

INFORMS Certified Analytics Professional (CAP)

Glossary of Terms

Term	Domain Aligned to	Definition
5 Whys	Business Problem Framing	A problem-solving technique where one asks "Why?" five times to identify the root cause of an issue.
5S	Business Problem Framing	A workplace organization method involving five steps—Sort, Set in Order, Shine, Standardize, and Sustain.
80/20 Rule (Pareto Principle)	Business Problem Framing	A concept that 80% of effects come from 20% of causes.
Accuracy	Model Life Cycle Management	For classifications, accuracy is the percent correctly classified. For predictions, accuracy is the can be expressed in multiple ways, including mean absolute error, mean absolute percent error or mean square error.
Activity-Based Costing (ABC)	Business Problem Framing	A costing method that assigns costs to products based on the resources they consume.
Agent-Based Modeling	Methodology	A computational model for simulating actions of agents to assess their effects on a system.
Algorithm	Model Building	A step-by-step procedure for solving a problem or performing a computation.
Amortization	Business Problem Framing	The process of spreading out a loan or intangible asset cost into a set of payments using their net present value and a known interest rate over an agreed upon period of time.
Analytical Benefits	Analytics Problem Framing	The description, typically communicated in writing and in enough detail to be clearly understood, of how a business problem/question will be addressed using analytics.
Analytics	Analytics Problem Framing	The systematic process of discovering, interpreting, and communicating meaningful patterns and insights from data. It involves the use of statistical, mathematical, and computational techniques to analyze data, identify trends, make predictions, and support decision-making.
Analytics Professional	Model Life Cycle Management	An individual with skills in analyzing and interpreting data for decision-making.
ANCOVA (Analysis of Covariance)	Model Building	A statistical technique that combines ANOVA and regression.
ANOVA (Analysis of Variance)	Model Building	A statistical method for comparing means of different groups.
Artificial Intelligence (AI)	Methodology	The simulation of human intelligence in machines.
Artificial Neural Networks	Model Building	Computing systems inspired by biological neural networks for pattern recognition, natural language processing, speech recognition, and predictive analytics.
Assemble-to-Order	Deployment	A production strategy where products are assembled only after receiving customer orders.
Assignment Problem	Methodology	a mathematical optimization problem where the goal is to assign tasks to agents (or resources) in a way that minimizes cost or maximizes efficiency while ensuring each task is assigned to exactly one agent.
Automation	Deployment	The use of technology to perform tasks with minimal human intervention.
Average	Data	A general term for central tendency in a set of univariate data. Specific measures include the arithmetic mean, the median and the mode.
Batch Production	Deployment	Manufacturing method where products are produced in groups or batches.
Benchmark Data	Data	Reference data used to evaluate or compare system performance.
Benchmark Problems	Model Life Cycle Management	Standardized problems used for evaluating algorithms or methods.
Benchmarking	Business Problem Framing	Comparing performance metrics with best practices to improve efficiency.
Bias	Model Life Cycle Management	A systematic deviation from the true value in data or results.
Big Data	Data	Extremely large data sets that require advanced methods for analysis.
Box-and-Whisker Plot	Data	A graphical representation of data showing the distribution's minimum, quartiles, and maximum.
Branch-and-Bound	Methodology	an optimization algorithm used to solve integer programming. It systematically explores decision branches, pruning parts of the search space that cannot yield better solutions than the current best.
Business Analytics (BA)	Analytics Problem Framing	The process of using data and statistical methods to inform business decisions.
Business Benefits	Business Problem Framing	Advantages or value gained by the organization from a project, often measured in terms of revenue growth, cost savings, or market competitiveness.
Business Case	Business Problem Framing	A justification for a project or decision based on its expected benefits and costs.
Business Intelligence (BI)	Analytics Problem Framing	Technology-driven processes for analyzing data to inform business decisions.
Business Process Modeling (BPM)	Business Problem Framing	A method for representing the processes within an organization.
Business Process Owners	Business Problem Framing	Individuals responsible for managing and overseeing specific business processes to ensure they align with organizational goals.
Business Validation	Deployment	
Bystanders	Business Problem Framing	Individuals or groups indirectly affected by the outcomes of a project but not actively engaged in its processes.
Chief Analytics Officer (CAO)	Model Life Cycle Management	An executive responsible for data analysis and data-driven decision-making.
Chi-Squared Automated Interaction Detection (CHAID)	Model Building	A statistical technique used to create decision trees by splitting data into mutually exclusive groups based on the significance of chi-squared tests.
Classification	Model Building	The process of categorizing data into predefined classes.

