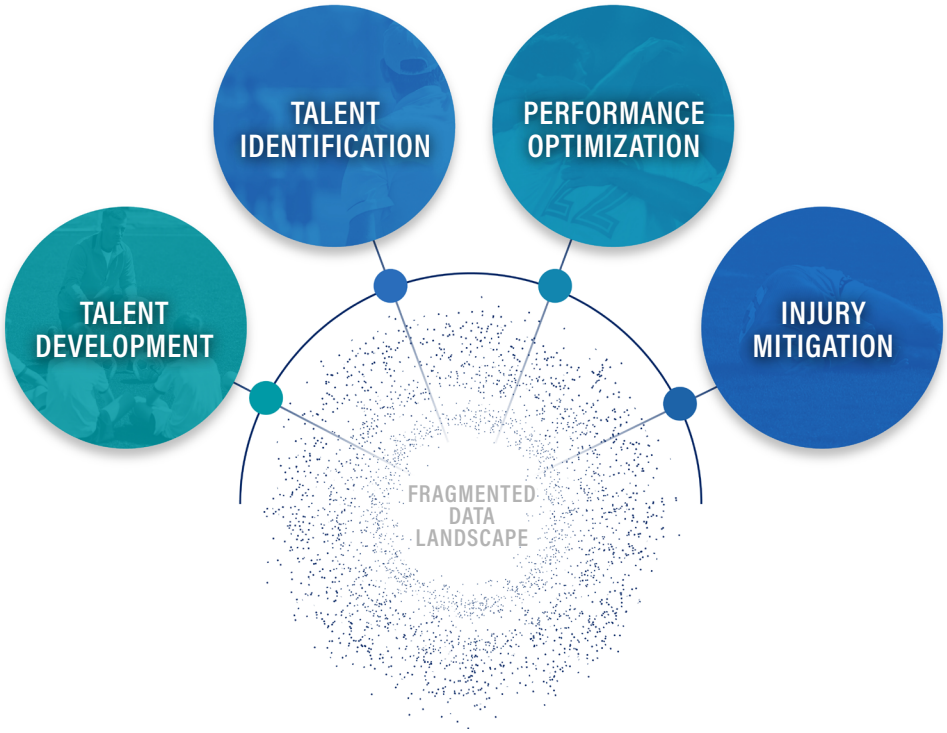
We begin by focusing on the *outcomes* we aim to achieve. In other words, where do we want to be? Thinking in outcomes can help teams focus and organize around a common mission and the actions required to get there.

Our analytics always start with these outcomes, which define the questions, and is ultimately measured by what degree these outcomes are achieved. Like our approach to data, we go deep with our teams to understand the relevant questions and the outcomes that will make a difference to their practice, and to their success.

When a team interacts with our analytics, they want to understand how it can make them better. Our focus is on offering *practical* and *actionable* solutions that result in better informed decisions so teams can take the appropriate actions given their unique situation.

Analytics isn't about math or algorithms, it's about *results*. From how we work internally, to how we work with teams and data, to how we present back the data, our sole focus is on providing intelligence that promotes conversation and collaboration, and supports purposeful action and meaningful impact.

Analytics is All About Outcomes



Analytics is a means to an end. We believe that everything about the analytics process, data science, and results must support an organization in learning, collaborating and achieving their desired business and performance outcomes.

Science is all about observation and experimentation. At Kitman Labs, we are performance scientists. We seek to be challenged and are on a continuous quest to learn and improve through experimentation to help advance the field of performance. It's about being curious and constantly looking for new approaches and new methodologies from inside and outside the sports analytics world that we can incorporate into our practice.

