

Share on your Social
Media



Data Science Full Stack Tutorial

Published On: July 12, 2024

Data Science Full Stack Tutorial

The exciting topic of data science is knowledge and insight extraction from data. We have designed a comprehensive data science full-stack tutorial to help you grasp every subject related to this field.

[Datat Science Full Stack Tutorial PDF](#)

Introduction to Data Science Full Stack

The Data Science Full Stack discusses AI and data science and gives a broad overview of their uses and significance across a range of sectors.

It includes an understanding of exploratory data analysis (EDA), data pretreatment procedures, data gathering strategies, statistical analysis, deep learning models, machine learning methods, and AI applications.

The following are the important things to learn to become a data science full-stack expert:

- Python programming basics
- Statistics
- SQL
- Exploratory Data Analysis and Machine Learning
- Power BI Basics

Featured Articles



Want to know
more about
becoming an
expert in IT?

[Click Here to Get Started](#)

100%
Placement
Assurance

AUTHORISED
CERTIFICATION
PARTNER

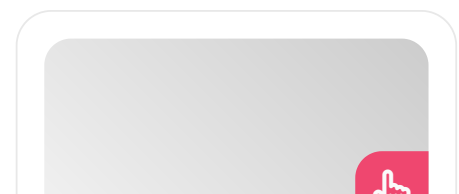
IBM

Quick Enquiry

Related Courses at SLA

- ➔ **Data Science Fullstack Training in OMR**
- ➔ **Data Science Fullstack Training in Chennai**
- ➔ **Data Science Full Stack Course**

Related Posts



- Deep Learning
- Natural Language Processing
- Computer Vision

Python Programming Basics

Data Science: Big data, AI, machine learning, predictive analytics, and decision-making are all greatly impacted by data science. DSM enables efficient analysis of big datasets through the marriage of mathematical models, algorithms, and subject-matter expertise. Working with data is only one aspect of data management! It is all about making sense of the data and constructively applying the results.

Importance of Python for Data Science

Python has been in high demand for the past few years, and a recent poll indicated that this is still the case. According to the TIOBE and PYPL indexes, Python is the top programming language. Nonetheless, there are five specific justifications for this.

- Easy to learn
- Cross-platform
- Portable
- Extensive Library
- Community Support

Data scientists favor Python for the following modules when it comes to application areas:

- Data Analysis
- Data Visualizations
- Machine Learning
- Deep Learning
- Image processing
- Computer Vision
- Natural Language Processing (NLP)

High-level programming languages like Python 3 are widely used and have a wide range of applications. You should be familiar with the following Python 3 fundamentals:

MongoDB Tutorial for Beginners

Published On: September 25, 2024

MongoDB Tutorial for Beginners In real-world contexts like e-commerce, finance, healthcare, journalism, and social networks,...



Mobile App Development Tutorial

Published On: September 25, 2024

Mobile App Development Tutorial Businesses looking to grow and scale can benefit greatly from mobile...



Microsoft Azure Tutorial for Beginners

Published On: September 24, 2024

Microsoft Azure Tutorial for Beginners We can use a wide range of services from Microsoft...



MEAN Stack Tutorial

for Beginners

Published On: September 24, 2024

MEAN Stack Tutorial for Beginners 'M' refers to MongoDB, 'E' for Express, 'A' for Angular,...

Variables: To create a variable in Python 3, provide a name and a value. For instance, the expression `x = 5` defines a variable named `x` and gives it the value 5.

Data Types: Integers, floats, strings, booleans, lists, tuples, and dictionaries are among the built-in data types that Python 3 offers.

Data items are categorized or classified using Python data types. It identifies the kind of value that represents the different operations that can be performed on a given collection of data. The standard or built-in, data types in Python are as follows:

- Numeric
- Sequence Type
- Boolean
- Set
- Dictionary
- Binary Types (memoryview, bytearray, bytes)

Operators: Python 3 has many different operators that it supports: logical operators (and, or, not), comparison operators (`>`, `<`, `==`, `!=`), and arithmetic operators (`+`, `-`, `*`, `/`).

