Gain expertise from experienced DATA SCIENCE professionals.

DIGI SAMURAI

DATA SCIENCE BOOTCAMP

Why Data Science?

Harnessing the power of data is no longer a luxury but a necessity in today's data-driven world, where insights are sought after and challenges abound.

A career in data science offers numerous advantages, including high demand, the opportunity to address the skills gap, diverse roles, challenging work, the ability to derive actionable insights from data, and the ability to contribute toward a data-driven world. Choosing a data science career can provide both professional satisfaction and the opportunity to make a significant impact in today's increasingly data-centric society.

Explore the Fascinating World of Data Science and Transform Your Career

Prepare yourself with critical knowledge to analyze data effectively, derive actionable insights, and navigate the data-driven world confidently and skillfully.

COURSE OVERVIEW

Statistics for Data Science

Module 1: Introduction to Statistics and Data

1. What are Statistics and why it is important in Data Science
2. Types of Data: Categorical, Numerical, Discrete, Continuous
3. Data Collection Methods: Sampling, Surveys, Experiments

Module 2: Python for Data Science

1. Measures of Central Tendency: Mean, Median, Mode
2. Measures of Dispersion: Range, Variance, Standard Deviation
3. Percentiles and Quartiles
4. Visualization of Data: Histograms, Box Plots, Scatter

Module 3: Probability

1. Introduction to Probability
2. Probability Distributions: Discrete and Continuous
3. Probability Rules: Addition, Multiplication, Complement
4. Conditional Probability and Bayes' Theorem

Module 4: Probability Distributions

1. Normal Distribution and Standard Normal Distribution
2. Binomial Distribution
3. Poisson Distribution
4. Central Limit Theorem and Sampling Distribution

Module 5: Sampling and Estimation

1. Population vs. Sample
2. Sampling Methods: Random Sampling, Stratified Sampling, Cluster Sampling
3. Point Estimation and Interval Estimation
4. Confidence Intervals

Module 6: Hypothesis Testing

1. Null and Alternative Hypotheses
2. Significance Level and p-values
3. Types of Errors: Type I and Type II
4. Common Hypothesis Tests: t-tests, Chi-Square tests

SQL and DBMS

Module 1: Introduction to DBMS and Relational Databases

1. What is a Database Management System (DBMS)?
2. Types of DBMS: Relational, NoSQL, NewSQL, etc.
3. Relational Model and its Components
4. Key Terminology: Tables, Rows, Columns, Relationships

Module 2: SQL Basics

1. Introduction to SQL (Structured Query Language)
2. SQL Data Types: Text, Numeric, Date/Time
3. Creating Databases and Tables
4. Basic CRUD Operations: INSERT, SELECT, UPDATE, DELETE

Module 3: Data Retrieval with SQL

1. SELECT Statement: Retrieving Data from a Single Table
2. Filtering Data with WHERE Clause
3. Sorting Data with ORDER BY
4. Limiting Results with LIMIT

Module 4: Filtering and Sorting

1. Filtering Rows: WHERE Clause and Logical Operators
2. Sorting Rows: ORDER BY Clause
3. Combining Filtering and Sorting

Module 5: Aggregate Functions

1. Understanding Aggregate Functions: COUNT, SUM, AVG, MIN, MAX
2. GROUP BY Clause for Grouping Data
3. HAVING Clause for Filtering Grouped Data

Module 6: Data Manipulation

1. Updating Data: UPDATE Statement
2. Deleting Data: DELETE Statement
3. Truncating Tables
4. Inserting Data: INSERT INTO Statement

Module 7: Data Integrity and Constraints

1. Primary Keys and Foreign Keys
2. UNIQUE Constraint
3. NOT NULL Constraint
4. CHECK Constraint

Module 8: Joins and Relationships

1. Understanding Relationships: One-to-One, One-to-Many, Many-to-Many
2. INNER JOIN and OUTER JOIN (LEFT, RIGHT, FULL)
3. Joining Multiple Tables
4. Self-Joins

Python Basics

Module 1: Introduction to Python Programming

1. Introduction to Python and its Features
2. Installing Python and Setting Up Development Environment
3. Basic Syntax: Variables, Data Types, Print Statements
4. Comments and Basic Input/Output

Module 2: Control Structures

1. Conditional Statements: if, elif, else
2. Loops: for and while loops
3. Loop Control Statements: break, continue

Module 3: Functions and Modules

1. Defining and Calling Functions
2. Function Parameters and Return Values
3. Introduction to Modules and Importing

Module 4: Data Structures

1. Lists: Creating, Indexing, Slicing, Modifying
2. Tuples: Creating, Accessing, Immutability
3. Dictionaries: Creating, Accessing, Modifying

Data Handling using Python

Module 1: Introduction to NumPy and pandas

1. Overview of NumPy and pandas
2. Installing and Importing NumPy and pandas
3. Basic Data Structures: Arrays (NumPy) and DataFrames (pandas)

Module 2: NumPy Fundamentals

1. Creating NumPy Arrays
2. Indexing and Slicing Arrays
3. Basic Array Operations: Arithmetic, Broadcasting

Module 3: Array Manipulation with NumPy

1. Reshaping and Resizing Arrays
2. Stacking and Splitting Arrays
3. Universal Functions (ufuncs) in NumPy

Module 4: Introduction to pandas DataFrames

1. Creating DataFrames
2. Indexing and Selecting Data
3. Basic DataFrame Operations: Filtering, Sorting
4. Reading and Writing CSV Files
5. Reading and Writing Excel Files
6. Working with SQL Databases

