



Searching for Insight Through Analytics

David Loshin, Knowledge Integrity



Gaining Insight through Analytics

The use of enterprise data as a value-adding business resource has grown and matured over the past decade, but while data warehouses and operational data stores have become common in many organizations, their use is often limited to standardized operational reporting or ad hoc queries. Yet, not only is there a growing interest in using warehoused data for more than simple reports, business analysts increasingly want to rely on more sophisticated analytical models to identify ways to improve operational efficiency, expose market opportunities, and generally add value to the organization's competitiveness. As the interest in exploiting analytics for business advantage grows, so does the desire to simplify the general ability to make use of multiple data sets and to put these capabilities into the hands of more front-line business users. This suggests the need to exploit different types of visualization methods for communicating insights, through an easy-to-use platform for iterative, undirected analysis that can incorporate information from many sources and not just rely on what is stored in a data warehouse.

The development of analytics-based models for business productivity improvement or optimization has always relied on the analyst's abilities in both understanding the data and the knowledge of the tools used for model development. Historically, analytics have been developed as a product of a series of iterations: developing theories based on the available variables, seeking correlation or even causality, refining the variable set for simplification purposes, checking the results, then taking some business action. However, the iterative nature of the process is impacted by the production cycle for model development: speculate on the dependent variables, perform some analysis, assert some theory, validate against a larger data set, review the results, and, based on those results, start the cycle again. Therefore, when the lion's share of the time is spent in wrestling with development tools, it decreases the time spent in analyzing and building effective models, and keeps analysis in the back office, not in the hands of front-line business users.

This poses two significant barriers to entry for mainstreaming business analytics: the sole reliance on established data silos (such as existing data warehouses), designed to address specific reporting needs, to support the creative requirements of analytics, and the need to master tools in order to craft the analytic models. In this paper, we will look at the iterative process analysts use to create analytic models, and consider the kinds of capabilities that business analytics tools should support in order to not only simplify the process of model development, but also reduce the need for subject matter expertise. In turn, reducing the dependence on traditional analytical tools requiring programming knowledge or deep application skills opens up the world of business analytics to a more "democratic" constituency, effectively equipping anyone in the enterprise to be empowered to develop and use analytics to achieve their business objectives.



The Iterative Aspects of Analytics

To understand the technology issues associated with business analytics, it is worthwhile to consider the iterative process associated with developing a model. Essentially, an analytics model captures the thought processes designed to answer specific questions in a way that will trigger actions to positively impact the business. Standard business intelligence reports are designed a priori to reflect productivity or performance metrics associated with predefined expectations, effectively answering known questions. The analytics approach differs from standard business intelligence in that the objective is to first figure out the right questions to ask, and then to find the answers. Ultimately, the model represents the result of figuring out the relevant questions coupled with the processes for deducing and visually presenting the answers.

The analytics process can be compared to web searching, which in its own right is essentially a “problem-solving” process. Individuals use search engines to locate document sources related to a concept embodied by a search term. Web searching is an iterative process, ending when the searcher has found the right resources to meet the objective. At the start of the process the searcher speculates as to the proper search terms to be used for the query. Based on the set of document references returned, the searcher determines whether the results are suitable, and if not, refines the search space by injecting precision based on terms that can be culled out of the referenced document descriptions.

This process is performed iteratively, sometimes introducing new search term combinations or restricting others. As phrases and terms are included or excluded from the query, the results are narrowed, until the documents returned reflect the desired goal, with a set of documents that address the original set of questions. In other words, as one reviews the results of each successive search, it is possible to understand more about the topic in question, possibly finding new and different aspects of the solution that had not been considered in the first place. In fact, one might consider that part of the process is figuring out the right questions to ask in order to draw the conclusions that provide the desired positive business impact. In essence, the analyst is evolving the question as a by-product of seeking the answer. Interestingly, not only will the searcher find the desired results, but the iterative process will also provide a query string used to find those results that can be re-used in the future! In this relatively simple analogy, the query string is the model, representing the “right question” to ask and the means by which the answers can be retrieved.

The same thought processes used for searching are employed when using richer and much more powerful frameworks employing business analytics. In fact, the value is magnified using an analytics framework, since the search process uses increasingly refined searches to retrieve information that is known to exist. The right kind of analytics platform empowers the analyst to search for and consequently exploit insights that may be previously unknown, yet can be derived from visualizing and interacting with the information retrieved.