**INFORMS Certified Analytics Professional (CAP)
Glossary of Terms**

Term	Domain Aligned to	Definition
Classification Problem	Model Building	A type of problem where the goal is to categorize data into specific groups.
Cleansing	Data	The process of identifying and correcting inaccurate data.
Clustering	Model Building	Grouping a set of objects into clusters based on similarity.
Combinatorial Optimization	Methodology	A field of optimization focused on finding the best solution from a finite set of possible solutions, often for problems involving discrete variables.
Confidence Interval	Data	A range of values that is likely to contain the population parameter.
Confidence Level	Data	The probability that a confidence interval contains the population parameter.
Confusion Matrix	Model Life Cycle Management	A performance evaluation tool used in classification problems to summarize the accuracy of a model by comparing predicted and actual outcomes.
Conjoint Analysis	Methodology	A statistical technique used to determine how consumers value different attributes.
Constraint (Mathematical)	Model Building	A restriction in optimization problems.
Constraint Programming	Methodology	A computational method used to solve complex combinatorial problems by defining constraints that must be satisfied and systematically exploring possible solutions.
Continuous Production	Deployment	A production method where products are manufactured continuously.
Correlation	Data	A measure of the association, both positive or negative, between two variables.
Cost of Capital	Business Problem Framing	The cost of funds used for financing a business.
Cube	Data	In data warehousing, a multi-dimensional array of data.
Cumulative Density Function (CDF)	Data	A function that shows the probability that a variable takes a value less than or equal to a certain level.
Customers	Business Problem Framing	End users or recipients of a product or service, often considered a central focus in decision-making and analytics projects.
Cutting Stock Problem	Methodology	An optimization problem that minimizes waste when cutting raw materials.
Data	Data	Factual information used as a basis for reasoning, discussion, or calculation.
Data Context	Data	Understanding the environment, purpose, and conditions surrounding data, ensuring its relevance and proper use.
Data Custodians	Data	Technical personnel responsible for storing, managing, and protecting data assets.
Data Mining	Data	The process of discovering patterns in large data sets.
Data Owner	Data	The person or entity responsible for the integrity, accuracy, and security of a dataset.
Data Stewards	Data	Individuals tasked with implementing and enforcing data governance policies to maintain data quality and compliance.
Data Warehouse	Data	A system for storing, retrieving, and managing large amounts of data.
Database	Data	An organized collection of structured information or data.
Decision Tree	Model Building	A model for decision-making that uses a tree-like structure.
Decision Variables	Model Building	Variables in optimization models that influence outcomes.
Descriptive Analytics	Analytics Problem Framing	Analysis focused on describing past data trends.
Descriptive Analytics	Analytics Problem Framing	Methods used to summarize historical data and identify patterns, trends, and insights.
Design of Experiments (DOE)	Methodology	A method for systematically testing factors and variables.
Diagnostic	Analytics Problem Framing	Analytics Techniques aimed at identifying the causes or factors behind observed outcomes or patterns.
Discrete Event Simulation	Methodology	A simulation method for modeling the operation of systems as sequences of events.
Dynamic Programming	Methodology	An optimization technique breaking down problems into simpler subproblems.
Effective Domain	Data	Refers to the subset of a problem's total domain in which a solution is feasible or relevant. It is the range of values or conditions under which a model, system, or process operates effectively and produces valid results.
Efficiency	Business Problem Framing	Achieving maximum productivity with minimum wasted effort or expense.
Engagement	Business Problem Framing	The involvement and participation of individuals in activities.
Enterprise Resource Planning (ERP)	Deployment	Integrated management of core business processes.
ETL (Extract, Transform, Load)	Data	A data integration process that involves extracting data from multiple sources, transforming it into a usable format, and loading it into a target data storage system, such as a data warehouse.
Experimental Design	Methodology	The plan for assigning treatments to units in a study.
Expert Systems	Methodology	Computer systems that emulate decision-making ability of human experts.
Factor Analysis	Model Building	A statistical method used to identify underlying relationships among observed variables by grouping them into a smaller number of factors or dimensions.
Failure Mode and Effects Analysis (FMEA)	Model Life Cycle Management	A systematic method for evaluating potential failures in a process.
Fixed Cost	Business Problem Framing	Costs that do not change with the level of production or sales.
Forecasting	Model Building	Predicting future events based on historical data.