Module 5: Data Cleaning and Preprocessing with pandas

1. Handling Missing Values
2. Removing Duplicates
3. Data Transformation: Renaming, Mapping

Module 6: Combining DataFrames

1. Concatenation of DataFrames
2. Merging and Joining DataFrames

Module 7: Grouping and Aggregation

1. Grouping Data with Group By
2. Applying Aggregation Functions
3. Multi-level Indexing

Data Visualization using Python

Module 1: Introduction to Data Visualization

1. Why Data Visualization Matters in Data Science
2. Types of Data Visualizations: Exploratory, Explanatory
3. Visualization Libraries: Matplotlib, Seaborn

Module 2: Getting Started with Matplotlib

1. Introduction to Matplotlib
2. Basic Line Plots and Scatter Plots
3. Customizing Plot Appearance: Titles, Labels, Legends

Module 3: Creating Bar and Pie Charts

1. Bar Plots: Vertical, Horizontal
2. Bar Plots and Stacked Bar Plots
3. Pie Charts and Donut Charts

Module 4: Exploring Seaborn

1. Introduction to Seaborn Library
2. Creating Distribution Plots: Histograms, Kernel Density Estimation (KDE)
3. Visualizing Relationships: Scatter Plots, Pair Plots

Module 5: Customizing Visualizations

1. Adding Annotations and Text
2. Customizing Colours, Styles, and Markers

3. Using Grids and Layouts

Module 6: Advanced Visualization Techniques

1. Box Plots and Violin Plots
2. Heatmaps for Correlation Analysis
3. Facet Grids for Multi-Plot Layouts

Machine Learning

Module 1: Introduction to Machine Learning

1. What is Machine Learning?
2. Types of Machine Learning: Supervised, Unsupervised, Semi-Supervised, Reinforcement Learning
3. Machine Learning Workflow: Data Collection, Preprocessing, Modelling, Evaluation, Deployment

Module 2: Supervised Learning: Linear Regression

1. Introduction to Linear Regression
2. Simple Linear Regression
3. Multiple Linear Regression
4. Evaluation Metrics: Mean Squared Error, R-squared

Module 3: Supervised Learning: Classification

1. Introduction to Classification
2. Logistic Regression
3. k-Nearest Neighbors (k-NN)
4. Model Evaluation: Confusion Matrix, Precision, Recall, F1-Score

Module 4: Model Evaluation and Cross-Validation

1. Train-Test Splitting
2. Cross-Validation: K-Fold Cross-Validation
3. Bias-Variance Trade-off

Module 5: Unsupervised Learning: Clustering

1. Introduction to Unsupervised Learning

2. k-Means Clustering
3. Hierarchical Clustering

Module 8: Dimensionality Reduction

1. Introduction to Dimensionality Reduction
2. Principal Component Analysis (PCA)

Module 9: Model Selection and Hyperparameter Tuning

1. Model Selection: Underfitting and Overfitting
2. Grid Search and Random Search for Hyperparameter Tuning
3. Hyperparameter Tuning with Scikit-Learn

Module 10: Ensemble Learning

1. Introduction to Ensemble Learning
2. Bagging and Random Forests
3. Boosting and Gradient Boosting

Module 11: Introduction to Neural Networks and Deep Learning

1. Basics of Artificial Neural Networks
2. Activation Functions, Loss Functions
3. Introduction to Deep Learning and Neural Network Architectures

Excel

Module 1: Introduction to Excel for Data Science

1. Overview of Microsoft Excel Interface
2. Basic Spreadsheet Concepts: Cells, Rows, Columns
3. Navigating Worksheets and Workbooks
4. Entering and Formatting Data

Module 2: Data Manipulation and Formulas

1. Basic Mathematical Operations
2. Cell References: Relative, Absolute, Mixed
3. Using Functions: SUM, AVERAGE, MIN, MAX

4. Text Functions: CONCATENATE, LEFT, RIGHT, MID

Module 3: Data Cleaning and Transformation

1. Removing Duplicates
2. Find and Replace
3. Text-to-Columns for Data Splitting
4. Handling Missing Values: Filtering, Filling

Module 4: Basic Data Visualization

1. Creating Simple Charts: Bar, Line, Pie
2. Formatting Charts: Titles, Axes, Legends
3. Adding Data Labels
4. Combining Multiple Charts on a Single Sheet

Module 5: Data Filtering and Sorting

1. Using AutoFilter
2. Custom Filters
3. Sorting Data: Ascending, Descending

Module 6: Conditional Formatting

1. Applying Basic Conditional Formatting Rules
2. Highlighting Cells: Text, Dates, Numbers
3. Using Icon Sets and Data Bars

Module 7: Data Analysis Tools

1. Using PivotTables: Creating, Modifying, Refreshing
2. Creating Pivot Charts
3. Data Validation: Input Messages, Error Alerts

Module 8: Advanced Formulas and Functions

1. Logical Functions: IF, AND, OR
2. Lookup Functions: VLOOKUP, HLOOKUP
3. Date and Time Functions: TODAY, NOW, DATE, TIME

Module 9: Data Visualization Techniques

1. Advanced Chart Types: Scatter, Bubble, Radar
2. Adding Trendlines and Error Bars
3. Creating Combination Charts

Course Pre-requisite

No prior knowledge of data science is required, but a basic understanding of statistics and the ability to wield a wand, or rather a programming language, will be advantageous.

Course Duration

30 Hours to Explore Data Science

CONTACT US

www.digisamurai.co.in

+91 8910632224

+91 7595887833