

Data Science:

The term Data Science has emerged recently with the evolution of mathematical statistics and data analysis. The journey has been amazing, we have accomplished so much today in the field of Data Science.

In the next few years, we will be able to predict the future as claimed by researchers from MIT. They already have reached a milestone in predicting the future, with their awesome research. They can now predict what will happen in the next scene of a movie, with their machine! How? Well it might be a little complex for you to understand as of now, but don't worry by the end of this blog, you shall have an answer to that as well.

Coming back, we were talking about Data Science, it is also known as data driven science, which makes use of scientific methods, processes and systems to extract knowledge or insights from data in various forms, i.e either structured or unstructured.

Data science is a way to try and discover hidden patterns in raw data. To achieve this goal, it makes use of several algorithms, **machine learning(ML)** principles, and scientific methods. The insights it retrieves from data lie in forms structured and unstructured. So in a way, this is like **data mining**. Data science encompasses all- data analysis, statistics, and machine learning. With more practices being labelled into data science, the term itself becomes diluted beyond usefulness. This leads to variation in curricula for introductory data science courses worldwide.

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.

Data science is a deep study of the massive amount of data, which involves extracting meaningful insights from raw, structured, and unstructured data that is processed using the scientific method, different technologies, and algorithms.

It is a multidisciplinary field that uses tools and techniques to manipulate the data so that you can find something new and meaningful.

Data science uses the most powerful hardware, programming systems, and most efficient algorithms to solve the data related problems. It is the future of artificial intelligence.

In short, we can say that data science is all about:

- Asking the correct questions and analyzing the raw data.
- Modeling the data using various complex and efficient algorithms.
- Visualizing the data to get a better perspective.
- Understanding the data to make better decisions and finding the final result.



Example:

Let suppose we want to travel from station A to station B by car. Now, we need to take some decisions such as which route will be the best route to reach faster at the location, in which route there will be no traffic jam, and which will be cost-effective. All these decision factors will act as input data, and we will get an appropriate answer from these decisions, so this analysis of data is called the data analysis, which is a part of data science.

Data Science has become the most demanding job of the 21st century. Every organization is looking for candidates with knowledge of data science. In this tutorial, we are giving an introduction to data science, with data science Job roles, tools for data science, components of data science, application, etc.

Why Data Science?

It's been said that Data Scientist is the "Most demandable Job of the 21st century". Why? Because over the past few years, companies have been storing their data. And this being done by each and every company, it has suddenly led to data explosion. Data has become the most abundant thing today.

But, what will you do with this data? Let's understand this using an example:

Say, you have a company which makes mobile phones. You released your first product, and it became a massive hit. Every technology has a life, right? So, now its time to come up with something new. But you don't know what should be innovated, so as to meet the expectations of the users, who are eagerly waiting for your next release?

Somebody, in your company comes up with an idea of using the user generated feedback and pick things which we feel users are expecting in the next release.

Comes in Data Science, you apply various data mining techniques like sentiment analysis etc and get the desired results.

It's not only this, you can make better decisions, you can reduce your production costs by coming out with efficient ways, and give your customers what they actually want!