Differentiating Traditional Business Intelligence from Analytics

There is often confusion as to the difference between business analytics and business intelligence. Business intelligence (BI) has solidified as the conceptual framework for consolidating operational and transactional data in a stand-alone repository, with the data reconfigured in a way that addresses the numerous reports expected by the management teams. Management relies on the reporting and dashboard capabilities of a BI framework to compare the business results against expectations.

BUSINESS INTELLIGENCE

Business intelligence essentially looks backwards, designed to capture information and report on what has happened in the past. BI usually relies on the use of a data warehouse, with the critical dimensions selected as part of a serious system requirements analysis activity. As a result of this structured process, the information asset that is delivered via the warehouse reflects a selection of variables and dimensions designed to address the specific reporting goals.

Unfortunately, since BI is limited to the data items, events, and aggregations that are defined to meet those reporting needs, the internal schemas and structures, and the data sets to be collected must be pre-considered before any analysis is done. The result of pre-selection and aggregation of data limits the body of information available for analysis. Data warehouse architectures, once solidified, are essentially static and are hard to change.

BUSINESS ANALYTICS

The focus of analytics, however, is seeking knowledge and insight that can be derived from the many ways that data can be brought together. As opposed to the confined information aspects of data warehousing, business analytics are not bound to predefined structures, models, or reports, and instead can be based on any available data sets. Analytics platforms provide an interactive interface allowing information to be presented using different visualization techniques that allows for the type of dynamic interaction that facilitates not only iterative problem-solving methods, but also the means by which the value of discovered insights can be communicated to other in the organization.



Communicating Concepts through Presentation

To highlight these differences, consider the ways that information is presented and that knowledge is conveyed. The presentation of standard BI reports is often limited to rectangular data grids or pivot tables, with supporting data reflected through a catalog of visual “widgets.” Dashboards and scorecards are configured using visualization objects to reduce the information that is being conveyed into a summary of what is being reported, providing the analyst with the means to drill through the graphic to gain a broader view of what is being reported. Typical dashboard graphics become paradigms for directing the user’s focus to the predefined dimensions of aggregation, enabling the search for greater detail but not necessarily any greater insight.

Alternatively, analytics platforms offer flexibility in information presentation, offering the same kind of dashboard-style reporting provided by BI platforms while allowing for visual drivers that answer the questions already posed, but also set the stage for more effective visual cues for finding those insights you aren’t already seeking. Because of the greater flexibility in incorporating data from different sources, there is a greater ability to dynamically adjust and drill through views. User actions such as modifying the set of variables being displayed and linking sets of graphics together provides visual interaction for undirected analysis, exposing trends, outliers, or unexpected relationships in the data through the iterative process. Incorporating analysis techniques into the interface provides an additional advantage in that it eliminates the need for technical expertise in order to perform the business analysis – in effect, equipping anyone in the enterprise with essential analytics capabilities.

Practical Objectives for Business Analytics

Making the ability to search for insight broadly available across the organization empowers your staff to take on the role of analysts. And when analysts are no longer bound by historical constraints that limit usability, the degree of creativity in developing models can explode. In contrast to the confines of the standard desktop utilities, modern business analytics platforms are extensible, allowing people to go above and beyond what can be achieved out of the box by building specific tools for analysis within their own business domains. Instead of viewing the analysis software as a “one-size fits all” tool, agile analysts are able to adapt the tools’ capabilities within a familiar business domain, thereby customizing the analytics based on their organization’s own objectives.

Here we look at some of the practical objectives for achieving this level of agility. Those goals include:

- Shortening the iteration cycle
- Breaking the dependency on the data warehouse
- Understanding the data
- Adapting your information resources
- Exploiting visualization for interaction
- Communicating the results in a repeatable way



SHORTENING THE ITERATION CYCLE

Developing analytics models is an iterative process, driven by the analyst's agility in intellectually processing the answers to one set of questions and triggering ideas for the next set of questions. While the analyst's mind works quickly, often the production cycles for the analysis platforms do not, due to slow turn-around in system and data availability. For example, when driving analysis off a data warehouse, the process may be slowed as the analyst requests adjustments to the underlying data sets, changes to the model, delays in availability of newly integrated data sets, or modifications to the dimensions and aggregations.

