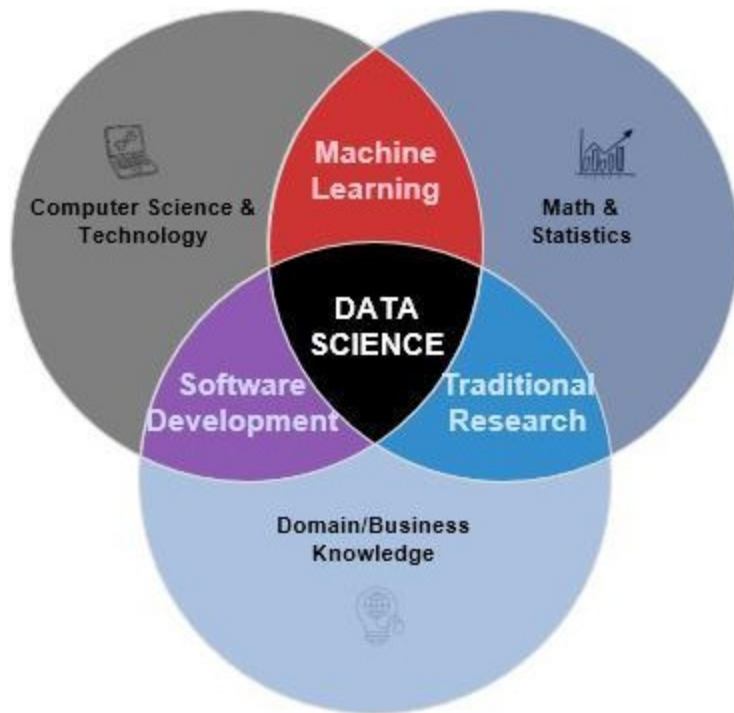


[← Go Back to Pre Work](#)[☰ Course Content](#)

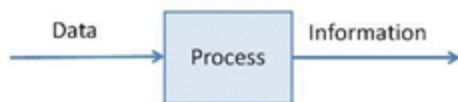
Introduction to Data Science

What is Data Science?

- Data Science is the science of extracting hidden patterns from large data sets
- Hidden patterns can appear in form of trends, cycles, associations, rules, groups etc. in the data
- Data sets usually refer to large volume of cleansed, structured data prepared for the analysis
- Science refers to the statistical tools and techniques employed to understand the data and reliability of the identified patterns
 - That part of statistics which is used to understand the data is called descriptive statistics. Descriptive statistics give vital insights into the data in terms of central values, spread and distribution (shape) of the data.
 - The part of statistics which is used to establish the reliability of the potential patterns identified, is called inferential statistics.
- This can easily be represented pictorially by following *venn* diagram



Data Vs Information:



What is Data?

Data is footprint of happenings – natural or human driven. Traditionally footprints used to be physical – in form of records in files or registers or receipts. Today majority of these footprints are digital – your browsing behavior, purchasing behavior etc. Data is the raw material that has information contained within but it has a lot of noise in it as well. When processed correctly, it can give us meaningful information.

- Ticket sales on a band on tour
- Survey data: Different companies collect data by survey to know the opinion of people about their product

What is Information?

Information is processed data. Information is basically the data plus the meaning of what the data was collected for minus the noise that got collected unintentionally.

Example:

- Sales report by region and venue - tells us which venue is the most profitable
- Survey Reports and Results: Survey data is summarized into reports/information to present to management of the company

Key Differences:

1. Data is the input and information is the output
2. Data is unprocessed records but information is processed data which has been made sense of

The Ascendancy of data

We live in a world that's drowning in data. Websites track every user's every click. Your

Proprietary content.©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

© 2023 All rights reserved

[Privacy](#) [Terms of service](#) [Help](#)