INFORMS Certified Analytics Professional (CAP)

Glossary of Terms

Term	Domain Aligned to	Definition
Fuzzy Logic	Methodology	A method of reasoning that resembles human reasoning with inexact data.
Game Theory	Methodology	A mathematical framework for analyzing strategic interactions where the outcome for each participant depends not only on their own decisions but also on the choices made by others.
Genetic Algorithms	Methodology	Search heuristics inspired by the process of natural selection.
Global Optimal	Model Building	The best possible solution among all feasible solutions in optimization.
Goodness of Fit	Model Life Cycle Management	A measure of how well data fit a statistical model.
Graphical User Interface (GUI)	Deployment	A user interface that includes graphical elements for interaction.
Greedy Heuristics	Methodology	An approach to solving optimization problems by making a series of choices, each of which looks best at the moment, with the hope of finding the global optimum.
Heuristic	Methodology	A problem-solving strategy or approach that simplifies decision-making by using practical methods or rules of thumb to find a solution that is good enough, especially when finding an optimal solution is too time-consuming or difficult.
Histogram	Data	A graphical representation of the distribution of numerical data using bars.
Holdout Data	Model Life Cycle Management	A subset of data used to test a model's performance after training.
Hyperparameters	Model Building	Configuration variables set before training a model, affecting its performance.
Hypothesis Testing	Model Building	A method of making statistical inferences about a population based on sample data.
Impact Analysis	Deployment	The evaluation of potential or actual effects of changes, decisions, or projects on processes, systems, or outcomes.
Influence Diagrams	Analytics Problem Framing	Visual representations showing relationships between variables, decisions, and outcomes to assist in decision-making processes.
INFORMS	Analytics Problem Framing	The Institute for Operations Research and the Management Sciences, a professional society for analytics and OR.
Innovative Applications in Analytics Award	Analytics Problem Framing	An INFORMS award recognizing impactful uses of analytics.
Integer Program	Methodology	An optimization model where some or all variables are restricted to integer values.
Integrity	Model Life Cycle Management	The accuracy, consistency, and reliability of data throughout its lifecycle.
Internal Rate of Return (IRR)	Business Problem Framing	The discount rate that makes the net present value of cash flows equal to zero.
KDD (Knowledge Discovery in Databases)	Data	The process of discovering useful knowledge from a collection of data.
Knapsack Problem	Methodology	Involves selecting a subset of items, each with a given weight and value, to maximize the total value of the selected items without exceeding a specific weight limit (the "capacity" of the knapsack).
Lead Time	Deployment	The time between initiating and completing a process.
Lean Production	Business Problem Framing	A production method aimed at reducing waste and increasing efficiency.
Lift Curve	Model Life Cycle Management	A graphical tool showing the performance of a classification model.
Lineage	Data	A record of data origins, transformations, and movement across systems to ensure its integrity and reliability.
Linear Program	Methodology	An optimization model with a linear objective function and linear constraints.
Little's Law	Methodology	A formula relating the average number of items in a system, arrival rate, and processing time.
Local Optimal	Model Building	The best solution within a limited area of the solution space.
Logistic Regression	Model Building	A statistical model used for predicting binary outcomes.
Machine Learning	Methodology	A field of AI that uses algorithms to learn patterns from data for predictions or decisions.
MANOVA (Multivariate Analysis of Variance)	Model Building	A statistical test for comparing multiple dependent variables across groups.
Mean	Data	A measure of central tendency for a distribution of data / random variable that involves giving weight to every value before aggregating. Means could be arithmetic or geometric.
Mean Squared Error (MSE)	Model Life Cycle Management	A measure of prediction accuracy in a model, averaging the squared differences between actual and predicted values.
Mean Time Between Failures (MTBF)	Model Life Cycle Management	The average time between system breakdowns or failures.
Median	Data	The middle value of a dataset when values are ordered from smallest to largest.
Merge/Join	Data	Combining datasets or tables based on shared keys to enrich data or create comprehensive datasets for analysis.
Metaheuristics	Methodology	High-level procedures guiding other heuristics for solving complex optimization problems.
Method = Plan	Methodology	A detailed outline or plan for implementing a specific solution or achieving an objective.
Methodology = Approach	Methodology	A structured framework or strategy used to address a problem or execute a project.
Mode (of a Probability Distribution)	Data	The value with the highest probability in a distribution.
Mode (of Variable Data)	Data	The value that appears most frequently in a dataset.
Monte Carlo Simulation	Methodology	A computational technique that uses random sampling to approximate complex calculations.
Navigating	Business Problem Framing	The process of identifying and following pathways to address challenges or achieve project objectives.
Negotiating	Business Problem Framing	The act of reaching agreements between parties to address conflicting interests or goals within a project.

