**Data Science**

**What is Data Science?**



Data Science is the process of applying scientific computations to extract meaningful insights from the billion and trillion bytes of data by using appropriate statistical methods.

Data Science is the discipline that is everyone’s word of mouth these days. The hype of Data Science has increased exponentially in recent years because of the enormous volumes of data that is getting generated from multiple sources.

Later in this article, we would look at how Data Science has affected our lives and how you could also be a Data Scientist with the right attitude and mastering the specific skills required for it.

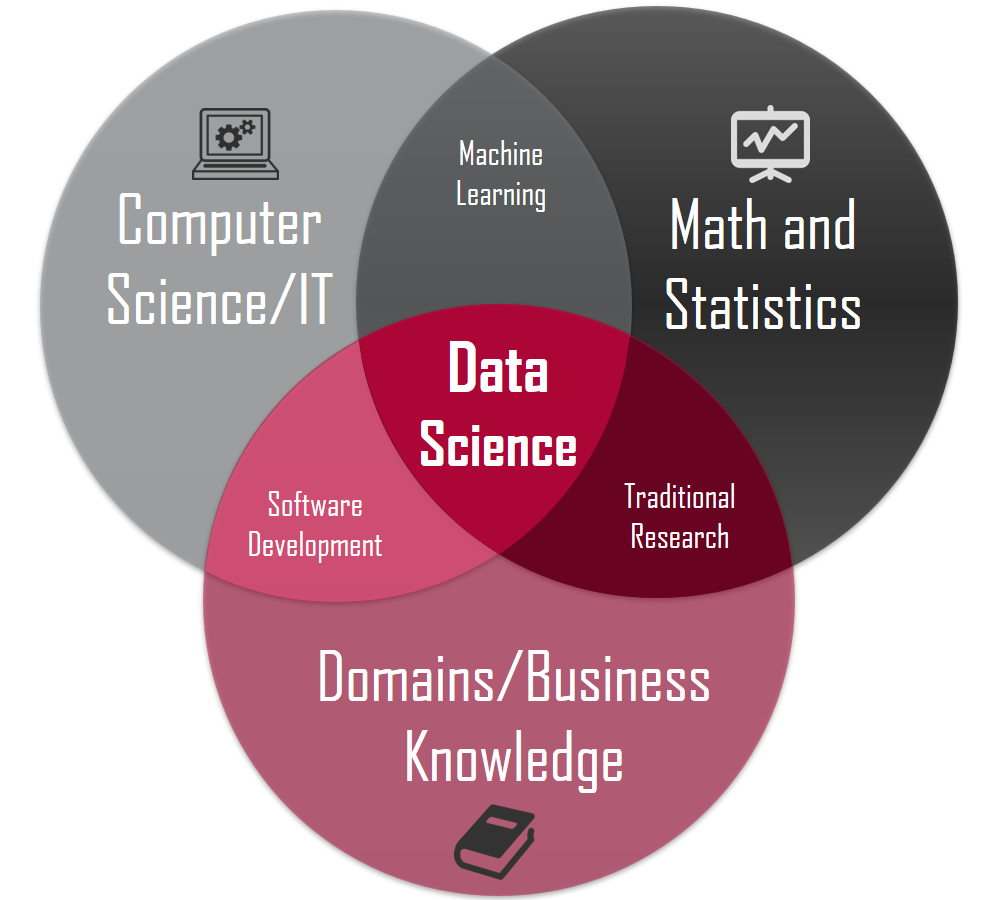
**Definition**

There is a massive debate about the exact definition of Data Science. In hindsight, there isn’t any formal definition which could be attached to the Data Science ecosystem, and different fields perceive Data Science differently.

Suppose, anyone working as a software engineer would often term the data visualization using a tool as a Data Science role, whereas someone working in the healthcare industry and dealing with sensitive patient data to predict cancer from the cells, would call that a Data Scientist job.

In layman terms, due to the diversity of its application, Data Science is defined differently by people belonging to different fields but all point to that one thing – extracting information from data using some methods.

**The various subsets of Data Science**



Data Science is a mixture of Mathematics, and Statistics, Machine Learning, Domain Knowledge, IT, and software development.

Math and Statistics is the core of Data Science as everything from Exploratory Data Analysis to Model Building requires dealing with numbers, vectors, probability, and so on.