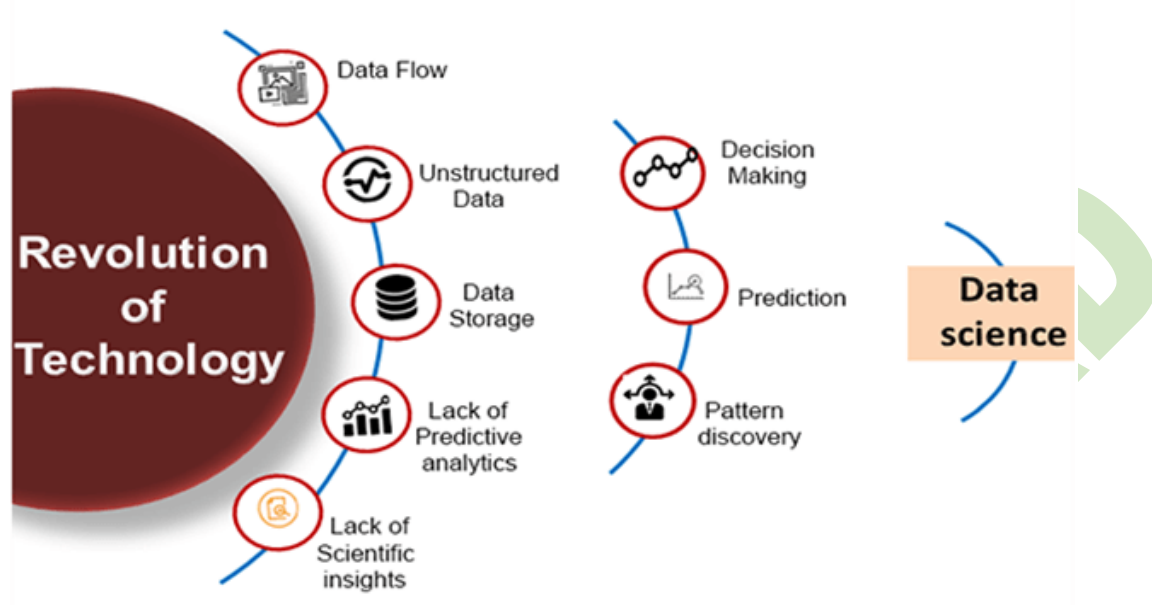
With this, there are countless benefits that Data Science can result in, and hence it has become absolutely necessary for your company to have a Data Science Team. Requirements like these led to “Data Science” as a subject today. Here, are significant advantages of using Data Analytics Technology:

- 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

Need for Data Science:



Some years ago, data was less and mostly available in a structured form, which could be easily stored in excel sheets, and processed using BI tools.

But in today's world, data is becoming so vast, i.e., approximately **2.5 quintals bytes** of data is generating on every day, which led to data explosion. It is estimated as per researches, that by 2020, 1.7 MB of data will be created at every single second, by a single person on earth. Every Company requires data to work, grow, and improve their businesses.

Now, handling of such huge amount of data is a challenging task for every organization. So to handle, process, and analysis of this, we required some complex, powerful, and efficient algorithms and technology, and that technology came into existence as data Science. Following are some main reasons for using data science technology:

- With the help of data science technology, we can convert the massive amount of raw and unstructured data into meaningful insights.
- Data science technology is opting by various companies, whether it is a big brand or a startup. Google, Amazon, Netflix, etc. which handle the huge amount of data, are using data science algorithms for better customer experience.
- Data science is working for automating transportation such as creating a self-driving car, which is the future of transportation.
- Data science can help in different predictions such as various survey, elections, flight ticket confirmation, etc.

Data Science Applications:



a. Image Recognition

Using the face recognition algorithm of data science, we can get a lot done. Did Facebook ever suggest people tag in your pictures? Have you tried the search-by-image feature from Google? Do you remember scanning a barcode to log in to WhatsApp Web using your smartphone?

b. Speech Recognition

Siri, Alexa, Cortana, Google Voice all make use of speech recognition to understand your commands. Attributing to issues like different accents and ambient noise, this isn't always completely accurate, though intelligible most of the time. This facilitates luxury like speaking the content of a text to send, using your virtual assistant to set an alarm, or even use it to play music, inquire about the weather, or make a call.

c. Internet Search

Search engines like Google, Duckduckgo, Yahoo, and Bing make good use of data science to make fast, real-time searching possible.

d. Digital Advertisements

Data science algorithms let us understand customer behaviour. Using this information, we can put up relevant advertisements curated for each user. This also applies to advertisements as banners on websites and digital billboards at airports.

e. Recommender Systems

Names like Amazon and Youtube will throw in suggestions about similar products aside or below as you browse through a product or a video. This enriches the UX(user experience) and helps retain customers and users. This will also take into account the user's search history and wishlist.

f. Price Comparison Websites

Websites like Junglelee and PriceDekho let us compare prices for the same products across different platforms. This facility lets you make sure you grab the best deal. These websites work in the domains of technology, apparel, and policy among many others, and use APIs and RSS feeds to fetch data.

g. Gaming

As a player levels up, a **machine learning algorithm** can improve or upgrade itself. It is also possible for the opponent to analyze the player's moves and add an element of difficulty to the game. Companies like Sony and Nintendo make use of this.

h. Delivery Logistics

Freight giants like UPS, FedEx, and DHL use practices of data science to discover optimal routes, delivery times, and transport modes among many others. A plus with logistics is the data obtained from the GPS devices installed.

i. Fraud and Risk Detection

Practices like customer profiling and past expenditures let us analyze whether there will be a failure. This lets banks avoid debts and losses.

