

Brought to you by:



Visual Analytics

for
dummies
A Wiley Brand



Understand visual analytics in simple terms

Identify state-of-the-art capabilities

See modern data analytics in action

Qlik
Special Edition

Jared Decker

About Qlik

Qlik® is the leading visual analytics platform and the pioneer of user-driven business intelligence. Its portfolio of cloud-based and on-premises solutions meets customers' growing needs from reporting and self-service visual analysis to guided, embedded, and custom analytics, regardless of where data is located. Customers using Qlik Sense® and QlikView® gain meaning out of information from multiple sources, exploring the hidden relationships within data that lead to insights that ignite good ideas. Headquartered in Radnor, Pennsylvania, Qlik does business in more than 100 countries with over 40,000 customers globally.

Visit qlik.com for more information.



Visual Analytics

Qlik Special Edition

by Jared Decker

**for
dummies®**
A Wiley Brand

Visual Analytics For Dummies®, Qlik Special Edition

Published by
John Wiley & Sons, Inc.
111 River St.
Hoboken, NJ 07030-5774
www.wiley.com

Copyright © 2018 by John Wiley & Sons, Inc., Hoboken, New Jersey

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

Trademarks: Wiley, For Dummies, the Dummies Man logo, The Dummies Way, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc., is not associated with any product or vendor mentioned in this book.

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: THE PUBLISHER AND THE AUTHOR MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES OR PROMOTIONAL MATERIALS. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR EVERY SITUATION. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING LEGAL, ACCOUNTING, OR OTHER PROFESSIONAL SERVICES. IF PROFESSIONAL ASSISTANCE IS REQUIRED, THE SERVICES OF A COMPETENT PROFESSIONAL PERSON SHOULD BE SOUGHT. NEITHER THE PUBLISHER NOR THE AUTHOR SHALL BE LIABLE FOR DAMAGES ARISING HEREFROM. THE FACT THAT AN ORGANIZATION OR WEBSITE IS REFERRED TO IN THIS WORK AS A CITATION AND/OR A POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE AUTHOR OR THE PUBLISHER ENDORSES THE INFORMATION THE ORGANIZATION OR WEBSITE MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. FURTHER, READERS SHOULD BE AWARE THAT INTERNET WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ.

ISBN 978-1-119-44842-6 (pbk); ISBN 978-1-119-44850-1 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

For general information on our other products and services, or how to create a custom *For Dummies* book for your business or organization, please contact our Business Development Department in the U.S. at 877-409-4177, contact info@dummies.biz, or visit www.wiley.com/go/custompub. For information about licensing the *For Dummies* brand for products or services, contact BrandedRights&Licenses@Wiley.com.

Publisher's Acknowledgments

Some of the people who helped bring this book to market include the following:

Development Editor:

Elizabeth Kuball

Copy Editor:

Elizabeth Kuball

Executive Editor:

Katie Mohr

Editorial Manager:

Rev Mengle

Business Development

Representative: Karen Hattan

Production Editor:

Selvakumaran Rajendiran

Special Help:

Chris Mabardy,

Josh Good, Deborah Athens

Introduction

In today's world, data is produced at an incredible rate, and we're able to amass stores of data faster than we're able to analyze it. Data analysis methods have been developed over the past decade to help people make sense of all that data, turning it into actionable insight and competitive advantage.

Data analytics methods allow people to draw conclusions and insights from raw data in ways that are impossible with summary statistics or tabular reports, resulting in more-informed, data-driven decisions. In modern data analytics, robust data processing and user interface capabilities combine to enable analytic discovery through visual free association and exploration of data. The result is that teams (and individual team members) can sift through and gain insight into massive amounts of data with unprecedented speed and flexibility.

About This Book

Discussions of data analytics are often dominated by technical lingo, and it can be difficult even for practitioners to keep up with the latest advances. This book is for the rest of us. The book is about data analytics of the visual variety, and no technical background is assumed.

To start, I define visual analytics in simple terms and outline ways to leverage it. Then I dig into the benefits of visual analytics, discussing when and why it's used, and point out some unique features of visual analytics compared with other approaches to working with and analyzing data.

Next, I look at modern data analytics, covering paradigms that have come out in recent years and state-of-the-art data analytics platforms. In closing, I present ten takeaways on data analytics.

Icons Used in This Book

This book uses icons in the margin to draw your attention to certain kinds of information. Here's a guide to the icons:



TIP

I use the Tip icon to highlight anything that'll save you time or money, or just make your life a little easier.



REMEMBER

When I tell you something so important that you should commit it to memory, I mark it with the Remember icon.



WARNING

When I want you to avoid making a potentially costly mistake, I mark that material with the Warning icon.



TECHNICAL
STUFF

Sometimes, I get into the weeds, providing some information that's a bit more technical in nature. When I do, I mark it with the Technical Stuff icon.

