Redefine Your Analytics Journey With Self-Service Data Discovery and Interactive Predictive Analytics

§sas
THE POWER TO KNOW®

Contents

Self-service data discovery and interactive predictive analytics

What's so important about the convergence of self-service data exploration with interactive predictive analytics? Why does this matter to your organization?

Think of all the self-service things you use in a day. Gas pumps. ATMs. Online apps for stock trading, banking and shopping. The idea behind them is added convenience for consumers and reduced costs for providers. People choose what they want, when they want – without involving others in their minute-to-minute decisions.

The same goes for self-service data exploration and easy-to-use analytics. Increasingly, people want to explore data and search for answers – when they want, where they want and how they want to – without asking IT for data or reports. We're seeing it even in everyday data applications like the Fitbit or Apple Watch™. People are hungry for information they can find, see and use on their own.

So whether it's combining caloric intake with exercise output for that magic fitness number, digging into why weekly sales are trending downwards from the office or drilling deep into a report from a mobile device at an airport, giving people the opportunity to explore data on their own empowers them to make better decisions (e.g., skip the doughnut at the airport).

But it's not just about learning what happened and why. What about the future? Adding interactive predictive analytics to your organizational mix enables decision makers and data consumers to go beyond the past to discovering (for themselves) insights about the future.

Combining interactive predictive analytics with self-service data discovery also enables statisticians and data scientists to explore data relationships before they dive into model building and refinement. They can quickly and iteratively try different data and techniques to find the model that produces the best answer. When satisfied with the model, they can easily collaborate with business analysts to explore the modeling results from the same environment. Once identified, the champion model can be put into production by embedding it into operational applications and scoring new data to produce faster insights.

What does your organization really need?

Today, organizations continue to rank analytics as their top technology priority. Why? Because there's an increasing realization that analytics, especially predictive analytics, can drive growth and innovation, which leads to competitive differentiation.

But despite technology advances, many are still struggling to make business intelligence and analytics a pervasive part of their culture, especially for workers without analytical skill sets.

So what is needed to support the increasing need for faster, more precise analytics and more interactive data exploration and true business intelligence? What is needed to bring analytical power to the masses?

According to industry analysts Gartner, the No. 1 must-have for delivering self-service data discovery and analytics is ease of use. Gartner's survey found that 41 percent of organizations cite this as one of the top three features to look for when choosing an advanced analytics platform. This is especially true as the user base expands beyond a limited number of statisticians and data scientists to a broader set of quantitative business analysts and specialists – for example, marketing or financial analysts. Speed of model development was also important for 31 percent of the respondents.[1]

It's becoming clear that three factors are important for organizations that want to provide real value from data to everyone who needs it:

Interactivity and ease of use for users of all skill levels.

Powerful predictive analytics and consumable results.

An architecture that can handle big data and the need for speed.

---

1   Gartner. "Survey Analysis: Customers Rate Their Advanced Analytics Platforms (G00270213)," Oct. 28, 2014.

The perfect convergence: SAS® Visual Analytics and SAS® Visual Statistics

There's good news from SAS. The key features that provide the Holy Grail of self-service data discovery and predictive analytics are now available. SAS Visual Analytics provides a highly interactive user experience that combines advanced data visualization, an easy-to-use interface and fast in-memory technology. Even nontechnical users can explore data on their own, create visualizations and dynamically share results. SAS Visual Statistics is an integrated add-on to SAS Visual Analytics that provides interactive modeling for increased predictive analytics power.

The modeling capabilities of SAS Visual Statistics are included within the common user interface of SAS Visual Analytics. This means users can interactively explore data and create models from the same, easy-to-use environment. Combining these two products gives you a platform that provides:

- Interactivity and ease of use.
- Self-service data exploration and visualization.
- Powerful descriptive and predictive analytics.
- Dynamic group-by processing.
- An in-memory architecture for faster results.

Let's take a look at each of these capabilities.

Interactivity and ease of use

It only makes sense that ease of use is one of the top criteria for choosing an advanced analytics platform. The platform needs to be interactive and approachable so nontechnical users can explore data and find answers they need, on their own.