Practically, the analyst benefits when the cycle time is decreased and the analyst's information needs are met quickly, either due to quick turn-around or based on cues from the visualization and presentation of results. With an analytics platform that enables rapid interaction and development, the time between iterations will decrease, reducing the time spent mastering the platform thereby increasing the time spent doing actual analysis.

BREAKING THE DEPENDENCY ON THE WAREHOUSE

There is often an assumption that because a data warehouse is intended to be the repository for organizational information for querying and reporting purposes, its data is the best and only candidate for business analytics. Unfortunately, this assumption is flawed for three reasons:

- The determination of the sets of data imported into the warehouse are probably based on the business requirements for operational and performance reporting. The facts and qualifiers are predetermined, as are any aggregations that are performed within the warehouse environment. Unfortunately, this predetermination of warehouse contents limits undirected analysis.
- Presuming that even if the data in the warehouse is for the most part suitable for business analytics, any additional data needs would need to be satisfied through the warehouse development life cycle process. This implies a relatively long turnaround time for integrating additional facts and dimensions, in opposition to the goal of a shortened iteration cycle.
- The ability to include alternate data sources in different formats is part of the process of searching for insight. This is the virtue of the iterative approach, and is analogous to the way that the results of a content search enable the analyst to adjust the parameters and questions being asked to reach the desired solution.

Depending on the data warehouse as the sole data source will curtail creativity. The pragmatic approach, then, is to employ a business analytics platform that can import data from any source without forcing the analyst to rely on what has already been packaged in the warehouse models.

UNDERSTANDING THE DATA

When the ability to develop analytical models was restricted to the individuals willing to become experts in the use of the available products, there was a premium placed on tools expertise. Yet



without the corresponding subject matter expertise, it is difficult to build analytical models. Clearly, there is great value in not just understanding how the organization's data sets support the business processes, but also in how acquired data sets can enhance the value of the analytic process.

Pragmatically, our expectation is that as the iterative investigations are performed, more speculative analysis is done, which will lead to a better understanding of the data and how it can be manipulated and reviewed. The result of the efforts of an analyst armed with a better understanding of the data is an increase in precision and accuracy in the design of customized analytic models.

ADAPTING YOUR INFORMATION RESOURCES

We have already suggested breaking the dependency on the data warehouse; the other side of the coin involves identifying and using data sets from different sources that lie outside the range of traditional data feeds. Tools that require complex denotations and documented metadata associated with information sources do not provide a full spectrum of value. Tools to support dynamic business analytics must enable the analyst to import data sets from various sources in different structures and formats as needed and provide the capability to integrate selected different data sets together.

For example to perform market research in preparation for a product launch, the analyst may require geo-demographic profiles describing the core constituency within the market area. Later, the same analyst may want to explore how well the company is doing in comparison to its competitors, requiring acquired competitor sales data. Perhaps sluggish sales are related to the ability to move product from the factories to distribution centers to the stores, and logistic network data can be used to improve distribution.

Make it a best practice within your teams to determine during the iterative process that additional data can add value. To optimize the process, reduce the dependency on internal resources for data preparation. Instead, employ a platform that provides the ability to bring available data directly into the model.

EXPLOITING VISUALIZATION FOR INTERACTION

The dependence on individuals with expertise in traditional analytics systems should not be an impediment to analysis. Different products may rely on different underlying approaches to guide the development of the models, but the tools that enable the communication of ideas and the ability to interact with the data using a variety of visual approaches facilitate the rapid iteration that shortens the cycle for incorporating new streams of thought into the process.

Better yet, seek out the products that enable the linkage of multiple visual views layered on top of the same data sets. With linked views, user interaction within one view (such as applying filters to the data or selecting data points within a specific graphic region) will correspondingly adjust within the linked graphic components. Enhanced embedded visual cues will simplify the interaction as well as "train" the user in ways the platform drives innovation.



