**What is Data Science? Tutorial for Beginners**

**What is Data Science?**

Data Science is the area of study which involves extracting insights from vast amounts of data by the use of various scientific methods, algorithms, and processes. It helps you to discover hidden patterns from the raw data.

The term Data Science has emerged because of the evolution of mathematical statistics, data analysis, and big data.

Data Science is an interdisciplinary field that allows you to extract knowledge from structured or unstructured data. Data science enables you to translate a business problem into a research project and then translate it back into a practical solution.

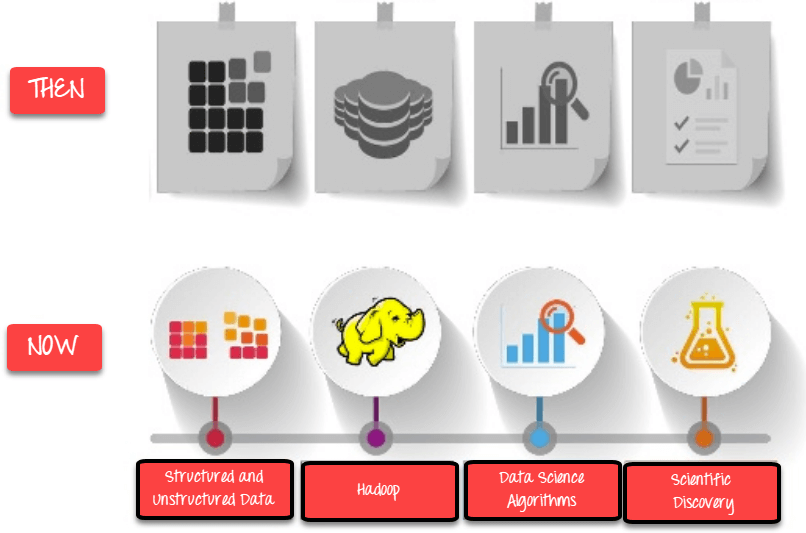
In this tutorial, you will learn

* [What is Data Science?](https://www.guru99.com/data-science-tutorial.html#1)
* [Why Data Science?](https://www.guru99.com/data-science-tutorial.html#2)
* [Data Science Components](https://www.guru99.com/data-science-tutorial.html#3)
* [Data Science Process](https://www.guru99.com/data-science-tutorial.html#4)
* [Data Science Jobs Roles](https://www.guru99.com/data-science-tutorial.html#5)
* [Tools for DataScience](https://www.guru99.com/data-science-tutorial.html#6)
* [Difference between Data Science with BI (Business Intelligence)](https://www.guru99.com/data-science-tutorial.html#7)
* [Applications of Data science](https://www.guru99.com/data-science-tutorial.html#8)
* [Challenges of Data science Technology](https://www.guru99.com/data-science-tutorial.html#9)

**Why Data Science?**

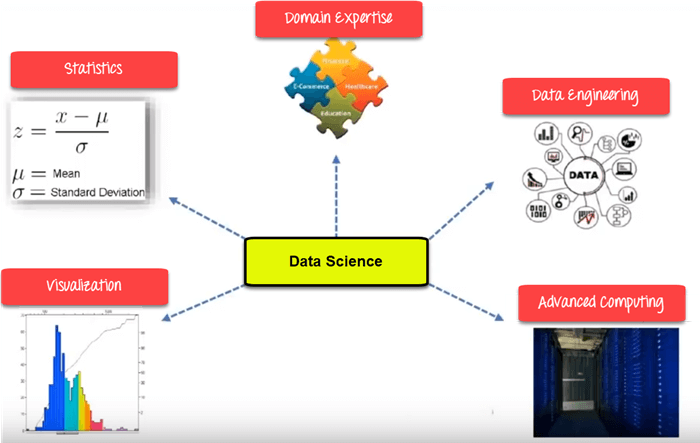
Here, are significant advantages of using Data Analytics Technology:

* Data is the oil for today's world. With the right tools, technologies, algorithms, we can use data and convert it into a distinctive business advantage
* Data Science can help you to detect fraud using advanced machine learning algorithms
* It helps you to prevent any significant monetary losses
* Allows to build intelligence ability in machines
* You can perform sentiment analysis to gauge customer brand loyalty
* It enables you to take better and faster decisions
* Helps you to recommend the right product to the right customer to enhance your business



Evolution of DataSciences

Data Science Components



### Statistics:

Statistics is the most critical unit in Data science. It is the method or science of collecting and analyzing numerical data in large quantities to get useful insights.

