Glossary: BI Analytics Landscape

Welcome! This alphabetized glossary contains many terms used in this course. Understanding these terms is essential when working in the industry, participating in user groups, or participating in other certificate programs.

Estimated reading time: 7 minutes

Term	Definition
Anonymization	The process of modifying personal data so it is no longer possible to identify the individuals associated with it.
Authentication	The process of confirming the identity of a user or application seeking access to data.
Authorization	The process of determining whether an authenticated user has the right to perform an operation.
Business intelligence	A combination of various technologies, tools, and methodologies that gather, analyze, and transform data into meaningful information.
Business intelligence tools	Proprietary or open-source application software that is used to collect, process, analyze, sort, filter, and report large data from the systems and transform raw data into useful information.
Causal relationship	A type of relationship between two events where one event influences the occurrence of another event.
Central tendency	The statistical measure that represents the central value of the entire distribution or a data set.
Corporate performance management (CPM)	Software that integrates information from various sources to support operational plans and align key performance indicators or KPIs to improve financial planning.
Customer relationship management (CRM) software	Software that helps companies measure and control their lead generation and sales pipelines.
Data analysis	The process that involves cleaning, transforming, and modeling data to uncover useful information to aid business decision-making.
Data analytics	The process that focuses on extracting valuable information from data using various tools, techniques, processes, and algorithms. It includes data analysis and the interpretation of the results, keeping in mind specific business objectives.
Data encryption	The process of translating data from a readable to an unreadable format, also known as ciphertext, to protect sensitive information during transmission or storage.
Data governance	The management of data in enterprise systems that involves ensuring availability, usability, integrity, and security, all of which are controlled by internal data standards and policies.
Data integration	A combination of technical and business processes that combine data from disparate sources into meaningful and valuable information.
Data marts	Data warehouses have data marts or subsets, which are like a partition to manage one specific business function, department, or subject area. Data marts make specific data available to a defined group of users, which allows those users to quickly access critical insights without wasting time searching through an entire data warehouse.
Data mining	The process of uncovering patterns and other valuable information from large data sets.
Data modeling	A visual representation of either a whole information system or parts of it to communicate connections between data points and structures.
Data privacy	Protection of personal data from those who should not have access to it and the ability of individuals to determine who can access their personal information.
Data repositories	Data sets identified to be mined for reporting and analysis. It is also known as a data archive or library.
Data security	The practice of protecting digital information, including personal information, from unauthorized access, tampering, or theft throughout its entire lifecycle.
Data storytelling	Skills that involve analyzing data to communicate insights and influence decisions.
Data visualization	Graphical representation of information and data. It helps data visualization to understand trends, outliers, and patterns in data.
Data warehouse	A storage architecture that pulls data from many sources into a single data repository for sophisticated analytics and decision support.
Descriptive analysis	The process of utilizing statistical techniques to explain or summarize a specific set of data. Descriptive analysis is also called descriptive statistics.
Descriptive analytics	A type of analytics that uses historical and current data to spot any trends and relationships between that data.
Diagnostic analytics	A type of analytics that helps identify the reason an event occurred. It lets you discover hidden correlations and connections between variables, determine causal relationships, detect anomalies, and isolate patterns.
Enterprise resource planning (ERP) systems	A type of software system that enables businesses to automate and efficiently manage their key business processes to gain optimal performance.
Extract, load, transform (ETL) process	The process that extracts, loads, and transforms data from multiple sources to a data warehouse or other unified data repository.
Flat files	A collection of data that is stored specifically in a two-dimensional database. It usually contains a series of records (or lines), where each record is a sequence of fields.
Key performance indicators (KPI)	A performance measure for a specific objective that provides targets, milestones, and insights to help teams and individuals make better decisions.
Machine learning	The branch of artificial intelligence (AI) and computer science focuses on using data and algorithms to imitate how humans learn, gradually improving accuracy. It helps systems learn and improve at forecasting, much like how people learn from experience.
Mean	A measure of central tendency. The arithmetic average is the sum divided by the number of cases.
Median	The median is a measure of central tendency not sensitive to outlying values (unlike the mean, which can be affected by a few extremely high or low values).
Metadata management	Best-practice processes and technologies that manage the metadata of your data. Metadata management helps provide insights for more effective data management.
Metrics	Another measure to track the performance of specific business processes. They help provide context to the performance of key business goals but are not critical to its success like key performance indicators (KPIs).
Mode	The most frequently occurring value. If several values share the greatest occurrence frequency, each is a mode.
Online analytical processing (OLAP)	Software used to conduct multidimensional analysis on large volumes of data from a data warehouse, data mart, or other centralized data store.
Predictive analytics	A type of analytics used to predict future outcomes. It relies on historical data and employs various techniques such as statistical modeling, data mining, and machine learning for predictions.
Prescriptive analytics	A type of data analytics that recommends the optimal course of action to achieve a specific goal, drawing from inputs from descriptive, diagnostic, and predictive analytics processes.
Probability theory	The mathematical framework that can help you find the likelihood of specific outcomes.
Python	An agile, dynamically typed, expressive, open-source programming language that supports multiple programming philosophies, including procedural, object-oriented, and functional. Python is a popular high-level programming language that is easily extensible through third-party packages and often allows powerful functions to be written with few lines of code.
Real-time BI	Processes and tools that provide up-to-the-minute information and analysis, which enables businesses to monitor KPIs in real-time.
Regression analysis	Statistical methods can help find relationships between variables, such as how changes in independent variables may affect a dependent variable.
Return on investment (ROI)	A performance measure that is used to evaluate the efficiency of an investment.
Self-service business intelligence (SSBI)	The process, tools, or software that allows business users to ask questions about their data, get insights, and analyze data without relying on IT, BI specialists, or SQL to take business actions.
Standard deviation	The result of summing the squared deviations of the ratios about the mean, dividing the result by the total number of ratios minus one, and taking the positive square root.
Statistical analysis	The process of collecting large volumes of data and then using statistics and other data analysis techniques to identify trends, patterns, and insights.
Statistical modeling	A method that uses mathematical equations to analyze data and identify patterns.
Statistical tool	A tool that converts, analyzes, interprets, and uses data in different forms and purposes.
Structured query language (SQL)	A computer language which is used to interact with a relational database.
Time-series analysis	A way of analyzing data that can help you uncover patterns, trends, and anomalies in time-based data.

Author(s)

- Sangeeta Srinivasan