STRUCTURED DATA,UNSTRUCTURED DATA ,TIME SERIES DATA

STRUCTURED DATA

Structured data is most often categorized as quantitative data, and it's the type of data most of us are used to working with. Think of data that fits neatly within fixed fields and columns in [relational databases](https://www.g2crowd.com/categories/relational-databases) and spreadsheets.

Examples of structured data include names, dates, addresses, credit card numbers, stock information, geolocation, and more.

Structured data is highly organized and easily understood by machine language. Those working within relational databases can input, search, and manipulate structured data relatively quickly. This is the most attractive feature of structured data.

The programming language used for managing structured data is called structured query language, also known as SQL. This language was developed by IBM in the early 1970s and is particularly useful for handling relationships in databases.

If it sounds confusing, the picture below should help visualize how structured data relates to each other within a database.

UNSTRUCTURED DATA

Unstructured data is most often categorized as qualitative data, and it cannot be processed and analyzed using conventional tools and methods.

Examples of unstructured data include text, video, audio, mobile activity, social media activity, satellite imagery, surveillance imagery – the list goes on and on.

Unstructured data is difficult to deconstruct because it has no pre-defined model, meaning it cannot be organized in relational databases. Instead, non-relational, or [NoSQL databases](https://www.g2crowd.com/categories/nosql-databases), are best fit for managing unstructured data.

Another way to manage unstructured data is to have it flow into a [data lake](https://learn.g2crowd.com/what-is-a-data-lake), allowing it to be in its raw, unstructured format.

**More than 80 percent of all data generated today is considered unstructured, and this number will continue to rise with the prominence of the internet of things.**

## **Time-series Data**

Time-series data refers to a set of observations taken over a given period of time at specific and equally-spaced time intervals. That the observations are taken at specific points in time means time intervals are [discrete](https://analystprep.com/cfa-level-1-exam/quantitative-methods/probability-distributions-random-variables-types-examples/).

A good example of time-series data could be the daily or weekly closing price of a stock recorded over a period spanning 13 weeks. Other appropriate examples could be the set of monthly profits (both positive and negative) earned by Samsung between the 1st of October 2016 and the 1st of December 2016.

Time-series data can be used to predict future values of a given financial vehicle. Although such past data may help estimate future values, it’s always important to note that the future and the past are independent and therefore, past performance may not always be indicative of future performance.

**time-series-data**

Time-series data has at least one systematic pattern with the most common patterns being either trends or seasonality. Since most trends are linear or quadratic, regression analysis and the moving average method are used to establish the linear relationship between variables. Seasonality, on the other hand, is a trend that systematically keeps on repeating itself over time. There are numerous modern computer-based programs that are used to analyze time-series data including SPSS, JMP, SAS, Matlab, and R.

## **Cross-sectional Data**

Cross-sectional data refers to a setoff observations taken at a single point in time. Samples are constructed by collecting the data of interest across a range of observational units – people, objects, firms – at the same time.

A good example of cross-sectional data can be the stock returns earned by shareholders of Microsoft, IBM, and Samsung as for the year ended 31st December 2015:

**cross-sectional-data**

It is possible to pool time series data and cross-sectional data. If we were to study a particular characteristic or phenomenon across several entities over a period of time, we would end up with what’s referred to as panel data. For example, suppose we study the GDP of 3 developing countries for a period spanning 3 years, from 2015 to 2017:

|  |  |  |
| --- | --- | --- |
| **Country** | **Year** | **GDP** |
| **Kenya** | **2015** | **–** |
| **Kenya** | **2016** | **–** |
| **Kenya** | **2017** | **–** |
| **India** | **2015** | **–** |
| **India** | **2016** | **–** |
| **India** | **2017** | **–** |
| **Brazil** | **2015** | **–** |
| **Brazil** | **2016** | **–** |
| **Brazil** | **2017** | **–** |

**Here, we would study a group of entities (Kenya, India, and Brazil) over a period of time (3 yrs).This would constitute panel data.**