Beyond the Book

For more information on visual analytics, go to www.qlik.com.

- » Getting to know visual analytics
- » Seeing what visual analytics can do
- » Understanding the visual analytics process

Chapter 1

Introducing Visual Analytics

A picture is worth a thousand words, but a visualization is worth countless data points. The same might be said of summary statistics, except that datasets that are shown to be statistically similar (or even identical) by standard statistical measures (average, variance, correlation) can be quite different when represented visually. Data visualization uses the visual processing power of the human brain to help people draw insights and conclusions from data quickly, in ways that sophisticated pattern detection algorithms and tabular reports can't match.

This chapter defines visual analytics in terms of the processes and capabilities that make it possible.

Defining Visual Analytics

Data visualization is important, but realizing the value of visual analytics is much more than just creating visualizations; it includes accessing and bringing together many different data sources into a format that's ideal for analysis, and providing users with the tools and methodologies that make analysis possible.

Visual analytics includes the following practices:

- » Data preparation and management
- » Creating data visualizations
- » Search, exploration, and analysis
- » Collaboration and sharing

I discuss these practices separately in the following sections.

Data preparation and management

Sometimes before data can be visualized, it must be identified, prepared, and then maintained for repeat analysis. In many cases, data from different sources must be brought together to form a cohesive picture. Yet raw data can be boundless and unwieldy. Without the right tools, data preparation and management can be a challenging, time-consuming part of the visual analytics process.

Some visual analytics tools facilitate the process by providing user-friendly data preparation tools that help people prepare and combine raw data for inclusion in the visual analytics user interface, as well as scheduling ongoing data preparation jobs to retrieve updated data.

Data visualization

Visual analytics deliver visual representations of data in a format ideal for human reasoning by harnessing the brain's ability to identify patterns, leading to discoveries that would otherwise go missed. When data is visualized, important features stand out, such as correlations, patterns, groups, outliers, and trends.



In science, business, public policy, and many other fields, data visualization helps people find insights, draw conclusions, and support decisions. Even agencies that have vast computing resources at their disposal rely on data visualization.

Search, exploration, and analysis

When data is brought together and visualizations are created, the visual analytics process is well under way. To be truly useful, however, data visualizations must be included in a user interface

that supports the business users' need to search, explore, and analyze.

Search has a known destination; specific terms, values, or conditions are being searched for. In the case of exploration, the destination is unknown; new areas are being investigated, or familiar areas are being explored at deeper levels.

Collaboration and sharing

The insights drawn from visual analytics are useful to the extent that they're shared with the people who can take action. As a result, visual analytics tools typically facilitate collaboration and sharing.

In the simplest cases, chart data or the images themselves can be downloaded from the visual analytics user interface for sharing via email or presentations. And often, visual analytics tools offer onboard storytelling features for communicating the insights contained in visualizations, in a linear yet interactive fashion.

Identifying the Benefits of Visual Analytics

Visual analytics represents a graphical way of working with data to understand information. Companies can now gain insights about their organizations in ways they never could before. For many businesses, visual analytics has become an indispensable part of informed decision-making.



TIP

To understand the benefits of visual analytics, consider the ways in which humans use information to make decisions. Humans are visual creatures, and when making decisions, we're likely to put extra weight on information that matches our visual perception capabilities. We detect patterns quickly, as well as notice certain details in milliseconds, such as colors that stand apart from other colors. Other types of visual features require more focus but stand out just as prominently in our perception.

In the era of big data, significant amounts of information are ready to be mined for insights. Given the complexities of dealing with data that can often seem boundless and unwieldy, getting from information to insight isn't always a trivial matter.

Data professionals, such as business analysts, are all too familiar with the challenges of working with large amounts of disparate data. Using whatever tools are available, an analyst often builds decision-support solutions that require an enormous amount of time to develop and maintain. When the analyst moves to a new position, those left behind have to figure out how to take ownership of the proprietary analytics solution.

Visual analytics tools offer data preparation and management capabilities that provide quality data visualization combined with search and exploration capabilities. The result is the ongoing, scalable capability to turn information into insight in the following ways:

- » Communicate data clearly.
- » Highlight patterns that would otherwise be difficult to see.
- » Make large or complex datasets comprehensible with minimal effort and space.
- » Increase user comprehension and engagement.

Employing the 3 A's of Analytics

Visual analytics is a process that leads to consistency in data collection and analytical processing, as well as clarity and insight from the data that is used to inform decisions. These steps can be summarized in a simple framework that I call the three A's of analytics: acquisition, analysis, and action.

Acquisition

The first step is identifying the right data. Data can come from company databases, emailed spreadsheets, and external sources. The possibilities are endless.

