Glossary: What Do Data Scientists Do?

Welcome! This alphabetized glossary contains many of the terms in this course. These terms are important for you to recognize when working in the industry, participating in user groups, and participating in other certificate programs.

Term	Definition	Video where the term is introduced
Comma-separated values (CSV) / Tab-separated values (TSV)	Commonly used format for storing tabular data as plain text where either the comma or the tab separates each value.	Working on Different File Formats
Data file types	A computer file configuration is designed to store data in a specific way.	Working on Different File Formats
Data format	How data is encoded so it can be stored within a data file type.	Working on Different File Formats
Data visualization	A visual way, such as a graph, of representing data in a readily understandable way makes it easier to see trends in the data.	Data Science Topics and Algorithms
Delimited text file	A plain text file where a specific character separates the data values.	Working on Different File Formats
Extensible Markup Language (XML)	A language designed to structure, store, and enable data exchange between various technologies.	Working on Different File Formats
Hadoop	An open-source framework designed to store and process large datasets across clusters of computers.	What Makes Someone a Data Scientist
JavaScript Object Notation (JSON)	A data format compatible with various programming languages for two applications to exchange structured data.	Working on Different File Formats
Jupyter notebooks	A computational environment that allows users to create and share documents containing code, equations, visualizations, and explanatory text. See Python notebooks.	Data Science Skills & Big Data
Nearest neighbor	A machine learning algorithm that predicts a target variable based on its similarity to other values in the dataset.	Working on Different File Formats
Neural networks	A computational model used in deep learning that mimics the structure and functioning of the human brain's neural pathways. It takes an input, processes it using previous learning, and produces an output.	A Day in the Life of a Data Scientist
Pandas	An open-source Python library that provides tools for working with structured data is often used for data manipulation and analysis.	Data Science Skills & Big Data
Python notebooks	Also known as a “Jupyter” notebook, this computational environment allows users to create and share documents containing code, equations, visualizations, and explanatory text.	Data Science Skills & Big Data
R	An open-source programming language used for statistical computing, data analysis, and data visualization.	Data Science Skills & Big Data
Recommendation engine	A computer program that analyzes user input, such as behaviors or preferences, and makes personalized recommendations based on that analysis.	A Day in the Life of a Data Scientist
Regression	A statistical model that shows a relationship between one or more predictor variables with a response variable.	Data Science Topics and Algorithms
Tabular data	Data that is organized into rows and columns.	A Day in the Life of a Data Scientist
XLSX	The Microsoft Excel spreadsheet file format.	Working on Different File Formats

Skills Network