INFORMS Certified Analytics Professional (CAP)

Glossary of Terms

Term	Domain Aligned to	Definition
Net Present Value (NPV)	Business Problem Framing	The present value of cash flows over time, discounted at a specific rate.
Network Optimization	Methodology	Optimization techniques applied to networks to improve flow, routing, or design.
Next Best Offer (NBO)	Model Building	The optimal offer tailored to a customer based on predictive analytics.
Nominal Group Technique (NGT)	Business Problem Framing	A group decision-making process used to generate and prioritize ideas.
Normalization (Database)	Data	The process of organizing data to reduce redundancy and dependency.
Normalized Data	Data	Data that has been adjusted in its scale or distribution to bring it into a consistent or standard range, for example by adjusting min/max values or adjusting by z-scores.
Objective Function	Model Building	A mathematical expression that defines the goal of an optimization problem.
OLAP (Online Analytical Processing)	Data	A technology that allows for fast analysis of multidimensional data.
OLAP Cube	Data	A data structure that allows data to be analyzed in multiple dimensions.
Operations Management	Deployment	The management of production and business operations.
Operations Research (OR)	Methodology	A discipline applying analytical methods to decision-making.
Opportunity Cost	Business Problem Framing	The loss of potential gain from other alternatives when one option is chosen.
Optimization	Methodology	The process of making something as effective or functional as possible.
Parameters (Model)	Model Building	The numerical factors that define a model's behavior and outcomes replace factors with coefficients
Pareto Concept	Business Problem Framing	The idea that a small number of causes are responsible for a large portion of outcomes (80/20 rule).
Pattern Recognition	Model Building	The identification of patterns in data.
Payback	Business Problem Framing	The time required to recover the cost of an investment.
Pie Chart	Data	A circular chart divided into sectors representing proportions of a whole.
Precision	Model Life Cycle Management	Precision in a classification problem is a metric used to measure the accuracy of positive predictions made by a model. Specifically, it is the ratio of true positive predictions to the total number of positive predictions (both true and false positives).
Predictive Analytics	Analytics Problem Framing	The use of statistical and machine learning models to forecast future outcomes or trends based on historical data.
Predictive Analytics	Model Building	Techniques for analyzing current and historical data to make predictions about future outcomes.
Prescriptive Analytics	Model Building	Analysis that suggests decision options to achieve desired outcomes.
Prescriptive Analytics	Business Problem Framing	Advanced analytics methods that provide recommendations or actions to achieve desired outcomes.
Pricing	Business Problem Framing	The process of determining the value at which a product or service will be sold.
Primary Measures of Success	Business Problem Framing	Specific metrics or criteria used to evaluate the effectiveness or achievement of project goals.
Primary Stakeholder	Business Problem Framing	The individual or group that has the most significant influence on or investment in the success of a project or decision-making process.
Principal Component Analysis (PCA)	Model Building	A technique for reducing data dimensionality by identifying principal components.
Probability Density Function (PDF)	Data	A function showing the likelihood of a continuous random variable taking certain values.
Problem Assessment/Framing	Business Problem Framing	The initial step in problem-solving, defining and understanding the problem.
Project Management	Deployment	The process of leading and organizing resources to achieve project goals.
Proprietary Data	Data	Data owned by a company or individual, typically not freely shared.
Queuing Theory	Methodology	The mathematical study of waiting lines.
Random	Data	Occurring by chance without a predictable pattern.
Range	Data	The difference between the highest and lowest values in a dataset.
Recoding	Data	The process of transforming or replacing values in a dataset to improve clarity, consistency, or usability.
Regression (Models)	Model Building	statistical techniques used to examine the relationship between one or more independent variables (predictors) and a dependent variable (outcome).
Regression Analysis	Model Building	A statistical technique used to understand the relationship between one dependent (target) variable and one or more independent (predictor) variables. The goal is to model this relationship so that the dependent variable can be predicted from the independent variables.
Regression Problem	Model Building	A type of predictive modeling problem where the goal is to predict a continuous (numeric) outcome or dependent variable based on one or more independent variables (predictors).
Response Surface Methodology (RSM)	Methodology	A method to optimize processes with a response influenced by several variables.
Return on Investment (ROI)	Business Problem Framing	A measure of profitability, calculated as net profit divided by investment cost.
Revenue Management	Business Problem Framing	The practice of pricing and capacity allocation to maximize revenue.