After relevant data is identified, additional steps are required to get it into a format appropriate for data analysis. These steps may include data cleansing, normalization, grouping, separating, pivoting, and integration (bringing together separate datasets).

After the raw data has been prepared, some or all of the data may be staged for future analysis, eliminating the need to perform the same steps on the same data every time.

Finally, some preprocessing is done on the data so that the visual analysis can be carried out at the speed of thought. The type of query preprocessing that's performed depends on the visual analytics tool being used (see Chapter 2).

Analysis

In visual analytics, data is visualized in standard chart types or custom visualizations. The standard types fall into four categories:

- » **Comparison:** Comparison views show high and low values on a relative basis, making it easy to see patterns and groupings. Typical comparison charts are bar, line, and circular area (sometimes called radar charts).
- » **Composition:** Composition views identify the relative differences between parts of a whole. Typical composition charts are treemap, stacked area, stacked bar, and pie.
- » **Distribution:** Distribution views show the general shape of a range and identify commonalities and outliers. Typical distribution charts are bar/line histogram and scatter plot.
- » **Relationship:** Relationship views show correlations, clusters, and outliers. A scatter plot is often used to show relationships.

Visual analytics tools allow users to customize standard chart types, such as by changing the colors and sizes of points on a scatter plot to show further dimensionality and measures. The degree to which charts can be tailored for maximum visual information is a useful measure of the power of a visual analytics tool. With some advanced visual analytics tools, users can also define custom visualizations programmatically in ways that greatly enhance the capabilities of the tool.

Users can employ other data visualization capabilities and tools, depending on the analysis goal. That goal tends to fall into one of three categories:

- » **Discovery:** When the analysis goal is discovery, visual analytics allow the user to freely explore the data in an analytical, unscripted manner. The user interface should support discovery through point-and-click actions and natural-language search that displays results visually. This process often leads to interesting discoveries, some of which lead to inquiry.
- » **Inquiry:** When the analysis goal is inquiry, visual analytics allow the user to investigate a specific event or condition, usually based on a question or hypothesis. Inquiry typically requires extensive use of conditions (or reverse conditions) that the user applies to shape the data per evolving hypotheses. Inquiry also requires robust search capability. This process sometimes creates a need for ongoing monitoring, which in turn leads to evaluation.
- » **Evaluation:** When the analysis goal is evaluation, visual analytics allows the user to show measurements relative to a benchmark or predefined target, often referred to as a key performance indicator (KPI). To support evaluation, visual analytics facilitates the creation of predefined dashboards and reports that can display KPIs that are important to the organization.

Action

Analysis often leads to action. In visual analytics, the first action usually is to share insights and conclusions with others. Sharing options depend on the visual analytics tools and range from downloading data or visualizations from the visual analytics interface to generating a hyperlink that calls up the same display of information, along with any filters that have been applied. Sharing information using visual analytics tools allows you to provide new insights to users in an easy-to-understand way.

When users spend time analyzing data and come across new insights, it's important to share those insights to educate relevant audiences. The extent to which insights can be shared with ease and clarity is a good measure of a visual analytics tool's ability to facilitate action.

In recent years, the concept of data storytelling has gained traction. In data storytelling, users assemble what amounts to presentations of data visualizations, along with supporting commentary and data highlights. Storytelling provides an opportunity to extract the most concise and impactful analytical content from the visual analytics tool so that it can be shared externally. Visual analytics tools often facilitate data storytelling.

The true power of data analytics goes beyond creating charts and graphs. Instead, users use the data for free-form, interactive exploration of the data to answer questions and spark curiosity for deeper engagement with the data. With this approach, users will start to use data storytelling techniques to collaborate with others to report on their insights. These stories strengthen the communication of data to make the facts more contextual, digestible, and memorable, which enables the users to better present, convince, and collaborate on their findings.

- » Defining modern data analytics
- » Knowing how modern data analytics works
- » Seeing modern data analytics in action

Chapter 2

Getting Ahead with Modern Data Analytics

As noted in Chapter 1, visual analytics is much more than just simple data visualization; it also includes the tools and processes that facilitate data preparation and management, search and exploration, and collaboration and sharing. With modern data analytics, the state of the art has come a long way.

Modern data analytics has two far-reaching capabilities that distinguish it from traditional visual analytics: the level of insight that can be achieved and the speed at which analytical activities can be carried out. In short, modern data analytics allows more users to achieve deeper insights faster. In this chapter, I cover some of the characteristics of modern data analytics.

Understanding Modern Data Analytics

In modern data analytics, data is not only visualized with high-quality graphics, but also open to exploration in ways that are impossible in traditional visual analytics. Visual analytics tools provide a great deal of feedback when users explore

data interactively, making the interface itself a means of gaining insight into the data. Modern data analytics allows users to immediately understand the relationships in their data, both associated and unrelated, based on the questions they ask. They can interact with data in one visualization and immediately see how data points in other analytics are affected. They can search to discover relationships and refine context across an entire application. And they can answer new questions and explore new ideas, without restrictions or boundaries. These types of analyses can be accomplished with little or no technical expertise.

