Data

A collection of raw facts and figures. It can be in the form of numbers, text, images, or any other format.

Clean data

Clean data is data that is accurate, complete, and ready for analysis. Data cleaning, an important step in the data analysis process, involves checking your data for inaccuracies, inconsistencies, irregularities, and biases.

Open data

Open data, also called public data, is data that is available for anyone to use. Exploring and analyzing open datasets is one way to practice data analysis skills.

Qualitative data

Qualitative data is data that describes qualities or characteristics. It's generally non-numeric data and can be subjective, for example eye color or emotions.

Quantitative data

Quantitative data is objective data with a specific numeric value. It's generally something you can count or measure, such as height or speed.

Structured Data

Structured data is formatted data, for example data that is organized into rows and columns. Structured data is more readily analyzed than unstructured data because of its tidy formatting.

Unstructured data

Unstructured data is data that is not organized in any apparent way. In order to analyze unstructured data, you'll typically need to implement some type of organization.

Database

A database is an organized collection of information that can be searched, sorted, and updated. This data is often stored electronically in a computer system called a database management system (DBMS). Oftentimes, you'll need to use a programming language, such as structured query language (SQL), to interact with your database.

Structured Query Language (SQL)

Structured Query Language, or SQL (pronounced "sequel"), is a computer programming language used to manage relational databases. It's among the most common languages for database management.

Data science

Data science is the scientific study of data. Data scientists ask questions and find ways to answer those questions with data. They may work on capturing data, transforming raw data into a usable form, analyzing data, and creating predictive models.

Data visualization

Data visualization is the representation of information and data using charts, graphs, maps, and other visual tools. With strong data visualizations, you can foster storytelling, make your data accessible to a wider audience, identify patterns and relationships, and explore your data further.

Data Analytics

is the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision making. Data analytics encompasses data analysis (the process of deriving information

from data), data science (using data to theorize and forecast) and data engineering (building data systems). Data analysts, data scientists, and data engineers are all data analytics professionals.

Dashboard

Data analysts use a data visualization tool, commonly known as a dashboard, to convert all the data they receive into charts and graphs.

It’s essentially their control room, and they’ve probably spent many painstaking hours constructing this data hub. Make sure to tell them it looks great.

Attribute

When working in a spreadsheet or database, an attribute is a common descriptor used to label a column. Labeling columns clearly and precisely can enable you to keep your data organized and ready for analysis.

Data integrity

Data integrity is the maintenance and protection of data over its entire lifecycle. It relates to security, backing up and removing duplicate data

Mapping

Mapping matches fields from different data sources so that data can be moved successfully. What appears in one field, for example a telephone number, needs to be accurately replicated in the corresponding destination field.

Data mining

is the practice of systematically analyzing large datasets to generate insightful information, uncover hidden correlations, and identify patterns.

ETL

ETL is the process used to copy, combine, and convert data from different sources and formats and load it into a new destination such as a data warehouse or data lake.

Data Warehouse:

A large, centralized repository of data from different sources that is used for reporting and analysis. It allows for efficient querying and reporting

Predictive analytics

Predictive analytics is a type of data analysis that uses statistics, data science, machine learning, and other techniques to predict what will happen in the future.

Prescriptive analytics

Prescriptive analytics answers the question “What should/can be done?” by using machine learning, graph analysis, simulation, heuristics, and other methods.