**INFORMS Certified Analytics Professional (CAP)
Glossary of Terms**

Term	Domain Aligned to	Definition
RFM (Recency, Frequency, Monetary Value)	Analytics Problem Framing	A marketing analysis technique used to segment customers based on their purchasing behavior. It is widely used to identify the most valuable customers, understand customer loyalty, and target specific segments with personalized marketing strategies.
Risk	Business Problem Framing	The potential for loss or adverse effects on an outcome.
Robust Optimization	Model Building	An optimization methodology used to handle uncertainty in mathematical models, particularly when certain parameters are unknown or subject to variability. Unlike traditional optimization methods that assume perfect knowledge of all parameters, robust optimization seeks to find solutions that perform well under a range of possible scenarios, especially when some of the data or conditions are uncertain.
Scatter Plot	Data	A type of data visualization that displays the relationship between two continuous variables. It uses dots to represent individual data points on a two-dimensional graph, where the horizontal axis (x-axis) represents one variable and the vertical axis (y-axis) represents another.
Scenario Analysis	Business Problem Framing	A process of analyzing potential future events by considering alternative possible outcomes.
Scheduling	Deployment	Planning the timing of tasks and resources in a process.
Sensitivity Analysis	Model Life Cycle Management	An analysis to determine how changes in variables affect outcomes.
Shadow Price	Business Problem Framing	A concept in optimization and linear programming that represents the change in the objective function (e.g., profit, cost, or utility) resulting from a one-unit increase in the availability of a resource, given the constraints of the problem. It indicates the value of relaxing or tightening a constraint in a linear optimization model. It applies to both linear and nonlinear models
Simulated Annealing	Methodology	An optimization technique that mimics the annealing process in metallurgy.
Six Sigma	Model Life Cycle Management	A set of techniques aimed at improving quality by reducing defects.
Skewness	Data	Advanced analytics methods that use data, statistical models, machine learning, and optimization techniques to recommend specific actions or strategies to achieve desired outcomes. Prescriptive analytics not only predicts future scenarios but also suggests decision options and their potential impact, enabling informed decision-making.
Societal Benefits	Business Problem Framing	Positive impacts on society resulting from a project, such as increased access to services, environmental improvements, or economic development.
Spreadsheet Analysis	Data	The use of spreadsheet software to analyze data.
Stakeholders	Business Problem Framing	Individuals or groups with an interest in the outcomes of a project, including customers, employees, investors, and external partners.
Standard Deviation	Data	A measurement of dispersion of a random variable or data that measures variability by computing the square root of the average squared distance from the mean (square root of the variance).