Modern data analytics continually expand ways to visualize data without requiring technical expertise. For example, with modern data analytics, powerful geospatial visualizations can be easily achieved that would have required extensive preparation and development even with special-purpose geospatial software. For more technical users, the visualizations used in modern data analytics are limited only by users' creativity. For example, a user may implement a chart that displays data as a network diagram that has been added as a custom type of visualization. These capabilities lead to deeper insights.

Another advance with modern data analytics is the speed at which analytical activities can be carried out. Speed can be measured in many ways, from how quickly a user can search for and understand specific characteristics of data to the agility of modern data analytics tools in accessing and combining diverse sources of data. With modern data analytics, speed also involves the capability to process larger amounts of data more efficiently, creating new analysis opportunities from information overload.

Seeing How Modern Data Analytics Works

Modern data analytics relies on several technical advances, which I describe in the following sections.

Robust data processing

Data processing can involve many different types of operations that can be broadly categorized as the cleansing, enriching, and shaping of data. Cleansing often involves removing unnecessary rows

or replacing incorrect values with correct ones. Enriching involves creating new data from existing data, for example by using if/then logic. Shaping entails bringing data together, breaking it apart, or changing the level of detail in the data. Regardless of the type of processing, it must be done efficiently so that end users can work with large amounts of data quickly. All of these activities require an engine that can operate on large data sets and for many concurrent users. Robust data processing includes the following capabilities:



» **Extensive data transformation support:** In the data acquisition process (introduced in Chapter 1), users transform data in a way that makes it useful for the required analysis. Disparate datasets may need to be combined, or a single dataset may need to be broken into separate datasets. A modern data analytics tool will allow users to perform most or all data preparation tasks without requiring additional processing in external tools or in databases using SQL.

SQL is the standard relational database query language; it is usually required to load data from disparate relational databases. However, many people don't know how to write SQL queries, and not every type of data processing operation can be achieved via SQL in a straightforward manner.

It is important to note that SQL and relational databases in general were not designed for some of the activities that are carried out with modern data analytics. Tools that rely solely on SQL for data processing operations can lead to a slower working pace for end users or possibly limited analysis operations. Certain tools offer additional data processing capabilities that don't impose the boundaries and limitations of SQL.

» **Preprocessing:** To provide insights at the speed of thought on even large data volumes, usually some preprocessing must occur. Data preprocessing concepts have been around for a long time, but with modern data analytics, the capabilities have greatly increased. Certain tools require more setup and technical expertise, while others accomplish preprocessing with very little user configuration or intervention.

» **Powerful dynamic calculations:** Modern data analytics allow for many types of real-time calculations to be carried out on the fly. Examples of dynamic calculations include values that change according to selections or other end-user actions, such as filtering, sorting, and pivoting of data.

Free visual association and exploration

Modern data analytics are carried out by exploring data freely in a visual, unscripted way, not just by repeating analyses that were prepared or coded in advance. As a result, users should be able to search for and filter data easily and on an ad-hoc basis. Tools that provide free visual association and exploration have several important features:



REMEMBER

- » **Search:** As search engines have matured, users have grown accustomed to immediate and comprehensive search results. Data analytics tools have varying levels of maturity with regard to search, but the best tools make quick, comprehensive search results available to users across all fields, allowing them to retrieve partial or fuzzy matches and to see possible matches while typing.

From a data analytics perspective, a comprehensive search means searching all numeric and textual fields in the underlying data. The search term *1001* might return results consisting of transaction amounts, street numbers, and so on. The visualized search results should make it clear where the data comes from so that the user can pursue the most relevant result.

- » **Visual association:** Modern data analytics highlights the associations in the data as the user explores. This approach allows users of all levels to easily understand what information is both related and unrelated to their selections as they interact. For example, if a user selects a customer, he can immediately see which products the customer purchased and which they did not, which channels they purchased through and which they did not, and so on. Traditional queries filter out unrelated values which often convey the best insights, such as products the customer did not buy. Having this feedback after every interaction allows users to spot insights and formulate new questions for further exploration.

The simplest tools display field filters that allow users to apply conditions, thereby shaping the data interactively. In more sophisticated tools, visualized data elements in charts are shaped not only by user interface filters (which may be created on the fly from any field), but also by user interaction with other data elements, such as clicking a data point on a chart.