### Visualization:

Visualization technique helps you to access huge amounts

of data in easy to understand and digestible visuals.

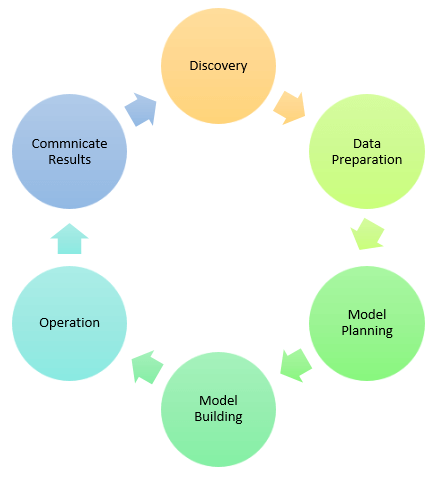
### Machine Learning:

Machine Learning explores the building and study of algorithms which learn to make predictions about unforeseen/future data.

### Deep Learning:

Deep Learning method is new machine learning research where the algorithm selects the analysis model to follow.

## Data Science Process



### 1.Discovery:

Discovery step involves acquiring data from all the identified internal & external sources which helps you to answer the business question.

The data can be:

* Logs from webservers
* Data gathered from social media
* Census datasets
* Data streamed from online sources using APIs

### 2.Data Preparation:

Data can have lots of inconsistencies like missing value, blank columns, incorrect data format which needs to be cleaned. You need to process, explore, and condition data before modeling. The cleaner your data, the better are your predictions.

### 3.Model Planning:

In this stage, you need to determine the method and technique to draw the relation between input variables. Planning for a model is performed by using different statistical formulas and visualization tools. SQL analysis services, R, and SAS/access are some of the tools used for this purpose.

### 4. Model Building:

In this step, the actual model building process starts. Here, Data scientist distributes datasets for training and testing. Techniques like association, classification, and clustering are applied to the training data set. The model once prepared is tested against the "testing" dataset.

### 5. Operationalize:

In this stage, you deliver the final baselined model with reports, code, and technical documents. Model is deployed into a real-time production environment after thorough testing.

### 6. Communicate Results

In this stage, the key findings are communicated to all stakeholders. This helps you to decide if the results of the project are a success or a failure based on the inputs from the model.

## Data Science Jobs Roles

Most prominent Data Scientist job titles are:

* Data Scientist
* Data Engineer
* Data Analyst
* Statistician
* Data Architect
* Data Admin
* Business Analyst
* Data/Analytics Manager

Let's learn what each role entails in detail:

### Data Scientist:

**Role:**

A Data Scientist is a professional who manages enormous amounts of data to come up with compelling business visions by using various tools, techniques, methodologies, algorithms, etc.

**Languages**:

R, SAS, Python, SQL, Hive, Matlab, Pig, Spark

### Data Engineer:

**Role**:

The role of data engineer is of working with large amounts of data. He develops, constructs, tests, and maintains architectures like large scale processing system and databases.

**Languages**:

SQL, Hive, R, SAS, Matlab, Python, Java, Ruby, C + +, and Perl

### Data Analyst:

**Role**:

A data analyst is responsible for mining vast amounts of data. He or she will look for relationships, patterns, trends in data. Later he or she will deliver compelling reporting and visualization for analyzing the data to take the most viable business decisions.

**Languages**:

R, Python, HTML, JS, C, C+ + , SQL

### Statistician:

**Role**:

The statistician collects, analyses, understand qualitative and quantitative data by using statistical theories and methods.