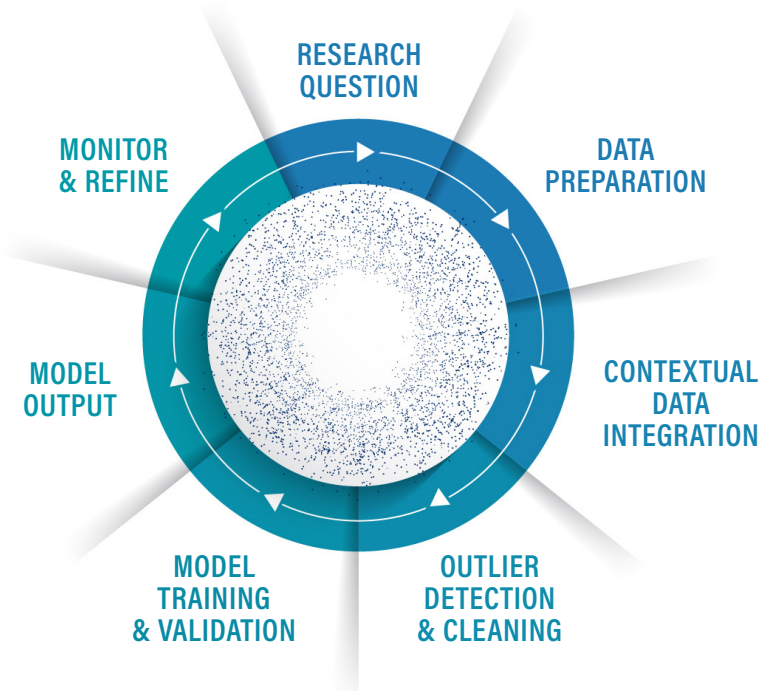
For that reason, when attempting to answer questions—e.g., *"why are injuries up?"* or *"why did he perform poorly?"* —we apply a scientific approach. Science is about experimentation and testing hypotheses to gain insights.

The process begins by understanding the overall team environment and context. We then probe deeper to come up with questions that will lead to meaningful action. Next, we put in place a model to collect, aggregate, and verify data accuracy. Then we build, validate, and iterate our models. Let's go through each of these steps briefly.

Data Preparation

A model is only as reliable as its data. We work collaboratively with our partners to identify data relevant to help answer their unique question. Sometimes, this information may already exist within our system. Other times, we must collect it. Regardless of its origin, we assess the health of the data for quality, consistency, and volume, using automatic product checks to make sure we understand outliers and know what's missing and why. Throughout this process, we stay closely aligned with the club to understand their metrics and frequency of capture. Once we're confident of its integrity, we categorize the data to facilitate feature engineering later.

There is More to Analytics Than Math



In order to achieve the desired outcomes, we believe analytics must employ a structured, scientific process that deeply investigates the question and the means to answer that question, integrates performance science and data science, and assesses the quality of results.

Contextual Data Integration

Equipped with verified data, our experts work alongside partners to explore areas of focus using data visualization, e.g., studying the team's injury profile and key metrics such as athlete injury history, age, days on feet, congestion levels, travel, and more. This process is also informed by our internal Applied Sports Science and Performance Science teams.

Using basic device data and domain knowledge, we create metrics directly relevant to the challenge at hand with the aim of strengthening the model's performance, more commonly known as feature engineering. In particular, we look for metrics capturing patterns (recurrency of events) and change over time—both short- and long-term. We also determine the universe of athletes to include in the analysis, e.g., just first team players, academy athletes up to age 18, etc.

Outlier Detection and Cleaning

Every club's situation is unique, and we are careful to develop metrics that consider each organization's specific context. We eliminate corrupted data from defective devices as determined by their data's validity and reliability. The result is a dataset ready for modeling.

Model Training & Validation

Model choice depends on the nature of the problem. They can range from a generalized linear model like logistic regression to machine learning approaches—such as gradient boosting or random forests—to more complex approaches, such as artificial neural networks for time-series data. Typically, we require a reasonable degree of interpretability from results, which drives model choice as much as model performance.

Our tuning of the model depends on the defined outcome a club seeks to achieve. Common factors for an injury challenge, for example, could include recall, precision, Brier score, cost, or other suitable scoring metrics.

02 SCIENTIFIC

Once constructed, we look to validate the model by performing out-of-sample testing. This is an area where we set a high bar for ourselves because we want to prevent data leakage which would overestimate our results and conceal real model performance.

One approach, in simple terms, is to omit some athletes' data when training the model, then comparing the model's outputs with what we know actually happened to the athletes that were omitted.

Finally, we monitor each model's performance over time to assess its ongoing relevance to the team and to ensure the model adapts to any changes in team context.