Who Is a Data Scientist?

Data scientists are a mix of mathematicians, trend-spotters, and computer scientists. The data scientist's role is to decipher large volumes of data and carry out further analysis to find trends in the data and gain a deeper insight into what it all means. Data scientists operate between the business and IT worlds and drive industries by analyzing complex datasets to tease out insights that companies can leverage into actions.



What Are Data Science Roles Out There?

To name a few, some of the most common job titles for data scientists include:

1. Business Intelligence Analyst

ABI analyst uses data to help figure out market and business trends by analyzing data to develop a clearer picture of where the company stands.

2. Data Mining Engineer

The data mining engineer examines not only the data for their own business but also that of third parties. In addition to analyzing data, a data mining engineer will create sophisticated algorithms to help analyze the data further.

3. Data Architect

Data architects work closely with users, system designers, and developers to create blueprints that data management systems use to centralize, integrate, maintain, and protect data sources.

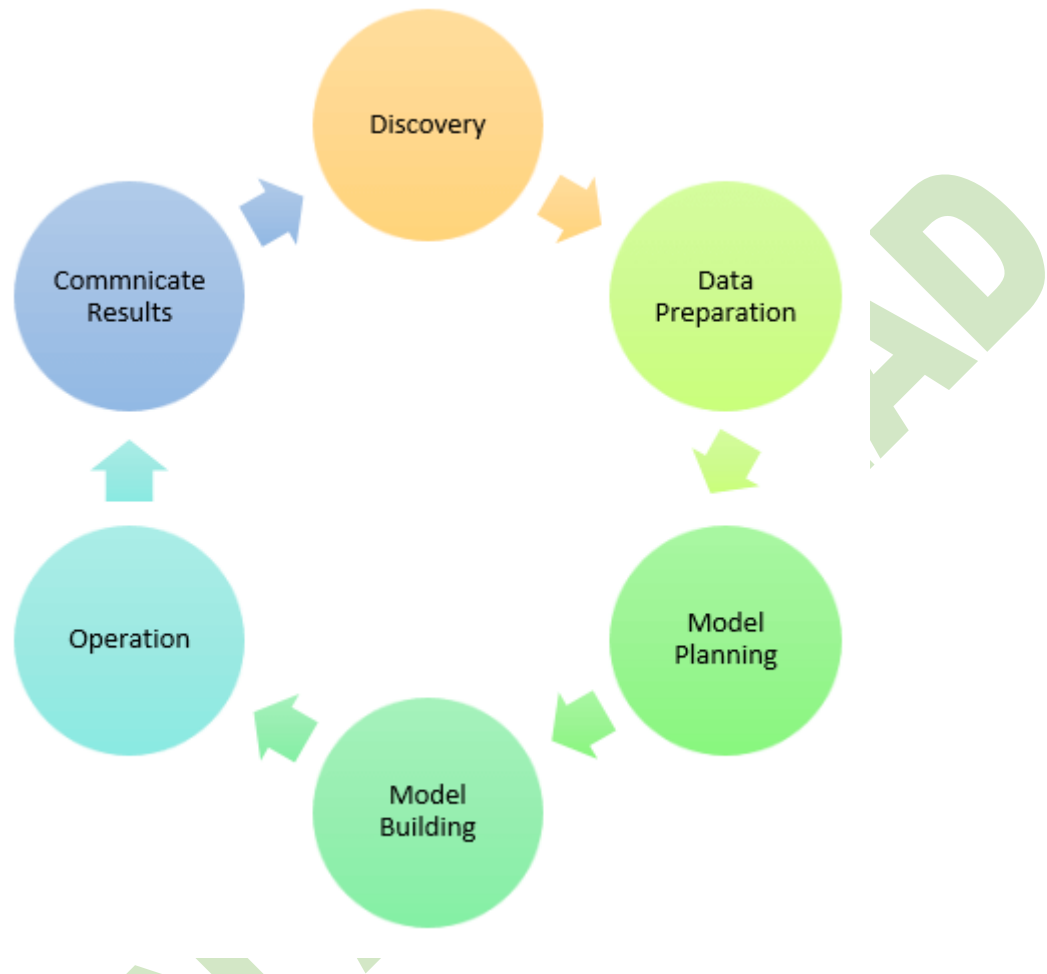
4. Data Scientist

Data scientists begin by translating a business case into an analytics agenda, developing hypotheses, and understanding data—as well as exploring patterns to measure what impact they will have on businesses. They also find and choose algorithms to help further analyze data. They use business analytics to not only explain what effect the data is going to have on a company in the future but can also help devise solutions that will help the company to move forward.

5. Senior Data Scientist

A senior data scientist can anticipate what a business's future needs will be. Apart from gathering data, they also analyze it thoroughly to resolve highly complex business problems efficiently. Through their experience, they can not only design but drive forward the creation of new standards, as well as create ways to use statistical data, and also develop tools to help further analyze the data.

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.

Designed by Abdur Rahman Joy - MCSD, MCPD, MCSE, MCTS, OCJP, Sr. Technical Trainer for VFX at IDB BISW (Scholarship program), and C#.net, R, Scala, Kotlin, JAVA, Android/IOS/Windows Mobile Apps, SQL server, Azure, Oracle, SharePoint Development, AWS , CEH, KALI Linux, Python, Data Science, Machine Learning ,Software Testing, Graphics, Multimedia and Game Developer at Joy Infosys and other premises like BITM, SkillsJob, PNTL, Leads Training and New Horizon inc , Cell #: +880-1712587348, email: jspaonline@gmail.com. Web URL: <http://www.joyinfosys.com/me>.