Machine Learning could be further divided into Deep Learning and Artificial Intelligence, and it is the model building subset of Data Science.

Additionally, essential software development and IT skills are deemed necessary to apply Data Science in those fields.

Finally, having the business or domain knowledge could go a long way in determining the accuracy of the result as different business uses different data for prediction and using the right data is of utmost importance in verifying the credibility of our output.

**Understanding Data Science**



Data Science is primarily the Science used to uncover hidden patterns from data. Those hidden patterns or insights could go a long way in achieving ground-breaking results in several fields and improve the lives of the people.

The image above shows the six stages in a Data Science workflow which helps in making predictions and build models to be used in the production. It's described in details in the next section.

**Working with Data Science**

A Data Science work would be divided into the following categories.

* *Understanding the Problem* – It is essential that the problem statement is clear before you dive in to the actual implementation part. The knowledge of what to find out is crucial to get the right data and to derive the perfect solution.
* *Getting the right data* – Once the problem is understood, it’s imperative to get the right data to perform the operation.
* *Exploratory Data Analysis* – It’s said that ninety percent of the work done by a Data Scientist is Data Wrangling. The term data wrangling refers to cleaning and pre-processing the data before feeding to the model. The steps involve checking for duplicate data, outliers, NULL values, and several other anomalies which don’t fall under the convention of a desired data for the business.
* *Data Visualization* – Once the data is cleaned and pre-processed, it’s necessary to visualize the data to find out the right features or columns to use for our model.
* *Categorical Encoding* – This step is applicable for those instances where the input features are categorical and needed to be transformed into numeric (0,1,2, etc.,) to be used in our model as the machine cannot work with categories.
* *Model Selection*- Selecting the right model for a particular problem statement is essential as every model cannot fit in perfectly for every data set.
* *Using the right metric*- Based on the business domain, the metric which would determine the perfectness of a model should be selected.
* *Communication-* The businessman, the shareholders, often don’t understand the technical know-how of Data Science, and hence it’s essential to communicate the findings in simple terms to the business who could then come up with measures to mitigate any foreseen risks.
* *Deployment-* Once the model is built, and the business is satisfied with the findings, the model could be deployed to production and used in the product.

**What can you do with Data Science?**

Data Science is rapidly consuming our daily lives. Starting from waking up in the morning to going to bed, there isn’t a single moment that the effects of Data Science don't influence us. Let’s look at some of the usages of Data Science which has made our life easy in recent times.

*Example 1:-*

YouTube is the favourite mode of entertainment, knowledge, news in our daily lives. We prefer to watch videos than going through slides of long articles. But how did we become so addictive to YouTube? What has made YouTube so unique and different?

Well, the answer is simple. YouTube uses our data to recommend the videos; we would like to see next. It uses a recommender system algorithm to track our search patterns and based on that; its intelligence system shows us those videos which are somewhat related to the one we have seen so that we are glued to the channel and continue surfing through the other videos.

So basically, it saves our time and energy to manually look for videos which might be helpful to us based on our liking.

*Example 2:-*

Similar to YouTube, the recommender system is also used in e-commerce websites like Netflix, Amazon.

In the case of Netflix, we are shown those TV shows or movies which are somewhat related to the one we have watched and thus saves our time to look for more similar videos.

Additionally, Amazon recommends the products based on our buying pattern, and it displays those products which other buyers have bought along with that product or what we could buy based on our shopping habits or patterns.

*Example 3:-*

One of the major breakthroughs in Data Science is Amazon’s Alexa or Apple’s Siri. Often we find tedious to surf through our phone for contacts or feel lazy to set up alarm bells or reminders.

In this regard, the virtual assistant systems do all the stuff for us only by listening to our commands. We tell Alexa or Siri about the things we want and the system convert our natural voice to text using the Natural Language Processing topology (we would see that later on) and extract insights from that text to solve our problems.

In layman terms, this Intelligent Systems uses Speech to Voice terminology to save time and solve our problems.

*Example 4:-*

Data Science has eased the life of athletes and people involved in Sports arenas as well. The enormous amount of data that’s available these days could be used to analyze a sportsman's health and mental conditions to prepare accordingly for a game.

Also, the data could be used to make strategies and outplay the opponent even before the match starts.