SAS Visual Analytics and SAS Visual Statistics share the same web-based interface, making it easy to visually explore data on the fly, spot patterns, run predictive analyses and quickly deliver insights. You also get point-and-click access to powerful SAS statistical modeling and machine-learning techniques. These techniques can be used to build and refine models on data of any size to predict outcomes at dramatic speeds that result in better, more targeted actions.

With integrated model building and visual data discovery, analysts and data scientists can maintain an uninterrupted workflow, cycling quickly between hypotheses and verification. These capabilities boost model development, productivity and accuracy in near-real time.

Self-service data exploration and visualization

SAS Visual Analytics enables users to discover, explore and analyze critical data without assistance, and then distribute dynamic charts and graphs to others. You can:

- Use autocharting capabilities and compelling data visualization techniques to quickly explore relevant relationships in data.
- Identify top variables and outliers affecting outcomes you are trying to predict.
- Interactively filter data, format variables, add ad hoc calculated columns and create dynamic hierarchies without the need for predefined data structures.

Powerful descriptive and predictive analytics

SAS Visual Statistics adds sophisticated analytical capabilities and works in tandem with SAS Visual Analytics. With SAS Visual Statistics, users can iteratively create descriptive and predictive models. They can refine models in seconds to find the one that produces the best results – in near-real time. It's really that easy. SAS Visual Analytics can be used **before** model development to explore data and **after** model development to explore the results. This helps reduce latencies during the iterative model development process.

Descriptive modeling

Clustering techniques segment your data into groups that share similar features so you can identify distinct groups based on multiple characteristics (see Figure 1). For example, market segmentation clusters customers with similar buying habits to determine which promotion will work best. SAS Visual Statistics lets you visually explore and evaluate segments using k-means clustering, detailed summary statistics, parallel coordinates and scatter plots. Scatter plots with cluster profiles are overlaid for small data sets. Heat maps with cluster profiles are overlaid for large data sets.

Predictive modeling

SAS Visual Statistics makes it easy to create analytical models using predictive techniques such as linear regression, generalized linear modeling, decision trees and logistic regression.

- **Linear regression** allows you to determine if a relationship exists between a response variable and a set of candidate predictor variable(s), and describes the nature and accuracy of this relationship (if it exists). Using linear regression, you can estimate numerical predictions such as customer spending, policy premiums, credit limits, etc.
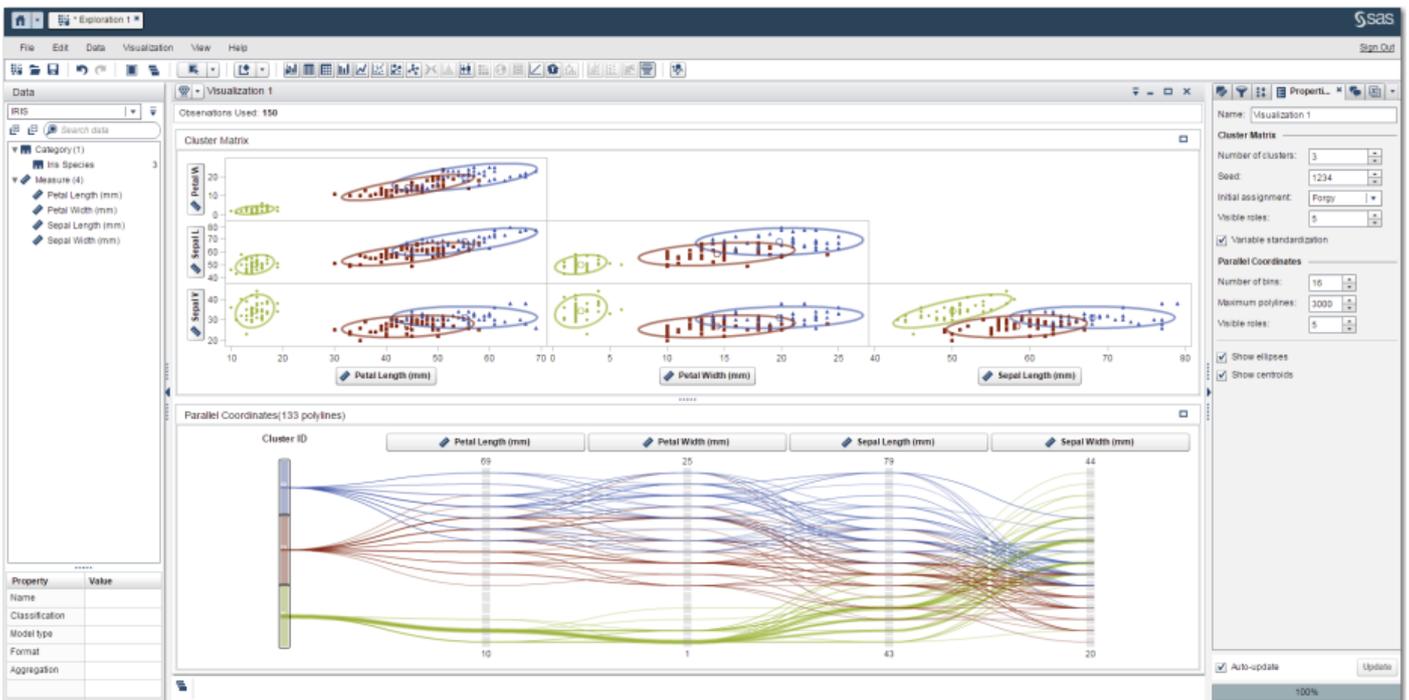
Figure 1: SAS Visual Statistics lets you easily explore and evaluate segments for further analysis using descriptive techniques such as k-means clustering.