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:

As per various surveys, data scientist job is becoming the most demanding Job of the 21st century due to increasing demands for data science. Some people also called it "the **hottest job title of the 21st century**". Data scientists are the experts who can use various statistical tools and machine learning algorithms to understand and analyze the data.

The average salary range for data scientist will be approximately **\$95,000 to \$ 165,000 per annum**, and as per different researches, about **11.5 millions** of job will be created by the year **2026**.

Types of Data Science Job

If you learn data science, then you get the opportunity to find the various exciting job roles in this domain. The main job roles are given below:

1. Data Scientist
2. Data Analyst
3. Machine learning expert
4. Data engineer
5. Data Architect
6. Data Administrator
7. Business Analyst
8. Business Intelligence Manager

Below is the explanation of some critical job titles of data science.

1. Data Analyst:

Data analyst is an individual, who performs mining of huge amount of data, models the data, looks for patterns, relationship, trends, and so on. At the end of the day, he comes up with visualization and reporting for analyzing the data for decision making and problem-solving process.

Skill required: For becoming a data analyst, you must get a good background in **mathematics, business intelligence, data mining**, and basic knowledge of **statistics**. You should also be familiar with some computer languages and tools such as **Python, MATLAB, SQL, Hive, Pig, Excel, SAS, R, JS, Spark**, etc.

2. Machine Learning Expert:

The machine learning expert is the one who works with various machine learning algorithms used in data science such as **regression, clustering, classification, decision tree, random forest**, etc.

Skill Required: Computer programming languages such as Python, C++, R, Java, and Hadoop. You should also have an understanding of various algorithms, problem-solving analytical skill, probability, and statistics.

3. Data Engineer:

A data engineer works with massive amount of data and responsible for building and maintaining the data architecture of a data science project. Data engineer also works for the creation of data set processes used in modeling, mining, acquisition, and verification.

Skill required: Data engineer must have depth knowledge of **SQL, MongoDB, Cassandra, HBase, Apache Spark, Hive, MapReduce**, with language knowledge of **Python, C/C++, Java, Perl**, etc.