*Example 5:-*

Data Science has eased the life in the Healthcare sector as well. The medics and the researchers could use Deep Learning to analyse a cell and stop a disease from occurring in the first place.

They could also prescribe adequate medication for a patient based on the prediction from the data.

**Top Data Science Companies**

Data Science is regarded as the sexiest job of the 21st century with professionals from different backgrounds embarking on the journey of becoming a Data Scientist.

Nowadays almost every company is trying to incorporate Data Science in its products to simplify the process and speed fast the operations to ensure accuracy in optimal time. The list of such companies is enormous, and it would be deemed unfair to pit one against the other in terms of the best as different companies use data for various reasons.

Along with the USA, the Data Science market in India is expanding, and it would only benefit professionals in the future. Here are some of the top companies where Data Science has an exhaustive usage:-

*JP Morgan, Deloitte, Bitwise, Salesforce, LinkedIn, Flipkart, WNS, Mc Kinsey & Company, IBM, Ola Cabs, Mu Sigma, Stripe, Amazon, Big Basket, Netflix, Wipro, Enterprise Bot, Accenture, Myntra, Manthan, TCS, Cisco, Cartesian Analytics, HCL, EDGE Networks, Walmart labs, Cognizant, [24]*[*7.ai*](http://7.ai/)*,Target Corporation, TEG Analytics, Citrix, Sigmoid, Facebook, Twitter, Google Inc., Gobble, Reliance, Square,* [*niki.ai*](http://niki.ai/)*, Dropbox , Airbnb, Khan Academy, Uber, Pinterest , Fractal Analytics.*

The sites where you could find several Data Science openings are - [LinkedIn](https://www.linkedin.com/), [Indeed](https://www.indeed.com/), [Simply Hired](https://www.simplyhired.com/), and [AngelList](https://angel.co/jobs" \t "_blank).

**Who is the right audience for learning Data Science technologies?**

Data Science is about working with data, and every field uses data in some way or another. Hence, you don’t need to belong to a specific discipline to be a Data Scientist.

However, what you do need is a curious mind-set and an eagerness to carve out insights from data.

**Advantages of Data Science**

• Data Science could help to mitigate time and budget allocation constraints and assist in the growth of the business.

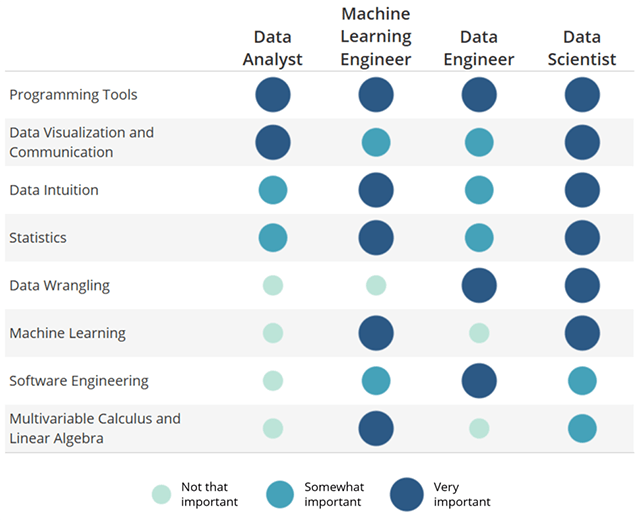
• Machine determined results of several manual tasks which could be better than human effects.

• Data Science helps prevent loan default, used in fraud detection, and several other use cases in the financial domain.

• Generate insights from raw, unstructured textual data.

• Predicting the future outcome could prevent the financial loss of many big corporations.

**Required Data Science skills**



The above image indicates the importance of the skills required based on different roles.

