

# WHAT DOES IT ALL MEAN???

## Your Guide to Common Data Science & Computer Science Terms

**Data** is plural- the singular unit of data is datum. Data are facts and statistics that are collected together for a reason, it could be for others to reference or for analysis and problem-solving. Data can be organized into tables, lists, charts, and graphs.

**A variable** is something that you measure- some examples are temperature, location, quantity, quality, color, size, or weight.

**An independent variable** is not changed by other variables you are trying to measure in an experiment. For example, if you want to measure housing prices in different zip codes, the independent variable would be the zip code- it doesn't change.

**A dependent variable** is affected by other factors- in our housing prices example, the price of the house would be dependent on a number of things, the size, age, condition, and location of the house.

**The x-axis** on a standard graph is usually horizontal and represents the independent variable.

**The y-axis** on a standard graph is usually vertical and represents the dependent variable.

**Features** are what data scientists call the columns in a dataset. A feature is a unique measurable property or characteristic. Features can be qualitative or quantitative.

**Observations** are what data scientists call the rows in a dataset. Each observation represents the measurement of a corresponding feature.

**Quantitative data** is measuring the quantity, or amount of something. It can be counted or compared on a numeric scale. For example, how many students attended EASTCon22?

**Qualitative data** is data describing attributes or qualities, something you can see or feel. The properties can be categorized and may be assigned numeric values to help organize it. For example, what was your favorite part of EASTCon22?

**A bit is a binary digit**- it is the smallest unit of information. A bit can only be one of two possible values and is commonly represented as 0 or 1, but could also be represented as on or off, yes or no, true or false.

**A byte** is a string of 8 bits, which is the number of bits it takes to encode a single letter or number in a computer. A byte is the smallest unit of memory in a computer.

**Nominal data** can be organized into categories that do not have a natural order. For example, colors (blue, green, yellow, red) or types of fruit (oranges, apples, bananas, pears).

**Ordinal data** can be organized into categories that do have a natural order. For example, school grade level- 5th grade, 6th grade, 7th grade, 8th, 9th, 10th, and so on. Another example is a Likert scale- a point scale used by researchers to take surveys and get people's opinion on a subject matter. How satisfied are you with this definition? You could choose very dissatisfied (1), slightly dissatisfied (2), indifferent (3), satisfied (4), or very satisfied (5).

**Numerical data** refers to the data that is in the form of numbers that can't be described with language or descriptive form.

**Data science** is solving problems with data. It involves collecting data, organizing it, analyzing it, and developing solutions.

**Computer engineering** is the science of building computers- the physical electronic parts, and how they all work together.

**Computer science** is solving problems with computers- understanding how they work and teaching them to do things for us, like writing software or programming algorithms.

**Algorithms** are complicated steps of instructions describing how to complete a task, such as solving a problem. Use the Sort it Out game as an example. When you examine the objects together, you will think and make a decision about how the objects should be sorted based on their attributes. You could sort them by color, material, or object type. Then you move the objects into the appropriate basket based on their attributes. If you wrote down the sorting instructions for someone else, you have written an algorithm. You could also teach a computer how to sort them.

**Artificial intelligence** is the intelligence of machines, as opposed to natural intelligence of animals and humans. Intelligence means the ability to learn and apply knowledge. The most common application of artificial intelligence is called machine learning, and many people use the terms synonymously or interchangeably.

**Machine learning** is the study and design of computer algorithms that can improve themselves automatically over several rounds of analysis, also known as epochs or iterations, by the use of data. Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so.