Five big data challenges

And how to overcome them with visual analytics

Big data is set to offer companies tremendous insight. But with terabytes and petabytes of data pouring in to organizations today, traditional architectures and infrastructures are not up to the challenge. IT teams are burdened with ever-growing requests for data, ad hoc analyses and one-off reports. Decision makers become frustrated because it takes hours or days to get answers to questions, if at all. More users are expecting self-service access to information in a form they can easily understand and share with others.

This begs the question: How do you present big data in a way that business leaders can quickly understand and use? This is not a minor consideration. Mining millions of rows of data creates a big headache for analysts tasked with sorting and presenting data.

Organizations often approach the problem in one of two ways: Build "samples" so that it is easier to both analyze and present the data, or create template charts and graphs that can accept certain types of information. Both approaches miss the potential for big data.

Instead, consider pairing big data with visual analytics so that you use all the data and receive automated help in selecting the best ways to present the data. This frees staff to deploy insights from data. Think of your data as a great, but messy, story. Visual analytics is the master filmmaker and the gifted editor who bring the story to life.

Data visualization is becoming an increasingly important component of analytics in the age of big data.
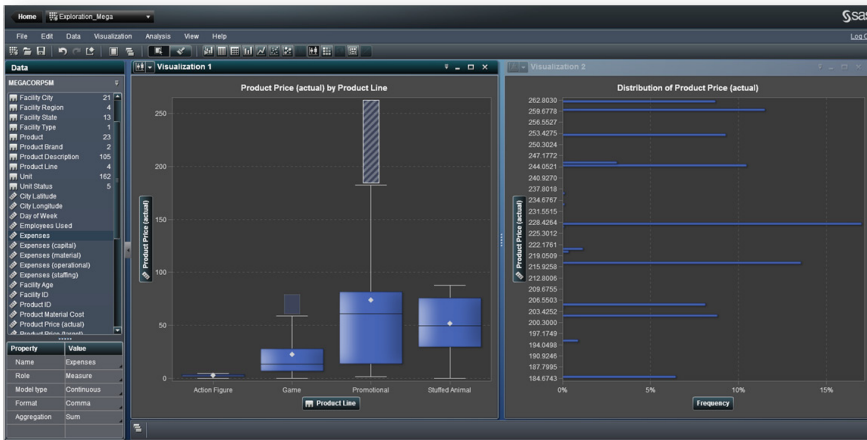
Grouping data together, or "binning," can help you easily visualize large quantities of data, including outliers.

To fully take advantage of visual analytics, organizations will need to address several challenges related to visualization and big data. Here we've outlined some of those key challenges – and potential solutions.

1 Meeting the need for speed

In today's hypercompetitive business environment, companies not only have to find and analyze the relevant data they need, they must find it quickly. Visualization helps organizations perform analyses and make decisions much more rapidly, but the challenge is going through the sheer volumes of data and accessing the level of detail needed, all at a high speed. The challenge only grows as the degree of granularity increases. One possible solution is hardware. Some vendors are using increased memory and powerful parallel processing to crunch large volumes of data extremely quickly. Another method is putting data in-memory but using a grid computing approach, where many machines are used to solve a problem. Both approaches allow organizations to explore huge data volumes and gain business insights in near-real time.

2 Understanding the data

It takes a lot of understanding to get data in the right shape so that you can use visualization as part of data analysis. For example, if the data comes from social media content, you need to know who the user is in a general sense – such as a customer using a particular set of products – and understand what it is you're trying to visualize out of the data. Without some sort of context, visualization tools are likely to be of less value to the user.

One solution to this challenge is to have the proper domain expertise in place. Make sure the people analyzing the data have a deep understanding of where the data comes from, what audience will be consuming the data and how that audience will interpret the information.

3 Addressing data quality

Even if you can find and analyze data quickly and put it in the proper context for the audience that will be consuming the information, the value of data for decision-making purposes will be jeopardized if the data is not accurate or timely. This is a challenge with any data analysis, but when considering the volumes of information involved in big data projects, it becomes even more pronounced. Again, data visualization will only prove to be a valuable tool if the data quality is assured. To address this issue, companies need to have a data governance or information management process in place to ensure the data is clean. It's always best to have a pro-active method to address data quality issues so problems won't arise later.

4 Displaying meaningful results

Plotting points on a graph for analysis becomes difficult when dealing with extremely large amounts of information or a variety of categories of information. For example, imagine you have 10 billion rows of retail SKU data that you're trying to compare. The user trying to view 10 billion plots on the screen will have a hard time seeing so many data points. One way to resolve this is to cluster data into a higher-level view where smaller groups of data become visible. By grouping the data together, or "binning," you can more effectively visualize the data.

5 Dealing with outliers

The graphical representations of data made possible by visualization can communicate trends and outliers much faster than tables containing numbers and text. Users can easily spot issues that need attention simply by glancing at a chart. Outliers typically represent about 1 to 5 percent of data, but when you're working with massive amounts of data, viewing 1 to 5 percent of the data is rather difficult. How do you represent those points without getting into plotting issues? Possible solutions are to remove the outliers from the data (and therefore from the chart) or to create a separate chart for the outliers. You can also bin the results to both view the distribution of data and see the outliers. While outliers may not be representative of the data, they may also reveal previously unseen and potentially valuable insights.

Conclusion

As more and more businesses are discovering, data visualization is becoming an increasingly important component of analytics in the age of big data. The availability of new in-memory technology and high-performance analytics that use data visualization is providing a better way to analyze data more quickly than ever. Visual analytics enables organizations to take raw data and present it in a meaningful way that generates the most value. Nevertheless, when used with big data, visualization is bound to lead to some challenges. If you're prepared to deal with these hurdles, the opportunity for success with a data visualization strategy is much greater.

For more information and to test drive SAS® Visual Analytics, visit sas.com/visualanalytics