4. Data Scientist:

A data scientist is a professional who works with an enormous amount of data to come up with compelling business insights through the deployment of various tools, techniques, methodologies, algorithms, etc.

Skill required: To become a data scientist, one should have technical language skills such as **Python, R, SAS, SQL, Hive, Pig, Apache spark, MATLAB**. Data scientists must have an understanding of Statistics, Mathematics, visualization, and communication skills.

Prerequisite for Data Science

Non-Technical Prerequisite:

- **Curiosity:** To learn data science, one must have curiosities. When you have curiosity and ask various questions, then you can understand the business problem easily.
- **Critical Thinking:** It is also required for a data scientist so that you can find multiple new ways to solve the problem with efficiency.
- **Communication skills:** Communication skills are most important for a data scientist because after solving a business problem, you need to communicate it with the team.

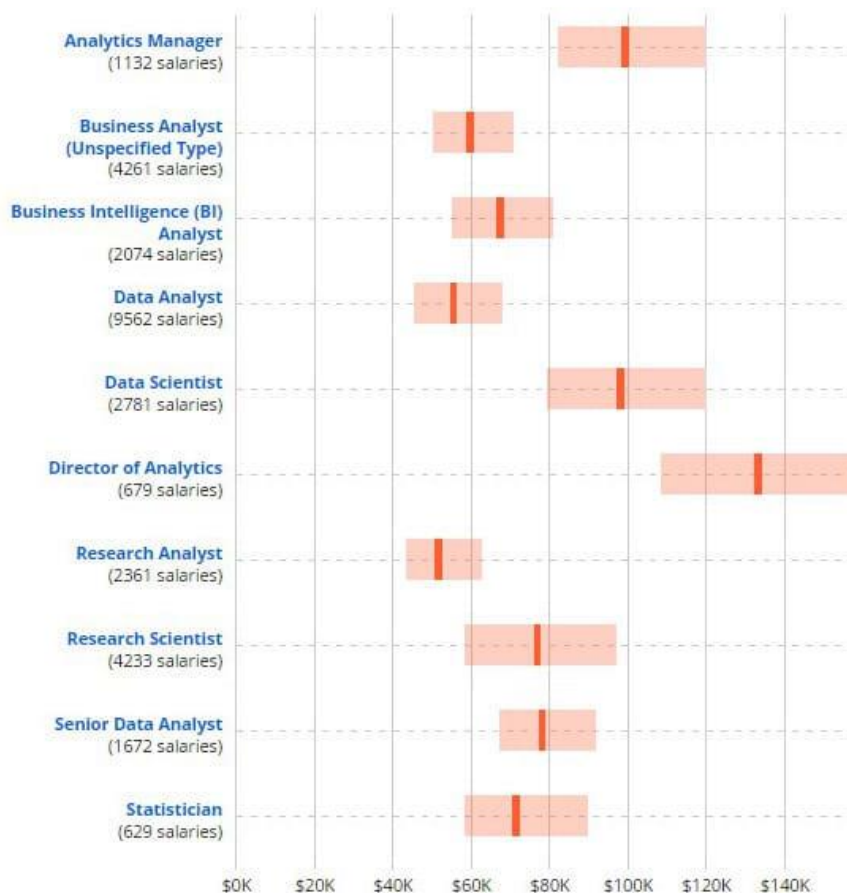
Technical Prerequisite:

- **Machine learning:** To understand data science, one needs to understand the concept of machine learning. Data science uses machine learning algorithms to solve various problems.
- **Mathematical modeling:** Mathematical modeling is required to make fast mathematical calculations and predictions from the available data.
- **Statistics:** Basic understanding of statistics is required, such as mean, median, or standard deviation. It is needed to extract knowledge and obtain better results from the data.
- **Computer programming:** For data science, knowledge of at least one programming language is required. R, Python, Spark are some required computer programming languages for data science.