We work with our applied sports science and performance science teams ***to convert investigation results into clear actions that can directly impact practices in the club.***

Model Output: Actionable Results

The availability of objective data allows practitioners to use analytics as a tool alongside their invaluable experiences and deep knowledge of their own organizations to enrich and improve outcomes.

We work with our applied sports science and performance science teams to convert investigation results into clear actions that can directly impact practices in the club.

This is where world-class sports Performance Experts come into play. These deeply experienced practitioners help teams leverage tailored intelligence platforms to unite their people around their data, strengthen instinct, and unlock greater performance.

03 UNBIASED

*The danger of bias is that **it can lead to incorrect conclusions.***

As data scientists, it's important to be impartial towards any specific data, algorithms, or methodology. At Kitman Labs, we are open to whatever we find, whether it confirms what we think, or is contrary to intuition or de facto truths.

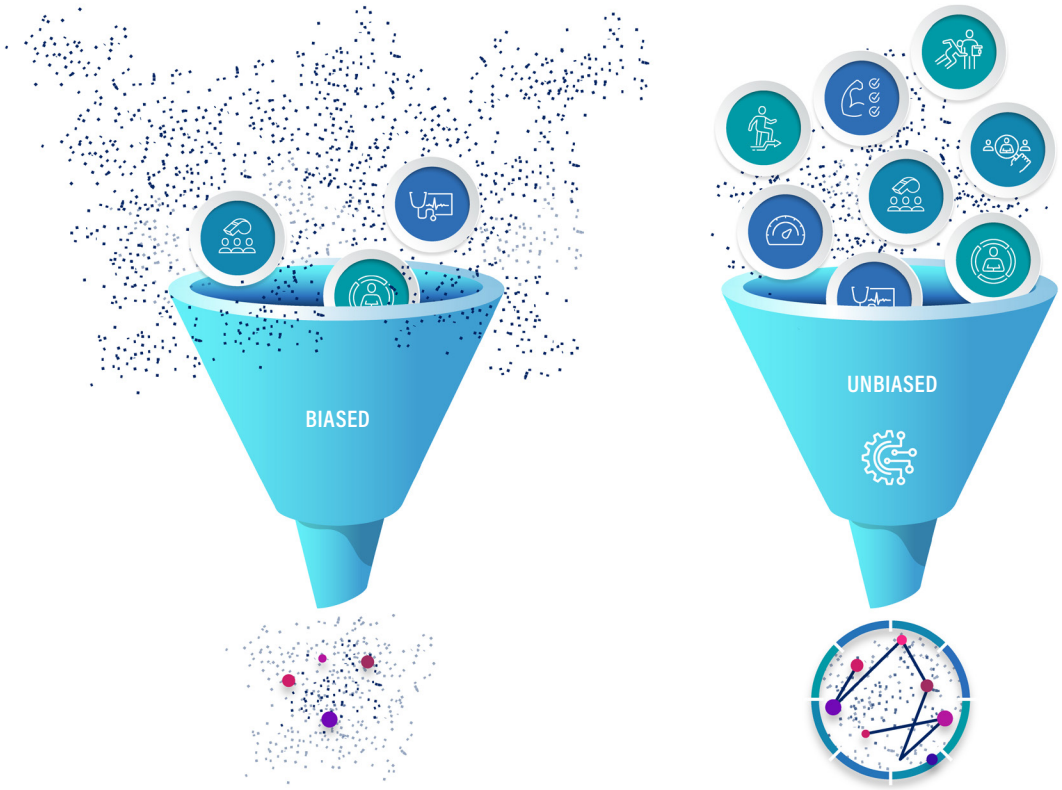
Our approach to analytics is to stand back from the question at hand and see where we can provide insight into potential solutions using all the available data. We don't come with preconceptions on what form the solution may take, or what metrics or techniques may help lead us to a solution.

In fact, it is less about finding the solution to the question, and more about how we can harness available data to give some direction towards reaching a practical solution.

We've learned a great deal in working with a broad range of diverse teams through close collaboration with our internal sports and performance science teams and the support teams within the clubs. This provides our analytics team with the contextual information they need to understand the team's data and questions much better. At the same time we keep an arm's length from the teams we work with to avoid specific biases about the data or potential solutions creeping in.