TIP

With this capability, dashboards can contain distinct views of data that are wired together via backend connections, allowing the user to interact with data and immediately see which data are related or not related. This allows the user to interact with data and immediately see the implications. This capability can be extremely powerful when the associations between distinct views of data occur at deeper levels than those that are visible. For example, a user might click a single point in a scatter plot that displays data at the customer level. This filtering action at the customer level in turn affects a bar chart that shows product categories, because the product categories have been filtered to those purchased by a single customer selected from the scatter plot. The same types of interactions can occur on any associated field.

Data analytics tools approach visual association differently. Most query-based tools simply filter out unrelated values, such as customers that did not buy or products that were not sold. In addition, they often require objects to be “wired” together on the front end in order to stay in context together, resulting in multiple queries and performance issues. All this can limit the users’ goal of free visual association and exploration. When you evaluate visual analytics tools, consider the degree of configuration that must be done in advance to approximate something that feels to end users like truly free visual association and exploration.

- » **User interface changes:** In modern data analytics, users can configure data visualizations and see the results interactively. Advanced tools allow users to create personal content that persists and remains in sync with global content. Users can also swap the measure on a bar chart interactively for a different measure (such as switching from sales to costs) or swap the dimension (such as switching from customers to stores). When it’s possible to drill down into additional levels of detail in a chart, users should be able to do so quickly and easily. Users should also be able to create their own charts from scratch.

Modern and open architectures

Modern data analytics can include big data platforms and tools, statistical methods, machine learning, and other sophisticated techniques. A single visual analytics tool can’t address every

possible need. Modern data analytics addresses this by providing open-architecture platforms that allow users to incorporate functionality from external platforms as well as develop new functionality that leverages open-source toolkits.

For example, functionality from an advanced analytics engine such as R or Python may be incorporated. A data analytics solution is the perfect vehicle to deliver the power of advanced analytics to business users in a way they can interactively explore. Data analytics tools may send a dataset to the advanced analytics engine in real time; the engine operates on the dataset via a prediction algorithm and returns the data to the data analytics solution for visualization. This approach gives users the best of both worlds without having to open multiple tools and interfaces.

In the case of big data, modern data analytics tools can connect to and extract data from flexible data architectures such as Hadoop and combine that data with other data sources such as relational databases, providing end users an efficient way to analyze necessary data via a unified approach.

New data processing capabilities can also be developed. A data subscription feed, for example, may allow connectivity via a program that automatically connects the data. The user finds or develops an appropriate module and adds it to the data analytics platform, making it possible to incorporate data from the subscription service into the data analytics solution.

Likewise, new kinds of visualizations may be developed. In one salesperson performance example, a scatter plot depicts number of deals on one axis and sales revenues on the other, summarized by salesperson (one dot per salesperson). In this case, salespeople are not as profitable on a deal-by-deal basis as their volume increases, but this does not show up in an obvious way at the level of detail of the visualization. To highlight high-margin salespeople, a “sweet spot overlay” is developed as a custom visualization that leverages D3 (a data visualization open-source JavaScript library) to extend a regular scatter plot to highlight high-margin salespeople (looking at profitability on a per-deal basis) and then putting a subtle rectangle around those salespeople on the scatter plot that fall into the “sweet spot.” With open architectures, these and many other types of deeper insights are possible.

With modern web architectures, data analytics allows more users to access analytic content from different types of devices such as

tablets and cellphones, rather than solely from their PCs. This makes it possible to access and explore analytics apps at the time and place most needed, or through mobile offline access when no connection is available.

PUTTING DATA ANALYTICS TO WORK

Many organizations have deployed data analytics successfully. Here are two case studies from two different industries.

Infrastructure maintenance

An international infrastructure maintenance company that specializes in protecting underground utilities from damage had poor visibility into its operations. This company, which produces and stores large volumes of data, had trouble retrieving data on resource use, which it needed to reduce fuel costs for its fleet of more than 5,000 trucks. At first, the company used a Microsoft Excel-based solution to gather the necessary data, but this approach required extensive data preparation.

Eventually, the company deployed a modern data analytics solution that greatly facilitated the data acquisition process. More important, the solution uncovered a previously unknown relationship between idle time and fuel expenses. Armed with this information, the company saved \$2.5 million within the first three months.

Retail furnishings

A multichannel modern furnishings retailer with more than 40 locations across the United States faced reporting challenges with regard to inventory analysis. Also, area sales managers needed improved reporting of sales, demand, and trends in stores.

The company deployed a modern data analytics solution that acquired relevant data from the existing system with minimal effort, including sales and inventory data for every product and location. This allowed the salesperson to accommodate on-the-go decisions about which inventory to use for the deal at hand. The real-time data analytics was provided via tablets. This solution allowed the company to save money by selecting the best product and location of the inventory for each new deal they completed.