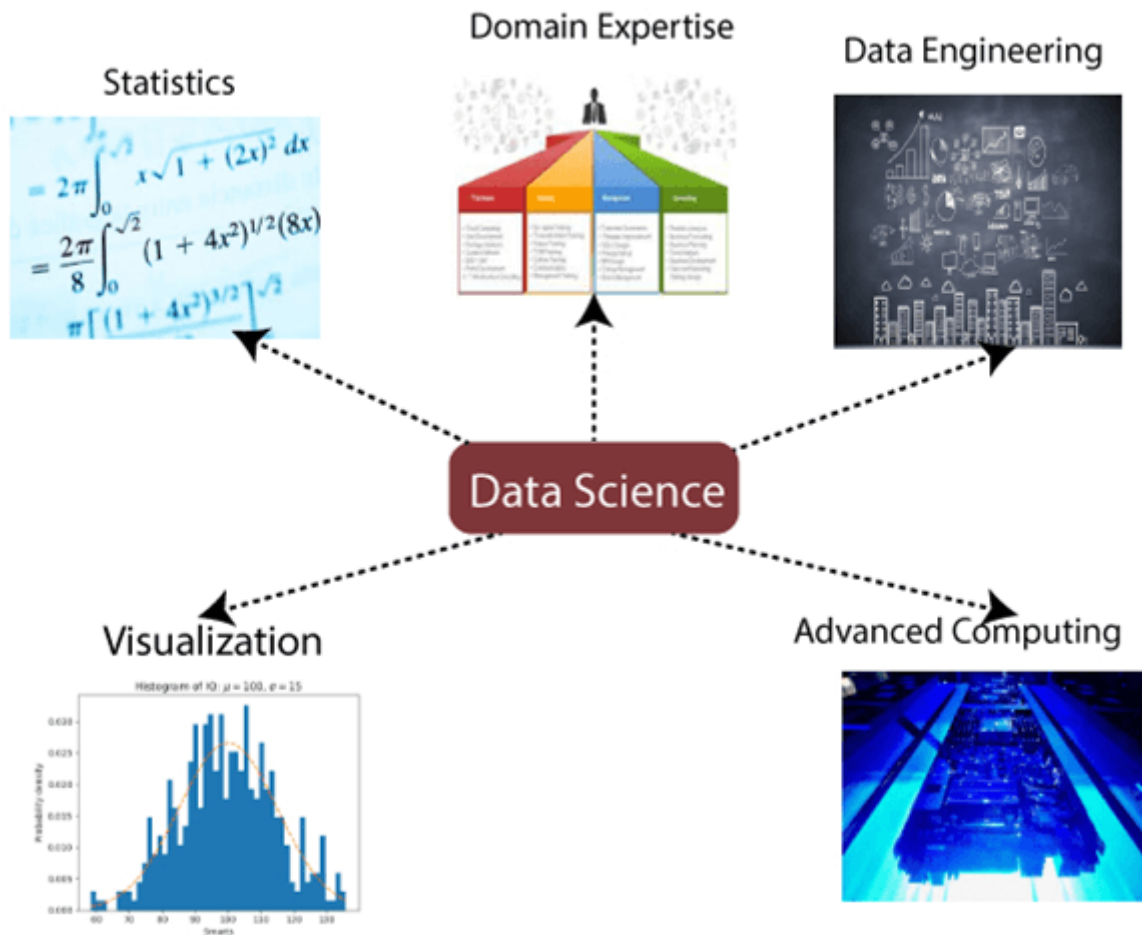
- **Databases:** The depth understanding of Databases such as SQL, is essential for data science to get the data and to work with data.

What's the Career Outlook for a Data Scientist?

With the right qualifications, you'll enjoy a bright career outlook as a data scientist. The demand for individuals with these skills will continue to increase, and those already in data science roles are sure to see their salaries increase in the future. As demand for skilled professionals to fill these positions increases, the salaries offered are also likely to increase—even those holding the lowest-paying data scientist jobs will still make a nice living: salaries range from \$60,000 and up.

