

# Data Science

Data Science is a study of analyzing and predicting the data, it is a way to analyze the pattern from the data and bring out meaningful insightful for business critical decisions and expansions.

Data Science is a blend of Mathematics, Statistics and Programming to analyze the data.

Data Science is all about the present and future. That is, finding out the trends based on historical data which can be useful for making present decisions and finding patterns which can be modelled and can be used for predictions to see what things may look like in the future.

## Why to learn Data Science?

With the amount of data that is being generated and the evolution in the field of Analytics, Data Science has turned out to be a necessity for companies. To make most out of their data, companies from all domains, be it Finance, Marketing, Retail, IT or Bank.

This field is such that anyone from any background can make a career as a Data Scientist.

Skills required to become a Data Scientist:



# Data Science Sense

## Data is big!

- 2.5 quintillion ( $10^{18}$ ) bytes of data are generated every day!

Everything around us collects/generates data.

- Social media sites
- Business transactions
- Location-based data
- Sensors
- Digital photos, videos
- Consumer behavior (online and store transactions)

## Business Analytics vs Business Intelligence vs Data Scientist

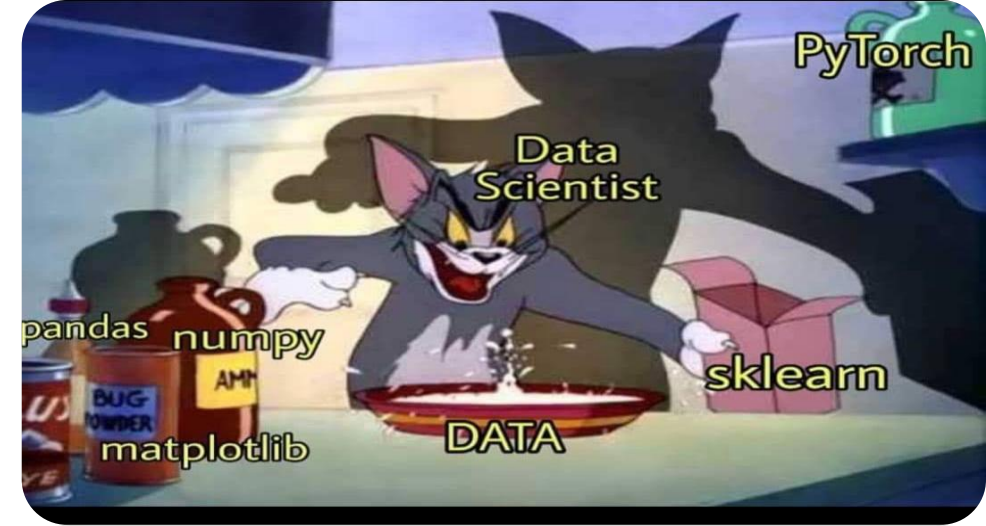
A business analytics professional has the skills to make use of the information from the data to generate insights about the business. To be a data focused business analytics professional, you must know the technical components related to managing and manipulating data.

A Business Intelligence Professional analyze the past trends using Data Visualization tools like Tableau, Power BI etc. to develop and implement business strategies. They also monitor all the performance metrics of the company and provide insight to the respective department.

Data Scientists help build complicated data models and simulations in a Big Data environment. Focusing more on math and statistics, these data scientists have a particular interest in reading statistics and building & deploying machine learning models.

# Data Scientists

Data scientists are big data wranglers, gathering and analyzing large sets of structured and unstructured data. A data scientist's role combines computer science, statistics, and mathematics. They analyze, process, and model data then interpret the results to create actionable plans for companies and other organizations.



Data scientists are analytical experts who utilize their skills in both technology and social science to find trends and manage data. They use industry knowledge, contextual understanding, skepticism of existing assumptions – to uncover solutions to business challenges.

Experienced data scientists and data managers typically have over ten years of experience and are tasked with developing a company's best practices, from cleaning to processing and storing data.

Data scientists don't need to just understand programming languages, management of databases and how to transpose data into visualizations – they should be naturally curious about their surrounding world, but through an analytical lens.

Possessing personality traits that resemble quality assurance departments, data scientists may be meticulous as they review large amounts of data and seek out patterns and answers. They are also creative in making new algorithms to crawl data or devising organized database warehouses.

# Data Science Business Domain

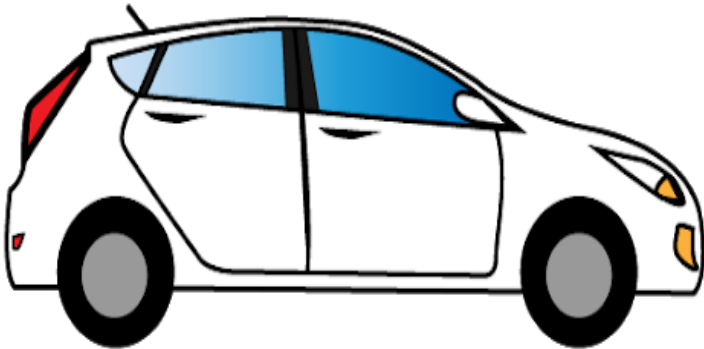
The term “Domain Knowledge” has been in play even before data science became popular. In software engineering, it means the knowledge about the environment in which the target (i.e. Data Scientist) operates.

We can use the same definition in data science to say — “Domain knowledge is the knowledge about the environment in which the data is processed to reveal secrets of the data”. In other words, the knowledge of the field that the data belongs to is known as Domain Knowledge.

## How does domain knowledge influence data science?

The true power of an algorithm and data can be harnessed only when we have some form of domain knowledge. Needless to say, the accuracy of the model also increases with the use of such knowledge of data.

For example, the knowledge of the automobile industry when working with the relevant data can be used like — Let’s say we have two features Horsepower and RPM from which we can create an additional feature like **Torque** from the formula. And also the weight of the car is directly proportionate to mileage.



# Data Science Project Lifecycle