- » Seeing why companies need data analytics
- » Working with strategic data assets
- » Reviewing technical capabilities

Chapter 3

Data Analytics for Organizations

Individual users can use data analytics tools to quickly draw conclusions and insights from raw data. The benefits are greater, however, and the total cost of ownership is lower when modern data analytics is deployed as a solution that benefits the entire organization. Individual efforts in any of the three A's of analytics (acquisition, analysis, and action; see Chapter 1) are multiplied when shared across teams.

Rarely do solutions fail for technical reasons. They almost always fail due to poor project management or change management. But in addition to strong project management, the necessary technical framework must be in place and must be capable of delivering at the enterprise level. This chapter discusses deploying data analytics solutions to address organizational needs rather than just individual users' needs.

Building and Maintaining Strategic Data Assets

As discussed in Chapter 1, data acquisition requires data preparation and management. A data analytics platform can greatly facilitate this process, but effort is still required to ensure that the right data is being used and then to configure it properly for analysis. The level of effort involved in acquiring data varies, depending on each user's technical abilities and familiarity with the data.

Modern data analytics makes it feasible for users to import data of their choosing for analysis on an ad-hoc basis, instead of relying solely on data that has been stored in centrally managed databases. When this happens repeatedly by more than one user, sometimes what started out as ad-hoc data starts to look more like a longer-term data acquisition requirement, kept up to date with the latest data, for the benefit of many. To minimize the organizational effort spent acquiring data, it often makes sense for people who have specialized knowledge of the data to be responsible for it (from an analytics perspective), choosing where and how to extract it, modeling it correctly in the data analytics platform, and ensuring its ongoing consistency. Generally, the people who have sufficient knowledge to prepare and manage data are the logical ones to secure it. Securing data means ensuring that the right organizational roles have access to data — no more and no fewer.



REMEMBER

IT departments call this responsibility *data governance*. The role of data governance is to foster trust in data quality by making sure that it's consistent, accurate, current, and secured.

When data is properly maintained and secured, it constitutes a strategic data asset. This asset is available in the data analytics platform for users to draw on to make discoveries and inquiries, as well as evaluate ongoing performance. In addition, they should be able to combine governed data with separate data that they may have acquired. As they draw conclusions and insights, users can easily take action by sharing their findings by creating and sharing interactive data stories with others in the organization — a process that's greatly facilitated if the platform is widely available. When users can share a link that retrieves the relevant information directly from the data analytics platform, you allow them to gain greater insights than they could with a screenshot or static dataset.

While analysis is being carried out on these strategic data assets, organizational leadership can also be confident that the data being used is accurate, consistent, and secure from unauthorized use.

Considering Technology Capabilities

When an organization decides to move beyond the individualized use of tools to a modern data analytics platform, it must consider the capabilities of multiple candidate platforms.

A modern platform approach enables users and companies to use one platform that provides the necessary core functionality, instead of supporting a plethora of products. This platform approach should support the full range of use cases in an organization (as well as embedded analytics for external users), all within a single solution enabling teams across the enterprise to collaborate freely without the risk of data chaos.

The following sections describe the functional areas that companies should look for in a data analytics platform.

Data processing speed and flexibility

Data processing should be equally capable on small as well as large data volumes, and should encompass a wide range of operations, from cleansing to enriching to shaping. In addition, data processing capabilities should be accessible to end users without requiring advanced development skills. When users understand the power and flexibility of an analytics platform to process disparate and voluminous data, the barriers of accessing powerful analytics are lowered for all.



WARNING

As I mention in Chapter 2, different platforms take different approaches to providing fast responses to user interactions. What can be difficult to see is the implications of each approach. For example, some approaches may require greater IT involvement, especially with larger data volumes. It is important to understand how a given platform handles user interface performance. Ideally, end users can accomplish their analytics goals without IT intervention.

The accessibility and flexibility of data processing capabilities can be more difficult to evaluate and should be investigated thoroughly. Most platforms offer intuitive approaches for simple operations such as joining disparate datasets. For more sophisticated operations, not every type of operation is supported by every tool. When a data processing operation (such as breaking data apart and recombining it differently) cannot be handled with the data analytics platform, it must be carried out externally, typically in a database, and so on. In such cases, data processing needs are not accessible. Both the power and accessibility of data processing are important considerations when evaluating a data analytics platform.

Collaboration facilities

The ability to share insights is paramount, so you should look for collaboration features in a data analytics platform, such as easy export of raw data and high-definition images of data visualizations.

Some platforms allow users to create *data stories*. This means that you can present PowerPoint-like presentations that tell a story about your information using charts, callouts (to describe interesting points), and descriptive text. This helps your audience more quickly understand the implications of your findings. You become a storyteller that provides insights that can make a difference to your company's bottom line. The stories can also link directly to live analytic content. This is achieved by simply clicking on an analytic in a story to go directly to the sheet where it was taken from, in context, to enable further exploration, without the need for any further setup.