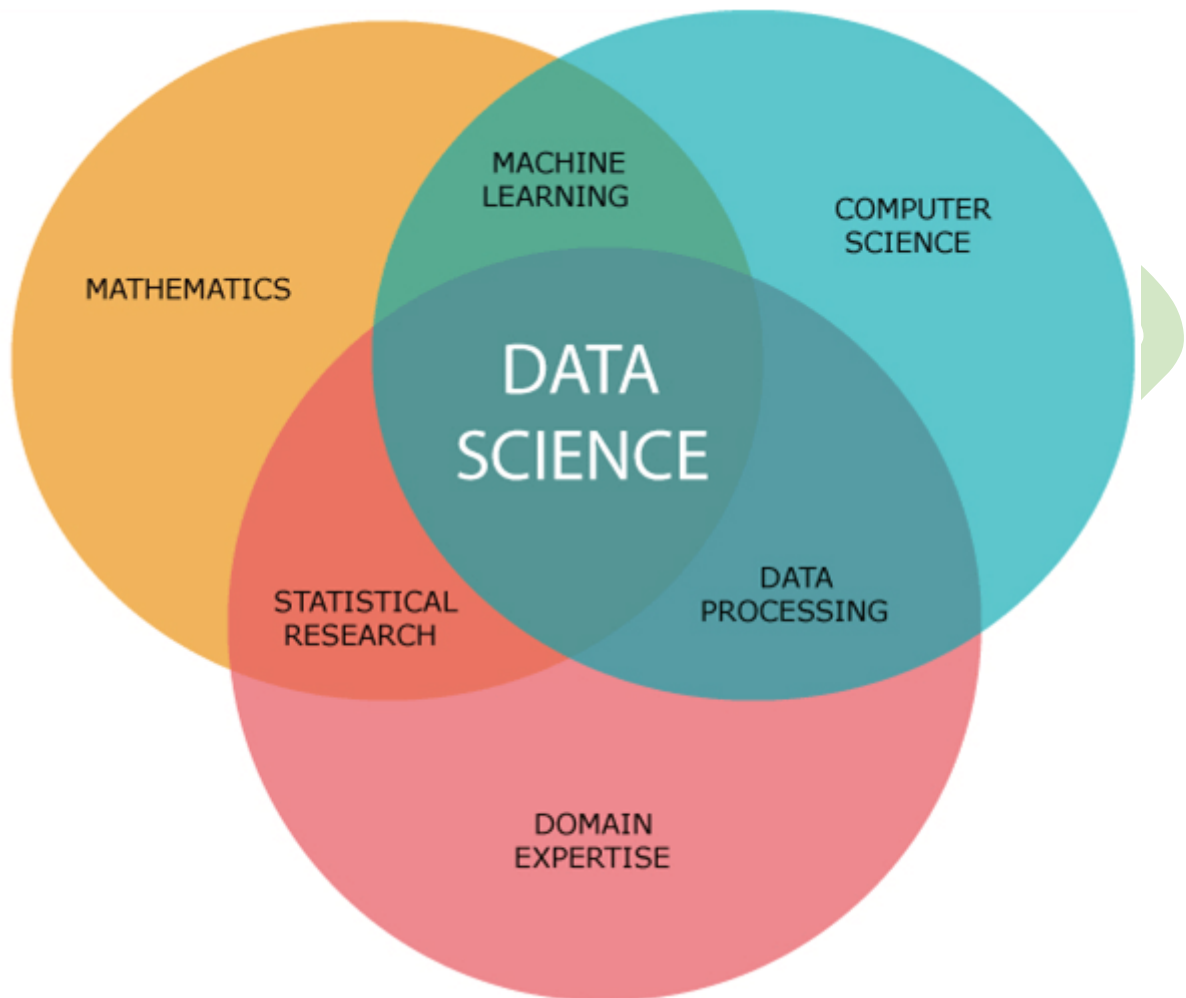


Data Science Components:



The main components of Data Science are given below:

- 1. Statistics:** Statistics is one of the most important components of data science. Statistics is a way to collect and analyze the numerical data in a large amount and finding meaningful insights from it.
- 2. Domain Expertise:** In data science, domain expertise binds data science together. Domain expertise means specialized knowledge or skills of a particular area. In data science, there are various areas for which we need domain experts.
- 3. Data engineering:** Data engineering is a part of data science, which involves acquiring, storing, retrieving, and transforming the data. Data engineering also includes metadata (data about data) to the data.
- 4. Visualization:** Data visualization is meant by representing data in a visual context so that people can easily understand the significance of data. Data visualization makes it easy to access the huge amount of data in visuals.
- 5. Advanced computing:** Heavy lifting of data science is advanced computing. Advanced computing involves designing, writing, debugging, and maintaining the source code of computer programs.