MAKING ANALYTICS REPEATABLE AND PACKAGEABLE

It can be frustrating to follow many chains of thought, find the appropriate result, but then forget how that result was achieved. The iterative approaches employed in developing the models provide greater impact when they can be turned into repeatable processes. Repeatability is defined across a number of dimensions, including:

- The ability to resolve to an analytics model that can be adapted against the same (or similar) data set on a periodic (quarterly, monthly, weekly, daily) basis
- The ability to exchange analytics models between analysts
- The ability to share the thought processes and results of the analysis
- The ability to validate the analysis and the conclusions by handing the process itself off to another analyst for review and verification

Analytic models can be made more effective when they can be reusable for both continual analysis with different instances of the data sets and for packaging the “story” behind the analysis. Again, this means more than just being able to generate “slideware,” but involves a comprehensive capture of the data sources (both within and outside of the data warehouse), the processes for integration, the analytical approaches taken, and the thought processes through which the results were derived. It is analogous to the query string in our search analogy – the iterations not only provide the sought-after results, they also home in on a query string that can be applied on a continual basis against the ever-growing scope of available content.

Enabling everyone within the enterprise with these types of repeatability is a critical success factor, suggesting the selection of an analytics platform that can capture history as well as packaging models in a standardized way. Doing so frees the analytics team to extend the basic platform with value-added algorithms and approaches customized to the organization’s way of doing business.



Bringing it all Together

There clearly are differences between the traditional approach to BI reporting and monitoring, and the use of analytics. Yet these two activities, despite their differences, are mutually dependent. Business intelligence systems are used for presenting dashboards or scorecards of collective metrics and for presenting those scorecards in a way that allows business analysts to monitor performance. Analysts can then use those metrics to drive the use of analytics tools to seek out patterns of behavior or activity that can be exploited for competitive advantage. There is a co-dependence, in which the creative approaches to speculative assessment uncovers the variables, dimensions, metrics, and measurements that will be used in the future to gauge business productivity.

Essentially, the business intelligence systems provide the starting point for developing new analytics. At the same time, the results of analysis are likely to uncover issues associated with business performance that are suited for ongoing monitoring. In other words, the analytics process in turn feeds the business intelligence process by introducing new metrics that can be added to the operational performance indicators incorporated into business intelligence dashboards. Analysis software can then be evolved from being viewed as a “one-size fits all” tool to providing specialized analytics services, as analysts begin adapting the tools’ capabilities within the context of the specific business domain to enable the creation of customized analytics driven by the organization’s own goals and requirements.

Organizations are increasingly under pressure to differentiate themselves from their competitors and seize opportunities to gain some advantage. The use of business analytics can fuel that process, and we have looked at a number of practical guidelines when seeking the right kinds of analytics tools. Shortening the iteration cycle, breaking the dependency on the data warehouse, and understanding and adapting your information resources are pragmatic suggestions. These approaches are actualized in products that allow the exploitation of visual graphics in real time, with feedback and interaction that facilitates a rapid iterative process for zeroing in on both the right questions and their corresponding answers. Lastly, being able to package the model and share it with other organizational stakeholders eases the process of communicating the results in a effective way.



About The Author

David Loshin is the president of Knowledge Integrity, Inc, (www.knowledge-integrity.com), a consulting company focusing on customized information management solutions including information quality consulting and training, business intelligence, metadata, and data standards management. David is among Knowledge Integrity's recognized experts in information management, contributing to Intelligent Enterprise, writing a monthly column for DM Review between October 2002 and December 2005, is a quarterly featured columnist for the Data Administration Newsletter (www.tdan.com), and is the channel expert of the B-EYE-Network's Information Quality and Data Integration channel.



David's book, "Business Intelligence: The Savvy Manager's Guide" (June 2003) has been hailed as a resource allowing readers to "gain an understanding of business intelligence, business management disciplines, data warehousing, and how all of the pieces work together." David also wrote "Enterprise Knowledge Management – The Data Quality Approach," which describes a revolutionary strategy for defining, managing, and implementing business rules affecting Enterprise-wide information quality management.

In addition, David has created courses for The Data Warehousing Institute (www.tdwi.org), has presented at the annual DAMA/Meta Data conference, has taught tutorials on data quality at a number of venues, and as a representative of Knowledge Integrity is often called upon to provide insights and thought leadership to the Information Management community.



TIBCO Spotfire,
TIBCO Software Inc.
212 Elm Street
Somerville, MA 02144

Tel: +1 617 702 1600
Toll Free: +1 800 245 4211
Fax: +1 617 702 1700
<http://spotfire.tibco.com>