



# Learning the Language of Data Analytics is No Longer Optional

By Don Carpenter, MSAcc/CPA

**T**he article “Is Accounting a Stem Field? Why it Matters” in the March/April 2023 issue of *Today’s CPA* magazine considered the implications to the profession of STEM classification. Support for STEM categorization is supported by the increased emphasis on data analytics for informed decision making that CPAs now provide to the larger business community. This article explores some of the basics of this ever-growing part of the profession.

It is often said that accounting is the “language of business.” The ability to navigate financial statements and understand the impact of decisions on an operation’s financial results is critical to a successful business career. But if accounting is the king’s English, then the language of data analytics is quickly becoming the Esperanto of commerce. As the ability to access, manage and analyze information has increased exponentially, the mystique of the process has been buttressed by terminology that even the most seasoned professionals may find bewildering.

Becoming familiar with some of the basic phrases in the rapidly growing field is a good first step to capitalizing on the potential it brings to decision making. Let’s delve into some of the most common vocabulary with the goal of making the world of “big data” a little less intimidating.

To begin, there is often a misunderstanding of just what **data analytics**, **data science** or even **data mining** entails. All three terms refer to accessing large amounts of data and often at a very granular level. For instance, advanced computing technologies now allow even the largest organizations to amass data and analyze it at the transactional level even though millions of transactions occur regularly. This is reflected in the fact that the growth in data is no longer measured in gigabytes but in zettabytes (trillion gigabytes).

Simplistically, data mining can be thought of as the somewhat indiscriminate process of combing through data in an effort to identify a meaning, interpretation or pattern.

Data analytics or data science may also follow this exploratory approach but can be more deductive, attempting to support a hypothesis with the available data. Some purists go even further by making a distinction between data analytics and data science based on the qualifications of the individual performing the work. Data scientists generally have more advanced statistical and mathematical credentials when compared to data analysts.

The use of data analytics in decision making generally can be segregated into four categories: predictive modeling, descriptive modeling, diagnostic modeling and prescriptive modeling.

**Predictive modeling** uses incidents (often transactions) where an outcome is known to predict an unknown outcome. This is often applied in marketing and customer behavior. For example, analysis of customer buying patterns can help a retailer target certain items to customers who buy other products based on the historical purchasing patterns of its existing customer base.

**Descriptive modeling** uses a full range of data points to yield a better understanding of a key business variable, such as data points that provide a better understanding of the customer base.

**Diagnostic modeling** seeks to determine the “why” or “how” behind a fact pattern. This approach could be very useful in fraud detection for example. Comparative analysis of costs across various sites could be used to isolate outliers to aid in detecting billing irregularities.

Finally, **prescriptive modeling** seeks to analyze the available data to determine how an outcome might be influenced and by its nature, it is less well defined.

The objective of any data analysis effort will determine the appropriate data to use. Organizations maintain business-critical information deemed necessary for analysis in a repository known as a data warehouse or data lake. A data warehouse might contain such diverse information as general ledger downloads, human resource records and environmental reports. The process of data analysis will “slice and dice” this information to draw out the relevant data and produce meaningful interpretation.

A clickstream database records the web traffic or sites visited within a search engine and would be useful in a predictive modeling analysis of customer behavior. A text database is a database of unstructured, continuous text such as the narrative in earnings’ releases, financial

statement footnotes or even tweets that can be searched and sorted to determine frequency of phrases or other recurring word patterns. And the more common numerical database is a collection of values representing a given variable (i.e., revenue, headcount) that can be compared or correlated to inform business decisions.

Once the appropriate database has been developed or accessed, a query is the mechanism used to extract a subgroup from the data set. The population is the entire data set that would be relevant to a particular analytical project. This subgroup should not be confused with a typical sample used in audit procedures for example. The strength of data analytics is its ability to process large volumes of data.

If all transactions in a general ledger comprised a population, the query might extract all transactions that increase expense accounts by greater than \$1,000,000. This subset can then be manipulated and analyzed with computer logic often referred to as bots. These threads of computer logic or algorithms can be quite complex and might be thought of as “turbo-charged macros.”



Data visualization refers to the method used to communicate the product of an analytical project in pictorial format such as bar/pie charts or linear graphs to allow for quick interpretation. The goal of data visualization is to succinctly convey the results without losing the message in the details.

Speaking the language of data analytics is critical to unlocking its full potential for anyone wanting to understand and use it to make informed business decisions.

**About the Author:** Don Carpenter is clinical professor of accounting at Baylor University. Contact him at [Don\\_Carpenter@baylor.edu](mailto:Don_Carpenter@baylor.edu).