Eliminating Biases To Find The Best Answer

Analytics should deliver objective intelligence that cuts through bias. We believe the analytics process should consider all data and all approaches (right graphic) to find the best answer, rather than employing preconceived notions (left graphic) of the data that matters or the data science solution that is best, which may result in a poor answer.



04 MULTIVARIATE

*With any athlete, there is an array of factors that interact **simultaneously.***

Sports is complex. Humans are complex. With any athlete, there is a huge array of factors that interact simultaneously—physical, psychological, technical, tactical, and contextual—that we can neither fully know nor measure. Therefore, analytics need to account for the multiple factors that affect outcomes and any model we build to understand these problems must be multivariate.

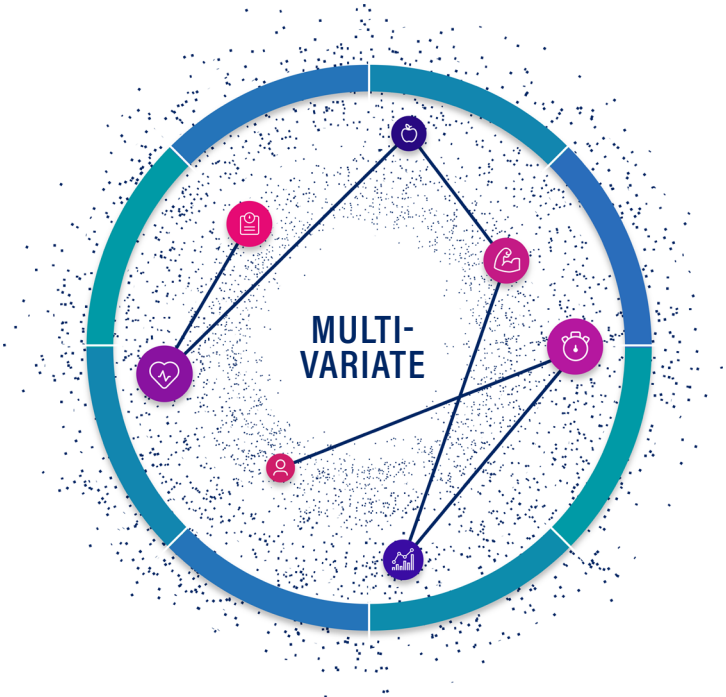
The sheer volume of metrics involved for any one athlete makes traditional statistical modelling, e.g., regression, difficult. We employ machine learning methods which are designed to handle the scale of the datasets involved. The sophistication of these algorithms allows us to probe the reasons behind the behaviors we observe, while still allowing a reasonable degree of interpretability.

For example, looking at the relationship between an individual metric and the desired outcome is useful to gain insight into whether that metric is relevant to achieving the outcome. Doing this with multiple metrics, individually, can inform what features to include in a model, but it becomes impractical when the number of metrics involved becomes very large.

Our goal is to understand the underlying data-generating process as best we can using the available data. Various interpretability methods allow us to assess which of the many factors and their combinations are most important to that process.

Analytics need to account for the multiple factors that affect outcomes ***and any model we build to understand these problems must be multivariate.***

A Multifaceted Environment, A Multivariate Solution



Humans are complex, the game is complex and the environment is complex. We believe there is never a silver bullet, and that organizations must understand the multiple and combinations of factors that impact desired outcomes.