You should learn the following

- **Arithmetic Operators:** Python has seven different arithmetic operators. They are addition, subtraction, multiplication, division, floor division, modulus, and exponentiation.
- **Comparison Operators:** The different types of comparison operators in Python are equality, inequality, greater than sign, less than sign, greater than or equal, and chaining comparison operators.
- **Logical Operators:** Python conditional statements (either True or False) use logical operators. They perform logical operations in AND, OR, and NOT.
- **Bitwise Operators:** Python bitwise operators

are used to perform bitwise operations on integers. "Bitwise operators" are operations on single bits or related pairs of bits that are carried out after the integers have been converted to binary. Subsequently, the outcome is returned in decimal notation.

- **Assignment Operators:** Operators provide the task variables. The value of the right side of the expression is assigned to the operand on the left side using this operator.
- **Identity Operators:** If two objects truly belong to the same data type and share the same memory address, they can be compared using the Python Identity Operators.
- **Membership Operators:** The membership operators in Python check if an object, like a text, list, or tuple, is a part of the sequence. Two membership operators are available in Python to verify or check if a value is a member.

Control flow statements: While loops, for loops, and if-else statements are just a few of the control flow statements that Python 3 offers. You can manage how your code is executed with these statements.

Functions: The `def` keyword in Python 3 is used to create functions.

Example: `def my_function(x):` generates a function with the name `my_function` and a single parameter named `x`.

Input and output: In Python 3, you may obtain user input by using the `input()` function and output text to the console by using the `print()` function.

Modules are collections of variables and functions that may be imported and utilized in other Python programs. Python 3 supports modules. The `import` keyword can be used to import modules.

[Data Science Full Stack Syllabus PDF](#)

Understanding of Statistics for Data Science Full-Stack

Data scientists must comprehend the basic ideas of descriptive statistics and probability theory, which include the ideas of probability distribution, statistical significance, hypothesis testing, and regression testing.

Machine learning also benefits from the use of Bayesian thinking, whose main ideas include maximum likelihood, priors and posteriors, and conditional probability.

- **Descriptive Statistics:** Using descriptive statistics, one can examine and pinpoint a data set's fundamental characteristics. In addition to providing a means of visualizing the data, descriptive statistics include summaries and descriptions of the data.
- **Probability Theory:** A field of mathematics called probability theory calculates the chance that a random event will occur. The probability of a genetic condition, actuarial charts for insurance companies, political polling, and clinical trials are only a few applications of statistical formulas connected to probability.
- **Statistical Features:** Data scientists frequently begin their investigations of data using statistical features. The statistical features (PDF, 21.6 MB) comprise data organization, quartile identification, and median, minimum, and maximum value determination.
- **Probability Distributions:** A probability distribution is the set of all potential outcomes for a random variable along with the associated probability values between zero and one. Data scientists use this to determine how likely it is to receive particular values or events.
- **Dimensionality Reduction:** The technique of lowering the dimensions of your data set is known as dimensionality reduction (PDF, 751 KB). There are several possible advantages to

dimensionality reduction, such as less data to store, quicker computing, fewer redundancies, and more precise models.

- **Over- and Under-Sampling:** When there is not enough data currently available, oversampling is used. When a portion of the data is over-represented, under-sampling is applied.
- **Bayesian Statistics:** In the Bayesian paradigm, the parameters are assigned a probability distribution, referred to as the prior distribution, which expresses the existing information about the parameters.

Data Science Full Stack Interview **Questions**

SQL for Data Science Full Stack

The data included in databases can be subjected to a variety of actions using SQL (Structured Query Language), including updating, removing, adding, and changing tables and views.

Additionally, SQL is the industry standard for big data platforms that employ SQL as their primary relational database API.

A database management system based on servers is called MySQL. Multiple databases can exist on a single MySQL server. There are two steps involved in building a MySQL database:

- Establish a connection with a MySQL host.
- To create the database and process the data, run different queries.