Stated Drivers	Business Problem Framing	Key factors or motivations explicitly identified as influencing a decision, strategy, or analysis.
Statistical Significance	Model Life Cycle Management	The probability that an observed effect is not due to chance.
Statistics	Data	The study of data collection, analysis, interpretation, presentation, and organization.
Stepwise Regression	Model Building	A method of fitting regression models by adding or removing predictors.
Supervised Models	Model Building	Machine learning models trained on labeled data, where input-output relationships are explicitly provided for prediction.
Supply Chain Management (SCM)	Deployment	Managing the flow of goods and services from production to delivery.
System Dynamics	Methodology	A modeling approach for understanding complex systems over time.
Technique = Procedure	Methodology	A defined set of steps or methods used to carry out a task or achieve a specific result.
Testing Data	Model Building	Data used to evaluate the performance of a model.
Tolerance	Model Life Cycle Management	Acceptable deviation in measurements or conditions.
Traceability	Data	The ability to track and verify the history, origin, and changes made to data throughout its lifecycle.
Training Data	Model Building	Data used to teach a machine learning model.
Traveling Salesman Problem (TSP)	Methodology	A classic optimization problem in which the goal is to find the shortest route visiting a set of locations once.
Uncertainty	Business Problem Framing	The lack of complete certainty about outcomes or data.
Univariate	Model Life Cycle Management	A type of analysis that focuses on understanding and summarizing a single variable or dataset.
Unsupervised Models	Model Building	Machine learning models that analyze and identify patterns in data without pre-labeled outputs.
Validation (of a Model)	Model Life Cycle Management	The process of confirming that a model meets its intended purpose.
Validation Data	Model Life Cycle Management	Data used to validate the accuracy of a model's predictions.
Variability	Data	The extent to which data points differ from each other.
Variable Cost	Business Problem Framing	Costs that vary with production levels.
Variance	Data	Measure of the dispersion of a set of values.

INFORMS Certified Analytics Professional (CAP)
Glossary of Terms

Term	Domain Aligned to	Definition
Variation Reduction	Model Life Cycle Management	Strategies aimed at reducing variability in processes.
Vehicle Routing Problem (VRP)	Methodology	An optimization problem to determine the most efficient routes for vehicles.
Verification (of a Model)	Model Life Cycle Management	Checking that a model is implemented correctly.
Web Analytics	Analytics Problem Framing	The measurement and analysis of web data to understand and optimize web usage.
Yield	Business Problem Framing	The output or results from a process relative to input.
Structure of the Data	Data	The organized format or arrangement of data, which defines how data is stored, accessed, and processed. The structure determines the relationship between different data elements and how they can be efficiently analyzed or manipulated.