- A **generalized linear model** is an extension of a traditional linear model and is useful when data is not normally distributed, variance of data is not constant for all observations or if the mean of the data is restricted to the range of values. Generalized linear models incorporate all interactions between the predictor variables being evaluated to determine effects on the response variable.

- **Decision trees** create a hierarchical partitioning of the input data based on a series of rules applied to each observation. This is a supervised machine-learning technique that uses a branching or tree-like approach to quickly model specific target variables or an outcome (e.g., to buy or not buy, fraudulent activity or not fraudulent activity, etc.). Decision trees help you understand how to target the right customers by analyzing the path of their decisions. (See Figure 2.)

- **Logistic regression** attempts to predict the probability of an event's occurrence (a binary variable, like Yes/No or 1/0) using the nonlinear relationship between response variable and the combined effects of the predictor variables. It can be used to predict things such as machine failures, which patients are high-risk or low-risk, customer churn (Y or N), a fraudulent transaction (Y or N), etc. (See Figure 3.)

Dynamic group-by processing

With SAS Visual Statistics, many users can concurrently build models and process results for each group or segment without having to sort or index data each time. This means granular models can be generated for each segment or group on the fly without additional processing overhead.

SAS® Visual Analytics can be used before model development to explore data and after model development to explore the results. This helps reduce latencies during the iterative model development process.
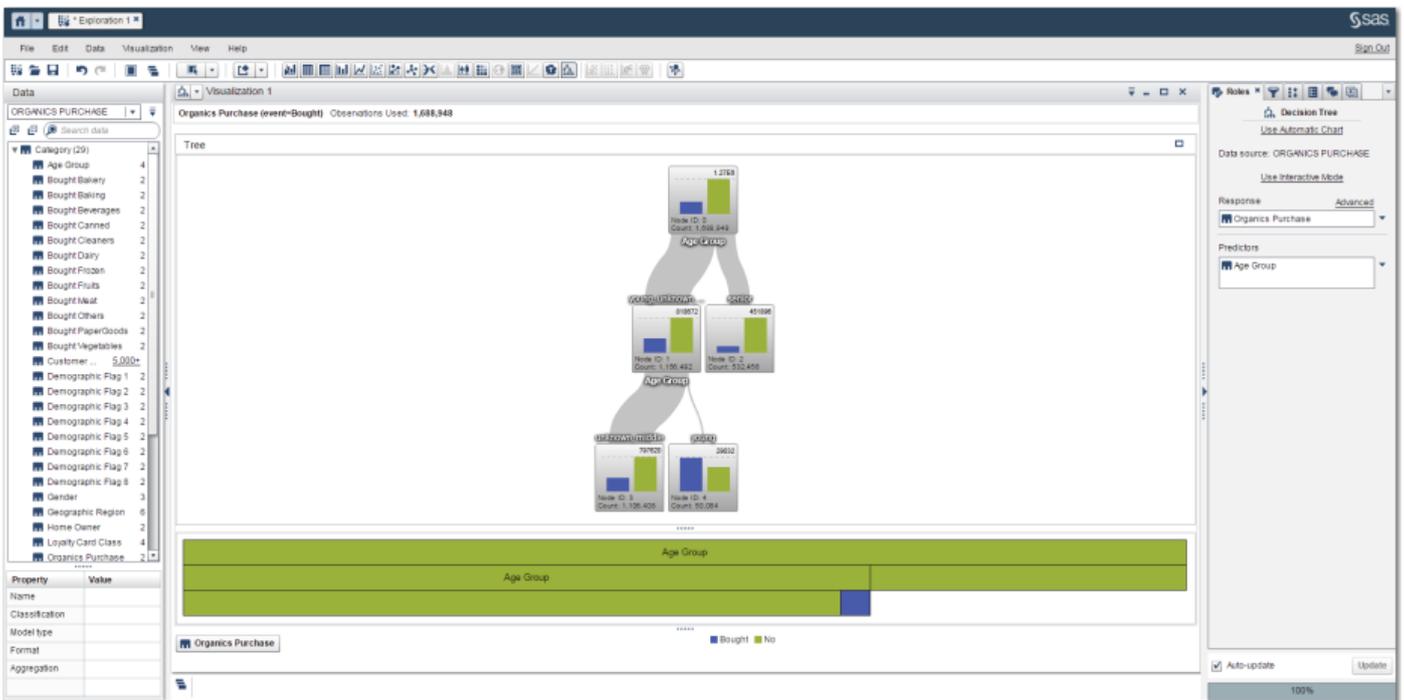
Figure 2: Interactively build decision trees to partition your data and understand actions.
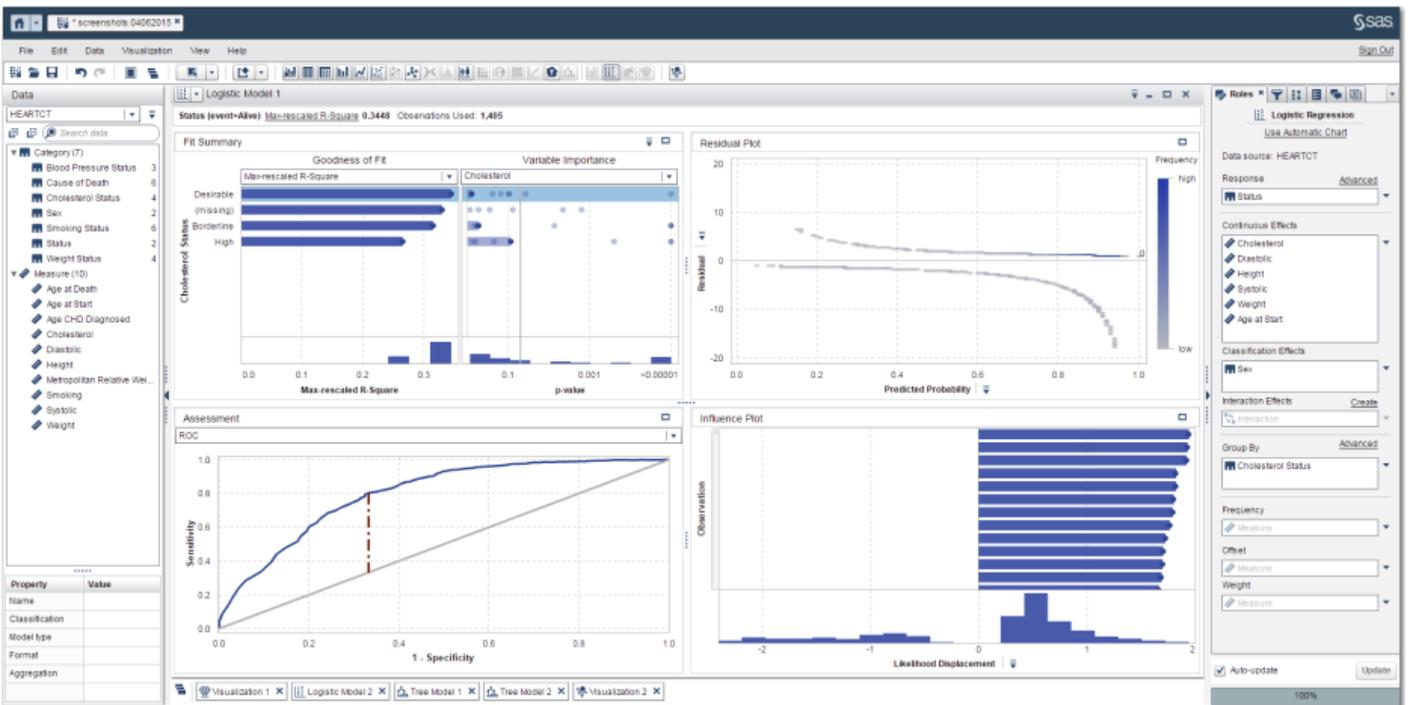