05 TRANSPARENCY

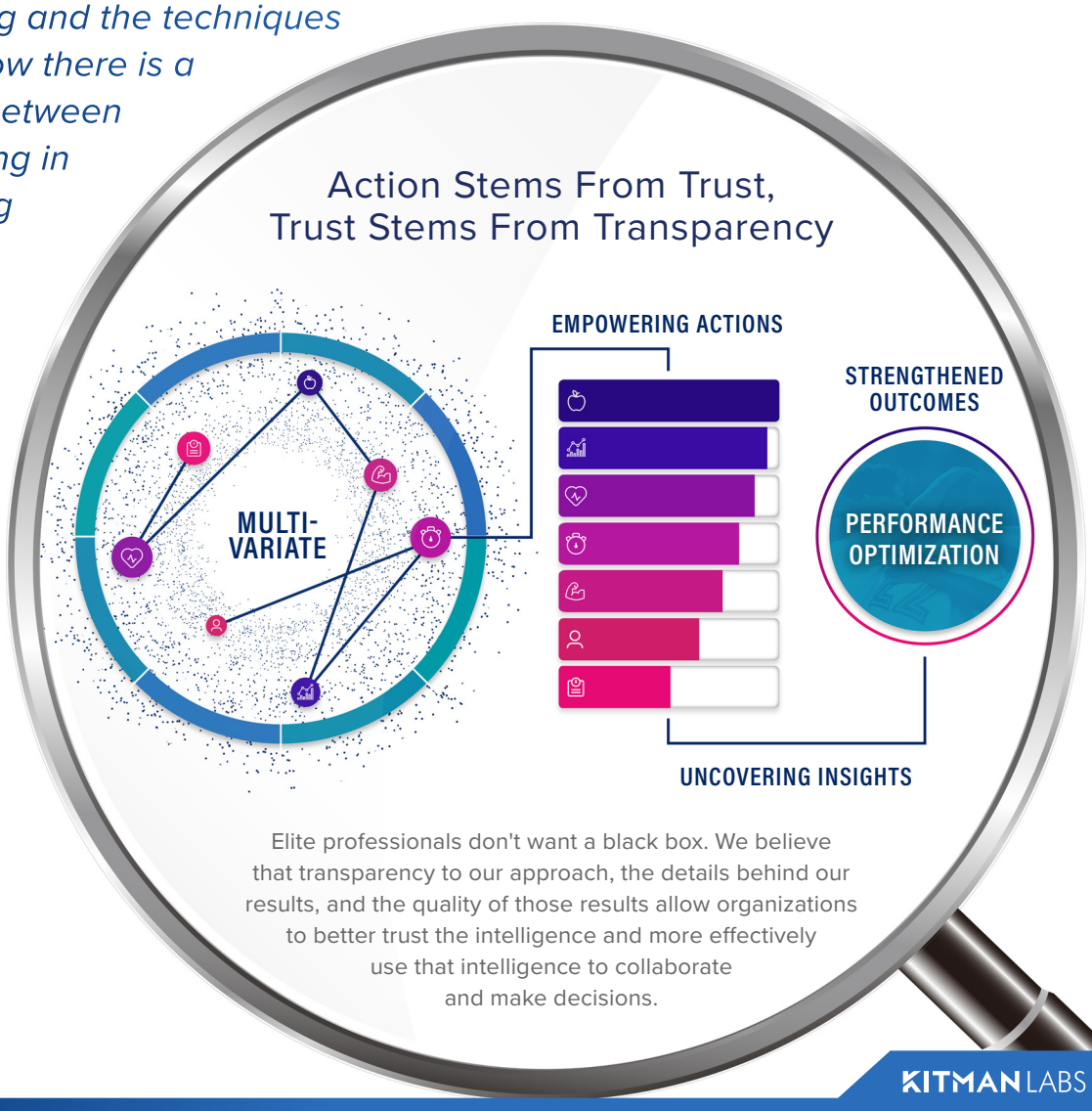
Trust is an essential element.

Transparency is at the heart of our approach. Transparency is being clear about the data being used, the methods used to model the data and the performance of the model, as well as clear, understandable outputs that lead to definitive action.

We do not create black box algorithms or try to hide our processes through claims of intellectual property. Our process is collaborative and fully transparent. For example, when we carry out a specific piece of analysis for a team, we are explicit in the steps we take with their data, the assumptions that have been made, and any limitations that exist. Most importantly, we share any uncertainties we may have regarding the results so that organizations are fully informed and aware of any information that may be lacking when making decisions.

Equally important, trust is an essential element. We ensure the data and insights are usable, and we create transparent mechanisms to allow teams to understand and trust the information being delivered through our tools so they can turn it into action.

> *At Kitman Labs, we always speak openly about what we are doing and the techniques we use as we know there is a huge difference between knowing something in theory and putting it into practice.*



05 TRANSPARENCY

Our modelling techniques are not the “secret sauce” of what we do—the reason our analytics have multiple steps, values, and principles, is because they are all important in determining meaningful insights. We will tell our partners when we don’t have solutions or when we think we are not able to deliver what is required. It’s all about positioning teams for incredible success.