SQLite: A serverless database is SQLite. It writes and reads information to files. That implies that we can do database operations such as MySQL and PostgreSQL without even installing and running a SQLite server!

Understanding of Exploratory Data Analysis for Data Science Full Stack

In data science initiatives, exploratory data analysis, or EDA, is an essential first step. It refers to the process of examining and analyzing record sets to recognize their defining characteristics, find patterns, find outliers, and determine relationships between variables.

It involves analyzing and visualizing data to understand its essential features, uncover patterns, and identify relationships between variables.

The essential elements of EDA consist of:

- Distribution of Data
- Graphical Representations
- Outlier Detection
- Correlation Analysis
- Handling Missing Values
- Summary Statistics
- Testing Assumptions

Univariate Analysis: To comprehend the internal structure of a single variable, univariate analysis is used. Its main goals are to characterize the data and identify patterns within a single characteristic. It includes the following:

- Histograms
- Box Plots
- Bar Charts
- Summary Statistics

Bivariate Analysis: Examining the relationship between variables is part of bivariate evaluation. It makes it possible to identify dependencies, correlations, and relationships between sets of variables. It includes the following:

- Scatter Plots
- Correlation Coefficient
- Cross-tabulation
- Line Graphs
- Covariance

Multivariate Analysis: The associations between

two or more variables in the dataset are examined using multivariate analysis. It seeks to comprehend the interplay between variables, which is essential to the majority of statistical modeling methods. It includes the following:

- Pair Plots
- Principal Component Analysis

Understanding of Machine Learning for Data Science Full Stack

Developing end-to-end machine learning systems with practical applications is known as full-stack data science development.

This covers every step of the process, from problem definition and data collection to model construction and training, production deployment, and post-use monitoring and maintenance.

The Process of Developing a Full-Stack Machine Learning System:

- Definition of the Problem
- Data Collection and Preparation using tools like Apache Spark, Apache Hive, Hadoop, etc.
- Model Building and Training using tools like TensorFlow, PyTorch, scikit-learn, etc.
- Model Deployment using tools like Flask, Django, Google Cloud AI Platform, Amazon SageMaker, etc.
- Model Monitoring and Maintenance using tools like TensorBoard, Google Cloud Monitoring, Amazon CloudWatch, etc.

Data Science Full Stack Developer
Salary

Understanding of Power BI Basics for Data Science Full Stack

With Power BI, a business intelligence tool, you can connect to several data sources, see the data in dashboards and reports, and share them with

whomever you choose.

Power BI consists of three primary components:

Power BI Desktop

It is a free desktop program for creating and designing reports.

The major elements of Power BI Desktop are:

Ribbon: The majority of the options and controls required to produce the report are located on the upper ribbon.

Views: The terms “views” refer to the data view, report view, and model view.

Canvas: The primary design area, or canvas, is where additional items and visualizations are applied.

Page selector: used to go to different report pages.

Filters: Fields can be placed here to create filters for the data.

Visualizations: This is a list of all the visualizations that are accessible.

Fields: The tables and fields that are present in the data model are included in this section.

Power BI Service

The online publishing tool for viewing and sharing dashboards and results.

You can publish your report to your Power BI workspace if you’re satisfied with it. You must first log into Power BI and choose Publish from the ribbon to accomplish this.

The report will be published to the Power BI Service once you select a workspace. After logging in, go to the workspace where your report was published in Power BI.

- Data
- Report
- Dashboard

Power BI Mobile Apps

They are used to examine dashboards and reports while on the go. Among the main attributes of the Power BI mobile apps are:

- Access to Reports and Dashboards
- Offline Access
- Interactivity
- Alerts and Notifications
- Collaboration
- Security

Understanding of Deep Learning for Data Science Full Stack

A branch of machine learning called “deep learning” teaches a computer to do tasks that humans do, like speech recognition, picture identification, and prediction.

It enhances the capacity to use data to categorize, identify, detect, and characterize. The hype around artificial intelligence (AI) is partly to blame for the present interest in deep learning.