Figure 3: Predict binary outcomes (often yes or no decisions) using logistic regression.

In-memory architecture: From a static, batch-oriented process to interactive, rapid predictive analysis

Traditional batch-oriented architectures are not designed for fast, complex analytical processing and on-the-fly changes to predictive models. Nor are they intended for the multiple passes through data that is needed for analytical jobs. This creates several issues.

- Statisticians, data scientists and business analysts can't easily experiment and refine their models with new information to reach a high level of confidence that they've produced the best model.
- They cannot quickly and easily build numerous models by segments or groups.
- And they can't achieve a fast response each time they adjust model properties. Building and testing models to find the best performer takes much longer.

All of these factors are great inhibitors in the world of big data. If it takes too long to uncover insights, you're not going to get the answers you need when you need them.

You need an architecture that will not only scale as your data and processing needs grow, but one that also supports the ability to quickly and interactively adjust and experiment with variables in near-real time. It's difficult to create good models – much less on the first try. Often the best strategy is to produce fairly good results quickly, and then adeptly adjust what's still not ideal.

Fortunately, processing power is more economical than ever. And with SAS, your software combines in-memory processing with an iterative, interactive model development environment so you can build and refine models and test what-if scenarios much faster and easier than ever before.

Based on in-memory technologies, SAS visual solutions enable users to perform fast, ad hoc data exploration and visually identify patterns and relationships in large amounts of data.

Large data and analytic workloads are distributed across multiple processors for exceptionally fast speed. Data is persisted in memory for multiple sessions to produce near-instantaneous results. There's no need to write data to disk or shuffle data frequently across nodes. All data is loaded into memory once. It does not have to be reloaded each time a new task is performed or another authorized user is given access to data.

This means that **results** from changes made to models (e.g., adding new variables, eliminating outliers or changing the role of variables) are instantly visible. And multiple users can concurrently analyze complex data on Hadoop clusters, relational database systems or SAS servers.

A powerful duo for one-step exploratory modeling

SAS Visual Analytics and SAS Visual Statistics are a powerful duo, combining interactive data exploration and discovery with the ability to easily and quickly build predictive analytical models. Here's how it works.