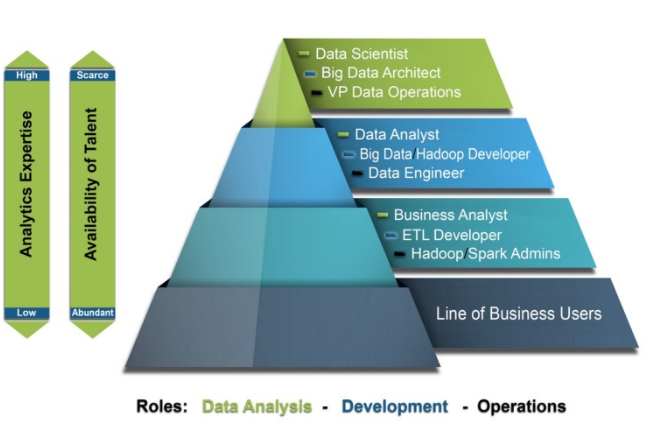
Programming, Data Visualization, Communication, Data Intuition, Statistics, Data Wrangling, Machine Learning, Software Engineering, and Mathematics are the required skills for anyone who wants to enter into Data Science space.

**Why should we use Data Science?**

The usage of Data Science in academia and in real life is vastly different. While in academia, Data Science is used to solve several cool projects like image recognition, face detection, etc.

On the other hand, in daily life, Data Science is used to prevent fraud, finger print detection, product recommendation, and so on.

**Data Science scope**



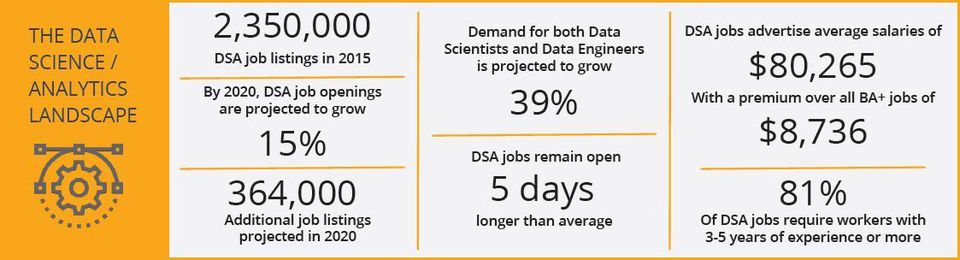
The opportunities or scope in Data Science is boundless. As shown in the image above, a professional could work in several different roles in Data Science depending on their skill set and the level of expertise.

**Why do we need Data Science?**

A lot of the work done nowadays is manual and takes a lot of time and resources which often causes hindrance to the budget allocated for the project. Big companies sometimes look for solutions to optimize such tasks and ensure the budget and resource constraints are mitigated.

Data Science gives the opportunity to automate the tedious processes and produce such outstanding results which might not have been possible in manual work.

**How this technology would help you in career growth?**



This survey by Forbes shows that Data Science is the future and it is here to stay. The days of manual work is over, and Data Science would automate every such task. Hence, if you want to remain relevant in the industry in the future, it’s necessary that you learn the various aspects of Data Science and increase your chances of always being employed.

**Conclusion**

If you are a graduate or a working professional, it’s high time that you hop onto the Data Science ship and get yourself involved in the Data Science community.

Listed below are few of the resources which help you get the required skills in your journey.

*Linear Algebra -* [*https://www.khanacademy.org/math/linear-algebra*](https://www.khanacademy.org/math/linear-algebra)

*Statistics, and Probability -* [*https://www.khanacademy.org/math/statistics-probability*](https://www.khanacademy.org/math/statistics-probability)

*Multivariate Calculus -* [*https://www.khanacademy.org/math/multivariable-calculus*](https://www.khanacademy.org/math/multivariable-calculus)

*Data Analysis -* [*https://in.udacity.com/course/data-analyst-nanodegree--nd002*](https://in.udacity.com/course/data-analyst-nanodegree--nd002)

*Machine Learning -* [*https://www.coursera.org/learn/machine-learning*](https://www.coursera.org/learn/machine-learning)*?*

[*https://www.udemy.com/machinelearning/*](https://www.udemy.com/machinelearning/)

*Deep Learning -* [*https://in.udacity.com/course/deep-learning--ud730-india*](https://in.udacity.com/course/deep-learning--ud730-india)

*A complete Data Science track with Python* - <https://www.datacamp.com/tracks/data-scientist-with-python>

*A complete Data Science track with R -* [*https://www.datacamp.com/tracks/data-scientist-with-R*](https://www.datacamp.com/tracks/data-scientist-with-R)

Apart from all these, some of the other sites where you could visit to check articles and blogs about Data Science are –

Medium - <https://medium.com/>

KDnuggets - <https://www.kdnuggets.com/>

EDUCBA - <https://www.educba.com/>