**Languages**:

SQL, R, Matlab, Tableau, Python, Perl, Spark, and Hive

### Data Administrator:

**Role**:

Data admin should ensure that the database is accessible to all relevant users. He also makes sure that it is performing correctly and is being kept safe from hacking.

**Languages**:

Ruby on Rails, SQL, Java, C#, and Python

### Business Analyst:

**Role**:

This professional need to improves business processes. He/she as an intermediary between the business executive team and IT department.

**Languages**:

SQL, Tableau, Power BI and, Python

## Tools for DataScience



|  |  |  |  |
| --- | --- | --- | --- |
| **Data Analysis** | **Data warehousing** | **Data Visualization** | **Machine Learning** |
| [R](https://www.guru99.com/r-tutorial.html), Spark, [Python](https://www.guru99.com/python-tutorials.html) and [SAS](https://www.guru99.com/sas-tutorial.html) | [Hadoop](https://www.guru99.com/bigdata-tutorials.html), SQL, [Hive](https://www.guru99.com/hive-tutorials.html) | R, [Tableau](https://www.guru99.com/tableau-tutorial.html), Raw | [Spark](https://www.guru99.com/pyspark-tutorial.html), Azure ML studio, Mahout |

## Difference between Data Science with BI (Business Intelligence)

|  |  |  |
| --- | --- | --- |
| **Parameters** | **Business Intelligence** | **Data Science** |
| Perception | Looking Backward | Looking Forward |
| Data Sources | Structured Data. Mostly SQL, but some time Data Warehouse) | Structured and Unstructured data. Like logs, SQL, NoSQL, or text |
| Approach | Statistics & Visualization | Statistics, Machine Learning, and Graph |
| Emphasis | Past & Present | Analysis & Neuro-linguistic Programming |
| Tools | Pentaho. Microsoft Bl, QlikView, | R, [TensorFlow](https://www.guru99.com/tensorflow-tutorial.html) |

## Applications of Data science

### Internet Search:

Google search use Data science technology to search a specific result within a fraction of a second

### Recommendation Systems:

To create a recommendation system. Example, "suggested friends" on Facebook or suggested videos" on YouTube, everything is done with the help of Data Science.

### Image & Speech Recognition:

Speech recognizes system like Siri, Google assistant, Alexa runs on the technique of Data science. Moreover, Facebook recognizes your friend when you upload a photo with them, with the help of Data Science.

### Gaming world:

EA Sports, Sony, Nintendo, are using Data science technology. This enhances your gaming experience. Games are now developed using Machine Learning technique. It can update itself when you move to higher levels.

### Online Price Comparison:

PriceRunner, Junglee, Shopzilla work on the Data science mechanism. Here, data is fetched from the relevant websites using APIs.

## Challenges of Data science Technology

* High variety of information & data is required for accurate analysis
* Not adequate data science talent pool available
* Management does not provide financial support for a data science team
* Unavailability of/difficult access to data
* Data Science results not effectively used by business decision makers
* Explaining data science to others is difficult
* Privacy issues
* Lack of significant domain expert
* If an organization is very small, they can't have a Data Science team

## Summary

* Data Science is the area of study which involves extracting insights from vast amounts of data by the use of various scientific methods, algorithms, and processes.
* Statistics, Visualization, Deep Learning, Machine Learning, are important Data Science concepts.
* Data Science Process goes through Discovery, Data Preparation, Model Planning, Model Building, Operationalize, Communicate Results.
* Important Data Scientist job roles are: 1) Data Scientist 2) Data Engineer 3) Data Analyst 4) Statistician 5) Data Architect 6) Data Admin 7) Business Analyst 8) Data/Analytics Manager
* R, SQL, Python, SaS, are essential Data science tools
* The predictions of Business Intelligence is looking backward while for Data Science it is looking forward.
* Important applications of Data science are 1) Internet Search 2) Recommendation Systems 3) Image & Speech Recognition 4) Gaming world 5) Online Price Comparison.
* High variety of information & data is the biggest challenge of Data Science technology.