



Share on your Social  
Media



# Data Science Tutorial

Published On: August 9, 2024

## Data Science Tutorial

Combining statistics, data analysis, and machine learning, data science is an interdisciplinary field that analyzes data and draws conclusions and information from it. In this data science tutorial, you will learn everything needed to get started with your learning journey.

[Download Data Science Tutorial PDF](#)

## Introduction to Data Science

Data science is the study of data collection, processing, and interpretation. Finding patterns in data through analysis and forecasting the future are the goals of data science. We cover the following in this data science tutorial:

- Overview of Data Science

## Featured Articles

 **Want to know more about becoming an expert in IT?**

[Click Here to Get Started](#) >>

100%  
Placement  
Assurance

AUTHORISED  
CERTIFICATION  
PARTNER

IBI

## Related Courses at SLA

## Related Posts



### Tableau Developer Salary in Chennai

Published On: October 12, 2024

Introduction A Tableau Developer designs, develops, and maintains dashboards and visualizations using Tableau software. Key...

Quick Enquiry



- Components of Data Science
- Data Science Lifecycle
- Popular Tools for Data Science
- DataFrame in Data Science
- Data Science Functions
- Data Preparation

## Overview of Data Science

Today, data science is applied across a wide range of industries, including manufacturing, banking, consulting, and healthcare.

- Businesses can use data science to make better decisions, forecast future events, and identify patterns in data to uncover hidden information.
- A data scientist needs to be acquainted with machine learning, statistics, R or Python programming, databases, and mathematics.
- A data scientist's role is to look for patterns in the data. He or she needs to arrange the data in a standard way before they can start looking for trends.

## Data Science Interview Questions and Answers

## Applications of Data Science

**Image and Speech Recognition:** Data science is used for speech and image recognition. When you submit a picture on Facebook and begin receiving recommendations to tag pals in it.

**Healthcare:** Data science offers numerous advantages to the healthcare industry. Medical image analysis, virtual medical bots, medication discovery, tumor detection, and other applications are using data science.

**Gaming Industry:** Machine learning algorithms are being used in the gaming industry more and more daily.

### VMware Tutorial for Cloud Computing Aspirants

Published On: October 12, 2024

VMware Tutorial for Cloud Computing Aspirants VMware software allows you to run a virtual machine...

### VBA Macros Tutorial for Beginners

Published On: October 10, 2024

VBA Macros Tutorial for Beginners VBA macros are programs that automate repetitive operations in Microsoft...

### VB.Net Tutorial for Beginners

Published On: October 10, 2024

VB.Net Tutorial for Beginners Visual Basic is an easy-to-learn programming language that is type-safe. Get...

**Risk Detection:** The finance industry has long struggled with fraud and loss risk, but data science can help turn this around.

**Transport:** The transportation sector is also developing self-driving automobiles with data science technology. Reducing traffic accidents will be simple with self-driving automobiles.

**Internet Search:** We use a variety of search engines, like Google, Yahoo, Bing, Ask, and others, to look up information on the internet.

**Recommendation systems:** Data science technology is used when you search for something on Amazon and start receiving recommendations for related things.

## Major Components of Data Science

**Statistics:** One of the key elements of data science is statistics. A significant amount of numerical data can be gathered and analyzed, and useful insights can be drawn from it using statistics.

**Domain Expertise:** Domain expertise is the glue that holds data science together. Domain expertise refers to specific knowledge or abilities in a given field. We require domain specialists in many areas related to data science.

**Data Engineering:** Data science encompasses data engineering as well, which deals with gathering, storing, retrieving, and altering data. Metadata, or information about data, is also a part of data engineering.

**Data Visualization:** It is the process of presenting information in a way that makes it easier for others to see its importance. Accessing the vast amount of data in visual form is made simple by data visualization.

**Advanced Computing:** This refers to the computationally intensive aspects of data science.

Writing, developing, debugging, and maintaining computer programs' source code are all part of advanced computing.

**Mathematics:** A crucial component of data science is mathematics. Studying quantity, structure, space, and changes are all part of mathematics. Strong mathematical skills are crucial for a data scientist.

**Machine Learning:** The goal of machine learning is to train a machine to function like a human brain. Various machine learning techniques are used in data science to solve challenges.

[\*\*Download Data Science Syllabus PDF\*\*](#)

## Data Science Lifecycle

The following is the data science life cycle:

**Discovery:** Asking the appropriate questions is part of the discovery phase, which is the first stage. Any data science project should begin with a determination of the fundamental prerequisites, project budget, and project priorities.

We must ascertain all project requirements at this phase, including personnel, technology, time, data, and an ultimate objective, before we can formulate the business problem at the first hypothesis level.

**Data preparation:** Another name for data preparation is data munging. During this stage, the following duties must be completed:

- Data cleaning
- Data Reduction
- Data integration
- Data transformation

We can easily use this data for our subsequent operations after completing all of the aforementioned activities.

**Model Planning:** During this stage, we must identify

the different approaches and strategies for establishing the relationship between the input variables.

Using a variety of statistical formulas and visualization tools, we will apply exploratory data analytics (EDA) to comprehend the relationships between variables and determine what information the data might provide.

Typical instruments for model planning include:

- SQL Analysis Services
- R
- SAS
- Python

**Model building:** During this stage, the model-building procedure is initiated. To facilitate training and testing, datasets will be created.

To construct the model, we'll use a variety of methods, including association, classification, and clustering.

Here are a few typical tools used in model building:

- SAS Enterprise Miner
- WEKA
- SPCS Modeler
- MATLAB

**Operationalize:** We will provide the project's final reports during this phase, along with technical documents, code, and briefings. Before the full deployment, this phase gives you a comprehensive overview of the performance of the entire project as well as additional small-scale components.