To create models in SAS Visual Statistics, you start in the SAS Visual Analytics Explorer. (The predictive modeling techniques icons will be on your toolbar.) You can easily investigate the different variables influencing your response variable, understand relationships among variables, remove outliers or handle data manipulations here – **before** you start building models. **After** you create a baseline model, you can continue to refine the baseline model, add more variables, explore your model predictions or segments, evaluate the model's fitness and perform model comparisons, all with very fast results.

The process

1. Before building models, you'll first need to identify what predictors are needed. SAS Visual Analytics helps you visually identify relationships between variables, discover unusual interactions and identify outliers or variables that could be useful in the model-building process. Graphics such as heat maps, box plots, animated bubble charts, histograms and network diagrams show trends and patterns in large volumes of data.
2. SAS Visual Statistics can then be used to interactively create, evaluate and refine predictive models (no coding needed). The same environment can be used simultaneously by multiple users who can all get instant results.

3. After models have been created, they can be easily compared and assessed using a variety of statistical techniques (see Figure 4). Lift charts, ROC charts and misclassification tables help identify the model that performs best. You can then go back, remove outliers and make other modifications to your data or models. It's a very interactive and iterative process.

4. At this point, data scientists can pass model predictions or outputs and/or segments to business analysts for further explorations, analysis and profiling in SAS Visual Analytics (see Figure 5).

5. After you have determined which model performs best using the built-in model comparison task, your champion model can be easily applied against new data. And this provides the precise insights needed for better decision making and improved results.

Solution recap

SAS® Visual Analytics – what does it do?

- Interactive data exploration and visualization.
- In-memory processing for self-service analytics.
- Delivers results via the web, Microsoft Office applications, mobile reporting and dashboards.

SAS® Visual Statistics – what does it do?

- Interactive and exploratory predictive modeling.
- In-memory processing for fast model creation, refinement and comparison.
- Derives data segmentation and generates models by segments or groups.

Expand the possibilities across your organization

Marketing analysts

- Use SAS Visual Analytics to gain insights into revenue over time and determine which factors are related to customer churn.
- Use SAS Visual Statistics to identify customers who will potentially churn in the next six months and find out why.

Financial risk analysts

- Use SAS Visual Analytics to understand and describe customers who default on mortgage payments.
- Use SAS Visual Statistics to predict the risk of which customers will default and estimate more appropriate credit limits for each customer.

Engineers

- Use SAS Visual Analytics to explore equipment efficiency and investigate factors that contribute to machine failure using decision trees.
- Use SAS Visual Statistics to predict equipment or machine failure using logistic regression or decision trees (with more controls over growth and pruning of decision trees).