Final Thought: Analytics Limitations

Analytics aren’t perfect and we would be remiss if we didn’t mention a few of its limitations. Data quality will always be an issue, data context is critical, and data collection and data consistency are difficult, especially when dealing with complex human problems.

And oftentimes, all factors related to an event (i.e., type and quantity of data) have not yet been identified so our data scientists engage with both the internal Kitman Labs sports and performance science teams, and the organizations we’re working with to address these issues up front.

We’re only as good as the data we get, and we will never have access to all the data relevant to a problem. That said, we work proactively with teams to get *more* and *better* data in advance of analytics to mitigate the biggest limitations of data: quality and quantity.

But how much data is enough? A key consideration when building a model is that it should be trained on data that has “seen” examples of similar situations enough times that it can provide an answer with a reasonable degree of certainty.

Reducing uncertainty in the decision-making process is our key aim. We bring objectivity into the process by consolidating all the data in one place, harnessing the latest statistical and machine learning techniques to interrogate it, and narrowing the focus of the investigation. Once that is done, we can prescribe data-driven insights that matter most to practitioners facing problems around athlete injury risk and performance improvement.

Ultimately, there is no magic. Collectively, we all must continuously ask questions to reduce uncertainties and push the limits on what we can expect from data.

CONCLUSION

KEY TAKEAWAYS

01 After years of collecting data at a departmental level, the industry is now focusing on consolidating data and turning it into shared intelligence.

02 Organizations that can capitalize on their data to learn and improve will gain competitive advantage.

03 Certain principles are necessary to deliver on the promise of analytics and create the positive impacts organizations need to consistently succeed.

04 Analytics is not a math exercise. It's not a destination. There is no perfect score or silver bullet. It's a journey. It's about deep insights, learning, and improving.

05 As sports and data evolve and change, analytics will need to evolve and change as well. New research, new methods, new approaches will always be needed.

As analytics in elite sport matures, advocates must effectively communicate its true role: supporting—not replacing—practitioners' valuable experiences, expertise and authority. Analytics can be extremely helpful in unearthing new insights and strengthening instinct to ensure practitioners are making the best possible decision, one that is based on both data and human intuition/experiences. That's a conversation Kitman Labs is eager to help facilitate. And the best way to do that is via real-world applications, not theoretical beliefs.

We want to engage industry and invite you to share your experiences, give us feedback, and challenge us by reaching out to **science@kitmanlabs.com**.

ABOUT

Performance Intelligence Research Initiative (PIRI)

An arm of Kitman Labs, the Performance Intelligence Research Initiative (PIRI) is a research, education, and community engagement program designed to produce new evidence-based insights that will advance the fields of sports performance science, coaching, talent development and identification, and leadership. It also aims to accelerate the adoption of gold standard practices discovered through the research via interactive learning vehicles and engaging forums.

By working collaboratively with its partners and other researchers, PIRI can uncover the industry challenges that are most important and apply its performance science, data science, and organizational science research expertise to solve them.
Visit https://www.kitmanlabs.com/performance-intelligence-research-initiative/
or open up a dialog with our Performance Scientists at science@kitmanlabs.com.

Kitman Labs

Kitman Labs is the intelligence platform provider for elite sports teams. Founded in 2012 with the vision to fundamentally change how the sports industry uses data to achieve on- and off-field success, Kitman Labs has consistently set the standard for innovation, data integration, design, and analytics.

With over 50 Performance Experts with backgrounds in coaching, medical, sports science, and data science, Kitman Labs provides unparalleled expertise and support. Over 700 elite teams across the NFL, NHL, MLB, NCAA, English Premier League, La Liga, Serie A Bundesliga, Pro14, Premiership Rugby, and Japan's Top League use the company's platform to improve talent economics, enhance performance, produce more talent, and increase player health and longevity. The company has a presence in the United States, Ireland, United Kingdom, Central Europe, Australia, and Japan.
For more information visit: **https://www.kitmanlabs.com/**