**Communicate findings:** During this stage, we will determine whether or not the initial phase's goal was met. We will share the results and outcome with the business team.

## Popular Tools for Data Science

The tools needed for data science include the following:

**Data Analysis Tools:** MATLAB, Excel, R, Python, Statistics, SAS, Jupyter, R Studio, and RapidMiner.

**Data Warehousing Tools:** ETL, SQL, Hadoop, Informatica/Talend, AWS Redshift, and data warehousing.

**Data Visualization Tools:** Tableau, R, Jupyter, and Cognos.

**Machine Learning Tools:** Azure ML Studio, Mahout, Spark, and other machine learning technologies.

## Data Scientist Salary

## Creating DataFrame with Python

An organized data representation is called a data frame. Let's build a hypothetical data frame with three columns and five rows of numbers:

### Example

```
import pandas as pd

d = {'col1': [1, 2, 3, 4, 7], 'col2': [4, 5, 6, 9, 5], 'col3': [7, 8, 12, 1, 11]}

df = pd.DataFrame(data=d)

print(df)
```

### Explanation

- Bring in the Pandas library (pd).
- Create a variable called d and define data with columns and rows.
- Make a data frame by utilizing the function, "pd.DataFrame()."
- Five rows and three columns make up the data frame.
- Use the print() method to produce the data frame.

## Data Science Functions

The `mean()`, `max()`, and `min()` functions are frequently utilized in data science operations.

## The `mean()` function

The average value of an array can be found using the NumPy `mean()` method.

### Example

```
import numpy as np

Calorie_burnage = [240, 250, 260, 270, 280, 290, 300, 310, 320, 330]

Average_calorie_burnage = np.mean(Calorie_burnage)

print(Average_calorie_burnage)
```

## The `max()` function

The highest value in an array can be found using the Python `max()` method.

```
Average_pulse_max = max(80, 85, 90, 95, 100, 105, 110, 115, 120, 125)

print (Average_pulse_max)
```

## The `min()` function

To determine which value in an array is the lowest, use the Python `min()` method.

```
Average_pulse_min = min(80, 85, 90, 95, 100, 105, 110, 115, 120, 125)

print (Average_pulse_min)
```

## Data Preperation

A data scientist must first extract the data and clean it up before beginning any analysis.

### Extract and Read Data With Pandas

Data needs to be imported or extracted before it can be evaluated.

### Example

To import a CSV file containing the health data, we utilize the `read_csv()` function:

```
import pandas as pd

health_data = pd.read_csv("data.csv", header=0, sep=",")

print(health_data)
```

## Explanation

- Bring the Pandas library in.
- Put `health_data` as the data frame's name.
- `header=0` indicates that the variable names' headers can be located in the first row (keep in mind that in Python, 0 denotes the first row).
- `Sep=","` indicates that the values are separated by ",". This is a result of the file type that we are employing comma-separated values, or CSV.

**Data Science Training**

## Data Categories

Knowing the kinds of data we are working with is also necessary for data analysis.

Data can be divided into two primary groups:

- **Quantitative Data:** It is information that can be quantified or expressed as a number. It can be categorized into two smaller groups:
  - **Discrete information:** Example: The number of pupils in a class or the number of goals in a soccer match counts as "whole" numbers.
  - **Constant data:** There is no limit to the precision of numbers. For example, a person's weight, shoe size, and temperature.
- **Qualitative Data:** Cannot be quantified or put into numerical form. consists of the following two subcategories:
  - **Nominal data:** for instance, race, gender, and hair color
  - **Typical data:** For instance, financial



position (poor, middle, high) and  
academic grades (A, B, C)

You may choose the appropriate analysis  
technique for your data by understanding what kind  
of data you have.

## Conclusion

We hope this data science tutorial gives you the  
basic understanding of where to start your data  
science learning journey. Master in data science  
skills by enrolling in our [data science training in  
Chennai.](#)

Share on your Social  
Media



## Softlogic Academy

## Softlogic Systems

### KK Nagar [Corporate Office]

No.10, PT Rajan Salai, K.K. Nagar, Chennai  
– 600 078.

**Landmark:** Karnataka Bank Building

**Phone:** [+91 86818 84318](tel:+918681884318)

**Email:** [enquiry@softlogicsys.in](mailto:enquiry@softlogicsys.in)

**Map:** [Google Maps Link](#)

### OMR

No. E1-A10, RTS Food Street  
92, Rajiv Gandhi Salai (OMR),  
Navalur, Chennai – 600 130.

## Navigation

---

- [About Us](#)
- [Blog Posts](#)
- [Careers](#)
- [Contact](#)
- [Placement Training](#)
- [Corporate Training](#)
- [Hire With Us](#)
- [Job Seekers](#)
- [SLA's Recently Placed Students](#)
- [Reviews](#)
- [Sitemap](#)

## Important Links

---

- [Disclaimer](#)
- [Privacy Policy](#)
- [Terms and Conditions](#)

**Landmark:** Adj. to AGS Cinemas

**Phone:** [+91 89256 88858](tel:+918925688858)

**Email:** [info@softlogicsys.in](mailto:info@softlogicsys.in)

**Map:** [Google Maps Link](#)

## Courses

---

Python

Software Testing

Full Stack Developer

Java

Power BI

Clinical SAS

Data Science

Embedded

Cloud Computing

Hardware and Networking

VBA Macros

Mobile App Development

DevOps

## Social Media Links

---



## Review Sources

---

Google

Trustpilot

Glassdoor

Mouthshut

Sulekha

Justdial

Ambitionbox

Indeed

Software Suggest

Sitejabber