Marketing Analyst

Financial
Risk Analyst

Engineers

Figure 5: Use the modeling results to create visualizations for deeper exploration of outcomes.
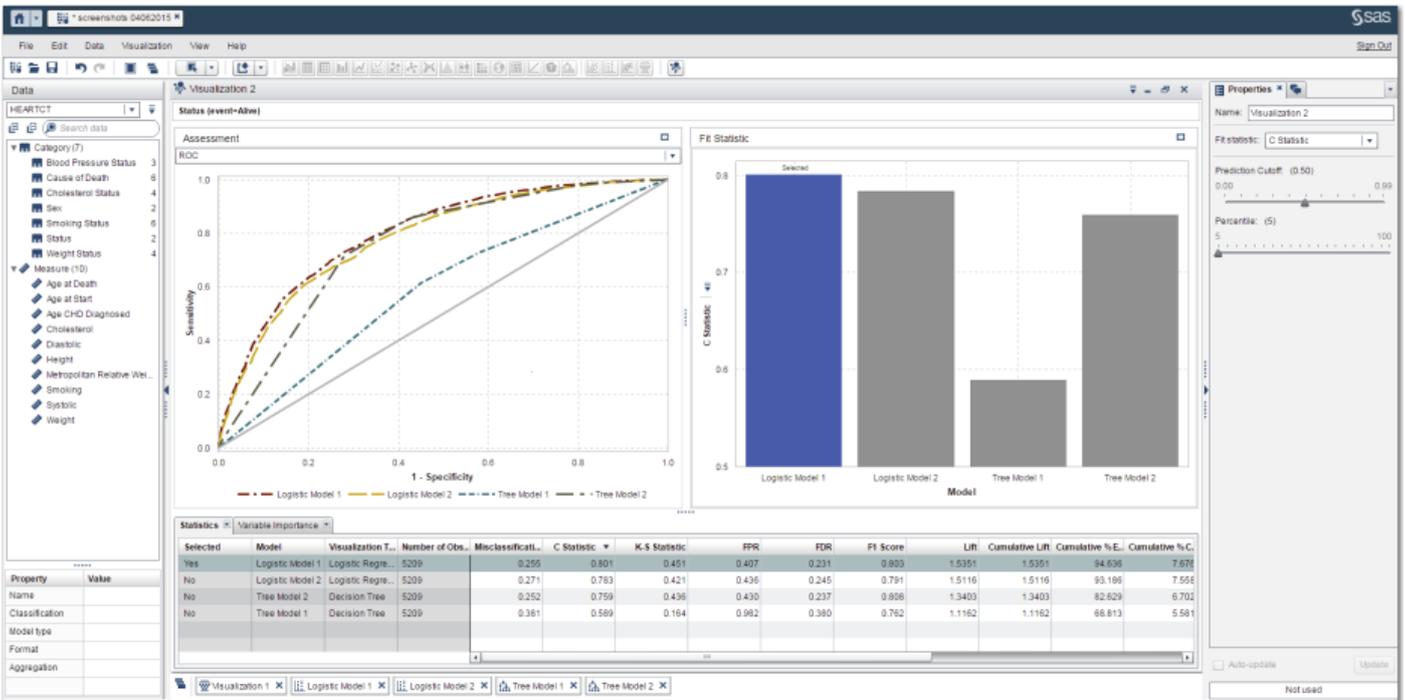

Figure 4: After your models have been created, visually compare competing models to determine which will produce the best outcome.

Conclusion

Merging self-service data exploration and reporting with advanced analytics serves an ever-growing group of users – statisticians, data scientists, line of business analysts, risk managers, marketing analysts and others. The combination makes any enterprise more agile.

With the growing emphasis on using data to create value, more organizations realize the value of increasing their analytics culture with a self-service BI model. They're looking to use increasingly sophisticated analytics to predict the outcomes of future events. And making those capabilities accessible to a wider audience.

So while an easy-to-use intuitive interface is necessary, the power behind the interface is even more important. SAS Visual Statistics, along with SAS Visual Analytics, empowers statisticians, data scientists, business analysts and even nontechnical users with a fast, fun way to explore data, build predictive models and assess them to find the best performer. By putting this software to work for your organization, you can:

- **Beat the competition with savvy insights**. Quickly surface insights hidden in vast data stores. Discover and evaluate new opportunities your competitors will miss. Find new ways to grow revenue. Because SAS Visual Analytics is included with the solution, business analysts and statisticians can use powerful, predictive analytics and visual data exploration capabilities to do more with data than ever before.

- **Run more models faster – with precision**. Our multicore processing environment reduces that to minutes. You can build models to target specific groups or segments, and run numerous scenarios simultaneously. Ask more what-if questions and get fast answers. Refined models produce better results.

- **Stay agile with in-memory computing**. Perform fast yet complex analytic computations using the in-memory engine. Modelers can quickly test new ideas, try different modeling techniques and refine models on the fly to produce the best results – using data volumes never before possible.

- **Boost your analytical teams' productivity.** Multiple users can quickly and interactively customize models – adding or changing variables, removing outliers, etc. – and instantaneously see how those changes affect model outcomes. Which model provides the most predictive power? It's now easy to find out – and in the process, get more value from your big data analytics.

The end result is a self-service data discovery and predictive analytical environment that delivers more precise insights to decision makers when and where they need them. Your analytics journey just got a lot easier.

☛ Try SAS® Visual Analytics

☛ Try SAS® Visual Statistics

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

sas

THE POWER TO KNOW®