Nowadays, many advancements are promoting deep learning:

- Deep learning techniques now perform better due to **analytical advancements**.
- **The accuracy** of deep learning models has increased due to new machine learning techniques.
- **Neural networks** have been redesigned into classes that operate well for tasks like picture classification and text translation.
- More data, such as text from social media, medical notes, transcripts of investigations, and streaming data from the **Internet of Things**, is accessible to create neural networks with numerous deep layers.

- We now have access to amazing amounts of computer power because of advancements in distributed **cloud computing and graphics** processing units.
- To train deep algorithms via deep learning, this amount of **processing power** is required.

Data Science Training

Understanding of Natural Language Processing

NLP includes techniques for producing and evaluating human language, which facilitates the use of chatbots and sentiment analysis in applications.

NLP is used in data science processing for the following reasons:

- Automate Routine Tasks
- Improved Search
- Search Engine Optimization
- Analyzing and organizing large document collections
- Social media analytics
- Market Insights
- Moderating content.

Some of the NLP techniques are as follows:

- Tokenization
- Bag-of-words models
- Stop word removal
- Stemming and lemmatization
- Part-of-speech tagging and syntactic parsing.

The most popular NLP libraries that are used in data science processing are as follows:

- TensorFlow
- PyTorch
- HuggingFace
- Spark NLP
- SpaCy NLP

Understanding of Computer Vision for Data Science Full Stack

The goal of the artificial intelligence (AI) field of study in computer vision is to enable computers to intercept and extract information from images and videos in a way that is comparable to human vision.

It entails creating methods and algorithms to interpret the visual world and extract relevant information from visual inputs.

Some of the examples of computer vision are:

Facial recognition: visual analysis is used to identify people.

Self-driving cars: Self-driving cars employ computer vision to navigate and steer clear of obstacles.

Robotic Automation: Automating tasks for robots and enabling them to make judgments based on visual input is known as robotic automation.

Medical Anomaly Detection: Finding anomalies in medical imaging to aid in a more accurate diagnosis is known as "medical anomaly detection."

Sports performance analysis: monitoring the motions of athletes to assess and improve performance.

Manufacturing Fault Detection: Detecting flaws in items during the production process is known as manufacturing fault detection.

Agriculture Monitoring: Using visual data, agricultural monitoring tracks crop growth, livestock health, and meteorological conditions.

Conclusion

This Data Science Full Stack tutorial covers everything you need to learn to become a data

scientist. Learn from scratch with complete hands-on exposure to real-time projects by enrolling in our **Data Science Full Stack 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

Map: [Google Maps Link](#)

OMR

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

Landmark: Adj. to AGS Cinemas

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

Email: info@softlogicsys.in

Map: [Google Maps Link](#)

Courses

Python

Software Testing

Navigation

[About Us](#)

[Blog Posts](#)

[Careers](#)

[Contact](#)

[Placement Training](#)

[Corporate Training](#)

[Hire With Us](#)

[Job Seekers](#)

[SLA's Recently Placed Students](#)

[Reviews](#)

[Sitemap](#)

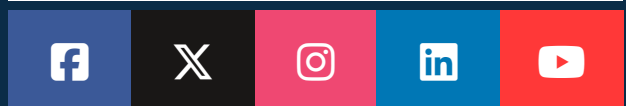
Important Links

[Disclaimer](#)

[Privacy Policy](#)

[Terms and Conditions](#)

Social Media Links



Full Stack Developer

Java

Power BI

Clinical SAS

Data Science

Embedded

Cloud Computing

Hardware and Networking

VBA Macros

Mobile App Development

DevOps

Review Sources

Google

Trustpilot

Glassdoor

Mouthshut

Sulekha

Justdial

Ambitionbox

Indeed

Software Suggest

Sitejabber

Copyright © 2024 - Softlogic
Systems. All Rights Reserved

SLA™ is a trademark of Softlogic Systems, Chennai.
Unauthorised use prohibited.