A good way to measure the collaboration facilities of a platform is to evaluate the ease of sharing insights, as well as the ease with which other users can pick up and carry the analysis further.

Deployment flexibility

You certainly want to know what options are available for enterprise deployment. Most likely, your IT department looks at deployment flexibility in terms of being able to support on-premises installations, cloud-based solutions, and hybrid deployments. If your organization has a large user base, you want to ensure that

the data analytics platform can be augmented as necessary to serve all users. In other words, scalability matters.

Another deployment consideration that may be brought up is how the data analytics UI can be provided to users in flexible ways. It's important to understand whether the UI will display in a professional and usable manner on different web browsers and screen sizes, from PC to tablet or smartphone. In other cases, some organizations may have a need to embed specific visualizations, UI controls, or dashboards from the data analytics solution into a proprietary web page that has a custom look and feel though enhanced with powerful capabilities from the underlying data analytics solution. Depending on the deployment flexibility of the platform, this could be easily achieved or quite difficult. Overall, deployment flexibility can be measured in terms of both flexibility in platform configuration and agility in delivering analytics to end users.

Customization and extensibility

A key feature to look for is the capability to add new functionality to the data analytics platform. (This feature has much to do with open architectures, which I discuss in Chapter 2.) A data analytics platform designed with an open architecture approach supports customization and extensibility. Data analytics platforms offer open architecture in different ways, such as open APIs or integrations with specific external engines, or a combination of both.



A good way to evaluate customization and extensibility capabilities is to understand the types of integration that the platform can achieve (such as integration with programming languages or with open-source libraries). When a platform offers flexibility and customization possibilities, it will adapt to the evolving needs of your organization. An equally important consideration is how easy it will be to incorporate additional data analytics tools and platforms in the future.

Data security and reliability

If data analytics are to be provided as an organization-wide solution, users must have confidence in the information coming out

of the platform. Different platforms have different approaches to ensuring data reliability. Here are some features to look for:

- » **Common measures:** Wherever there are calculations, it's important to define the measure expressions once and reuse them consistently. Otherwise, different departments may apply different calculation logic, resulting in inconsistent information. To get around this problem, look for a solution that centrally defines business logic.
- » **Governance:** Though not always the case, it's usually desirable to make specific groups responsible for the acquisition of data that can be analyzed by the rest of the organization. Data analytics platforms usually can be set up in a way that achieves this outcome, though some approaches are more flexible and maintainable than others. When evaluating data analytics platforms, be sure that you understand how capable and flexible the platform is with securing and maintaining strategic data assets. This will ensure that all users benefit from reliable and secure data for their data analytics activities.
- » **Security:** Security is mandatory, and it should be a critical focus of your evaluation of a data analytics platform. You want to ensure, for example, that users must be authenticated and authorized for access to specific data or applications. You also want to know that the platform can integrate with your organization's existing security protocols.



One important aspect of security that often gets missed by IT but that can be important to management is column- and row-based security. Column- and row-based security allows users to see only the data that is relevant to their particular department or function. So, the same dashboard report may show two users different totals or segments of a pie chart because those users are responsible for different departments.

Enterprise readiness

When you commit to providing data analytics across your organization, you must ensure that the platform is enterprise-ready. As discussed in the preceding sections, the capabilities that constitute enterprise readiness include data reliability and

security, powerful and accessible data processing, collaboration features, deployment flexibility, and customization.



WARNING

It may be difficult to decide which platform outperforms another because you can't use simple measures to compare performance. Many data analytics solutions display strong capabilities in standard functions such as quality of visualizations, but not equally well in all areas, such as visual free association. Because of this, you should understand both the range of capabilities and their depth as you look for an organization-wide solution. You don't want to get stuck with a simple tool when you need a substantial data analytics platform to support your needs.

- » Benefiting from data sharing
- » Developing strategic data assets
- » Enabling flexibility

Chapter 4

Ten Key Visual Analytics Takeaways

Visual analytics tools can provide exciting insights for your business. Here are ten takeaways that help companies be more successful deploying their visual data:

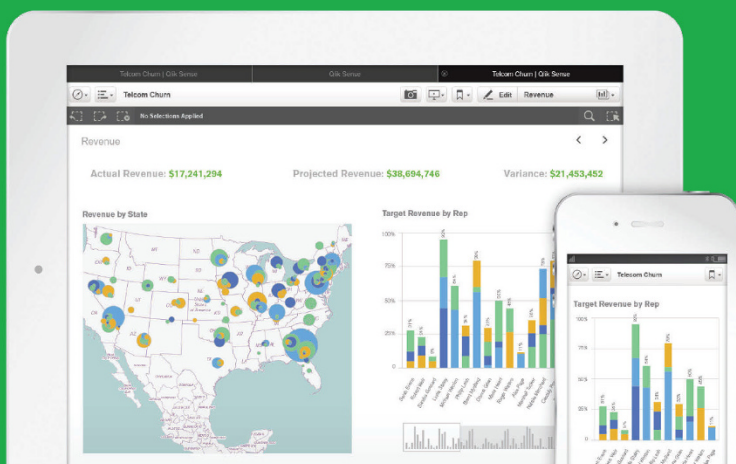
- » **Utilize the power of patterns.** Our brains are hard-wired to see patterns in data. Getting acquainted with the many visual ways people consume information and using this knowledge to build a data analytics platform will make your data more accessible to your audience. Make sure the visualizations and analytics you provide are strong in showing patterns, showing the overall shape of data, and pinpointing outliers.
- » **Deploy interactive capabilities.** High-quality analytics influence decisions and lead to new insights. Many software tools create compelling data visualizations, but data analytics is much more than creating visualizations. It allows users to acquire data and perform deep analysis on that data, which leads to conversations between users and, ultimately, action. When deploying data analytics, determine the types of data and questions that your users will be asking; then deploy

analytics that allow them to explore and uncover new insights using the platform's interactive capabilities.

- » **Reap the benefits of sharing.** Organizations derive greater benefits when staff share data analytics instead of siloing the data and performing separate data analyses in each department. The true value of analytics is fully realized when people collaborate on the data by sharing, discovering, and discussing their insights. This is also true when it comes to data acquisition. When many stakeholders are included in gathering and vetting data, those with specialized knowledge can enhance and improve the data for everyone. This increases the value of the data long-term.
- » **Build strategic data assets.** Data should be constantly evaluated to determine if it is of sufficient value and importance to be included as a strategic data asset. This requires agility to be able to add or remove data from the strategic data store. The best solutions can seamlessly use data from all sources — whether it's governed or ungoverned.
- » **Focus on solutions.** Many tools can build visualizations, but a platform approach is necessary to provide a long-term solution for the organization. When a solution is built on a strong data analytics platform, the three A's of analytics (see Chapter 1) can be performed over and over on different data and for different user groups.
- » **Address all analysis needs.** Your data analytics solution must be able to carry out the three primary types of analysis: discovery, inquiry, and evaluation (see Chapter 1). In other words, the same platform that you use to look at daily performance can also be used to investigate a decline in production.
- » **Embrace open architectures.** If you're not deploying data analytics solutions that leverage open architectures, you're missing something valuable. An array of open-source data analytics tools is available; in addition, you can develop custom capabilities in-house. Adopt a data analytics platform that you can customize and extend.
- » **Ensure deployment flexibility.** Your data analytics platform needs to support on-premises, cloud, and hybrid deployments (see Chapter 3), and it must be agile and capable of growing to serve large numbers of users and high volumes of data.

- » **Guarantee enterprise readiness.** Although many data analytics tools advertise their “enterprise readiness,” it’s important to analyze that claim to confirm that the solution under consideration can meet your requirements. Specifically, you should look for flexible capabilities regarding governance, security, and scalability.
- » **Focus on data processing power.** As data volumes increase and new types of disparate data become meaningful, data analytics solutions must process more and more data. Think carefully about how you use your platform’s data processing power to get the job done quickly and effectively, without limiting flexibility for users. Don’t rely on the same processes as volume increases. Work with your IT department to try new ways to transform your growing databanks.

Smarter analytics for every business.



Every business deserves modern, data-driven insights. Reach them faster than ever with Qlik®. Only Qlik has the Associative Difference™, which gives you the freedom to explore all of your data instantly, from any source, without the limits of predefined queries. Boost productivity. Enable collaboration. And give everyone on your team game-changing insights at their fingertips. Smarter business decisions have never been easier.

Learn more at qlik.com

Achieve deeper insights faster

See why visual analytics methods allow people to draw conclusions and insights from raw data in ways that are impossible with summary statistics or tabular reports, and how modern data analytics employs robust data-processing and user-interface capabilities to enable analytic discovery through visual free association and exploration of data, resulting in more informed, data-driven decisions.

Inside...

- Visual analytics versus other methods
- Employing the three A's of analytics
- Understanding modern data analytics
- Keeping an eye on data security
- Collaboration through analytics
- Building strategic data assets for analytics



Jared Decker is VP of Business Intelligence at Cyber Group and has worked in the field of analytics for more than 15 years. Jared has a degree in Decision Support Systems from the University of Tampa and an MBA from the University of Houston. Aside from frequent consulting and training, he has written several books and contributes regularly to industry publications.

Go to [Dummies.com](https://dummies.com)[®]
for videos, step-by-step photos,
how-to articles, or to shop!

**for
dummies**[®]
A Wiley Brand

ISBN: 978-1-119-44842-6
Not for resale

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.