

# Are you fluent in Geek?

*Is your business recruiting a data scientist or are you working on building your purple unicorn? Getting started can be daunting and it may seem like you need to learn a foreign language. Our data analytics glossary will help you understand some key terms and help you speak Geek!*

<b>Big Data</b>	Datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze.
<b>Business Intelligence</b>	The general term used for the identification, extraction, and analysis of data.
<b>Cloud</b>	A broad term that refers to any Internet-based application or service that is hosted remotely.
<b>Data Analyst</b>	A person responsible for the tasks of modeling, preparing and cleaning data for the purpose of deriving actionable information from it.
<b>Data Engineering</b>	The collection and storage of data which allows for batch or real-time processing for data scientists to query.
<b>Data Governance</b>	A set of processes or rules that ensure data integrity and that data management best practices are met.
<b>Data Mining</b>	Finding meaningful patterns and deriving insights in large sets of data using sophisticated pattern recognition techniques. To derive meaningful patterns, data miners use statistics, machine learning algorithms and artificial intelligence.
<b>Data Scientist</b>	Someone who can make sense of big data by extracting raw data, massaging it and come up with insights. Skills needed are statistics, computer science, creativity, story-telling and understanding business context.
<b>IoT (Internet of Things)</b>	The network of physical objects or “things” embedded with electronics, software, sensors and connectivity to enable it to achieve greater value and service by exchanging data with the manufacturer, operator and/or other connected devices. Each thing is uniquely identifiable through its embedded computing system but is able to interoperate within the existing Internet infrastructure.
<b>Predictive Modeling</b>	The process of developing a model that will most likely predict a trend or outcome.
<b>R</b>	An open source programming language used for statistical computing and graphics. It is a GNU project which is similar to the S language. R provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques, and is highly extensible. It is one of the most popular languages in data science.
<b>Structured Data</b>	Data that is organized according to a predetermined structure (rows and columns, as an example).
<b>Transactional Data</b>	Data that relates to the conducting of business, such as accounts payable and receivable data or product shipments data.
<b>Unstructured Data</b>	Data that has no identifiable structure, i.e. email message text, social media posts, audio files such as recorded human speech, music, etc.