6. Mathematics: Mathematics is the critical part of data science. Mathematics involves the study of quantity, structure, space, and changes. For a data scientist, knowledge of good mathematics is essential.

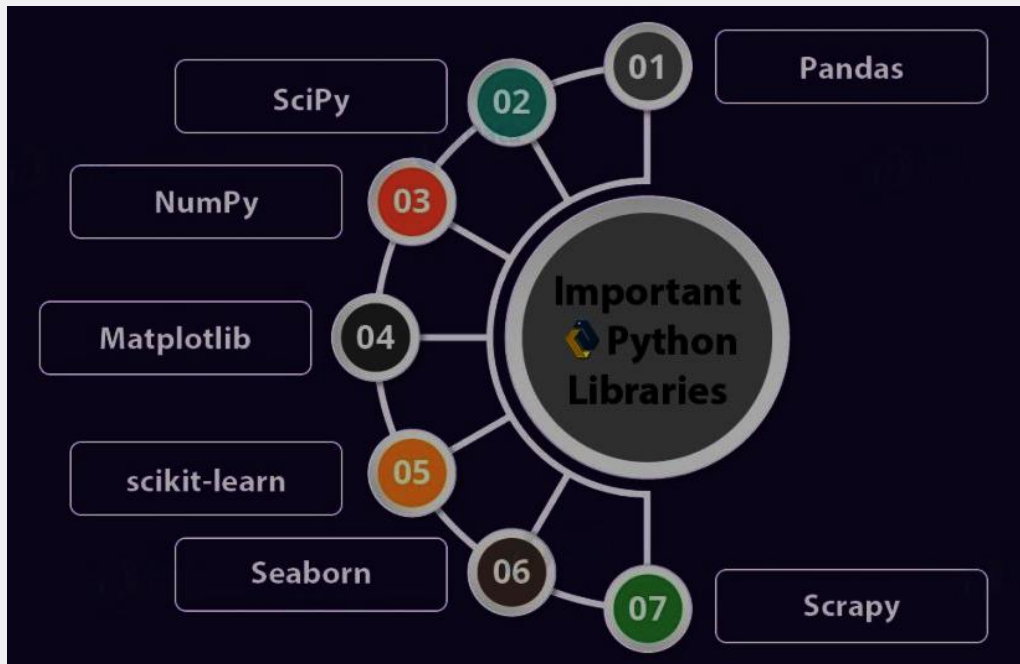
7. Machine learning: Machine learning is backbone of data science. Machine learning is all about to provide training to a machine so that it can act as a human brain. In data science, we use various machine learning algorithms to solve the problems.

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

Data Science – Python Libraries

For carrying out data analysis and other scientific computation, you will need any of the following libraries:



Data Science Tutorial – Data Science Libraries

a. Pandas

Pandas help us with munging and preparing data; it is great for operating on and maintaining structured data.

b. SciPy

SciPy (Scientific Python) stands on top of NumPy. With this library, we can carry out functionality like Linear Algebra, Fourier Transform, Optimization, and many others.

c. NumPy

NumPy (Numerical Python) is another library that lets us deal with features like linear algebra, Fourier transforms and advanced random number capabilities. One very import feature of NumPy is the n-dimensional array.

d. Matplotlib

Matplotlib will let you plot different kinds of graphs. These include pie charts, bar graphs, histograms, and even heat plots.

e. Scikit-learn

Scikit-learn is great for machine learning. It will let you statistically model and implement machine learning. The tools for these include clustering, regression, classification, and dimensionality reduction.

f. Seaborn

Seaborn is good with statistical data visualization. Making use of it, we can create useful and attractive graphics.

g. Scrappy

Scrappy will let you crawl the web. It begins on a home page and gets deeper within a website for information.