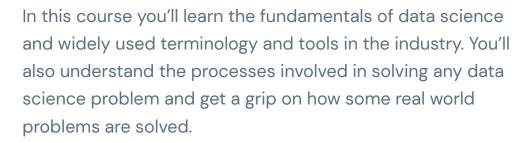# Data Science

Data Science is an interdisciplinary field that incorporates scientific knowledge from computer science, modelling, statistical analysis and mathematics. In this program, you will learn the art of uncovering insights and trends from data with several applications in the field of business, e-commerce, social networks and many more.

Foundations of Data Science

In this course you'll learn the fundamentals of data science and widely used terminology and tools in the industry. You'll also understand the processes involved in solving any data science problem and get a grip on how some real world problems are solved.

Introduction to Data Science

- What is Data Science?
- Applications of Data Science in various sectors
- Components of Data Science
- Python Libraries for Data Science

Mathematical Preliminaries

- Basic Probability theory
- Basic Statistics
- Bayesian statistics
- Probability distributions
- Normal distribution
- Poisson distribution
- Exponential distribution
- Mean, Median and Mode
- Central Limit theorem

Handling Data with Python

- Importing Data Sets into Python
- DataFrame Object in Pandas
- Dataframe Operations
- Grouping Data
- Aggregating Data

Data Cleaning

- Data types
- Handling Missing Values
- Handling Duplicates
- Handling Inconsistent Data
- Handling Outliers

Data Visualization fundamentals

- Principles of Information visualization
- Charting Fundamentals
- Applied Visualizations

Matplotlib, Seaborn

- 2D Plots and Subplots
- 3D plots
- Animations and Interactivity
- Styling
- Complex layouts with Pandas and Seaborn

Data Munging and Exploratory Data Analysis

- Data Formatting and Normalization
- Feature Scaling
- Dimensionality Reduction
- Summarizing Data Frames
- Descriptive Statistics
- Correlation
- Scores and Rankings

Course Project
Ecommerce Sale – An Exploratory Analysis

Develop a model and study the top products with best ratings and sales performance on the e-commerce platform.

Advanced Data Science

In this course, you'll learn advanced concepts in data science using statistical models. During this course, you'll get to apply the learnt processes on different data sets and also draw interesting insights from the data.

Statistical Analysis

- Inferential Statistics
- Statistical significance
- Z–test
- T–test
- ANOVA
- F Test
- Chi Square Test
- Multivariate Analysis
- Hypothesis Testing

Building and Validating Models

- Linear regression
- Logistic regression
- Nearest neighbor methods
- Clustering methods

Design of Experiments and Surveys - I

- Data science in the ideal versus real life
- Components of Experimental Design
- Causality
- Confounding
- Sampling bias and random sampling
- Blocking and adjustment
- Multiplicity
- Effect size, significance, & modelling

Design of Experiments and Surveys - II

- Comparison with benchmark effects

- Negative controls

- Non-significance

- Design Process and Guidelines

- One Factor Experiments

- Multi-factor Experiments

- Taguchi Methods

- Report writing

Course Project
Unemployment Trends due to Covid

Are young generations most affected by Covid 19? Develop hypotheses to statistically analyze the unemployment trends across various european countries.

Data Science at Scale

In this course, you will learn how to build predictive data science models for product teams using different cloud environment tools. Throughout the course, you'll have plenty of exercises to work with a diverse range of toolset. You'll also explore model workflows that move data between different cloud environments and also build real-time data pipelines.

Fundamentals of Data Engineering

- What is Data Engineering
- Data Engineering Problems
- Tools of a Data Engineer
- Cloud Providers and Cloud Computing

Data Engineering Tools

- Fundamentals of databases
- Parallel computing frameworks
- Spark, Hadoop and Hive
- Workflow scheduling frameworks
- Airflow DAGs

Building Data Engineering Pipelines in Python

- Scalable computing with Python
- Cloud Environments
- Coding Environments
- Data ingestion with Pandas
- Data pipelines
- ETL process

Data Modelling Tools

- Data Modelling with PostgreSQL
- Data Modelling with Apache Cassandra
- Storing Data in cloud warehouses
- Building Data Lakes

Prototyping Data Models

- Linear Regression
- Logistic Regression
- Keras Regression
- Automated Feature Engineering

Data Models as Web Endpoints

- Web service
- Echo service
- Model persistence
- Model Endpoints
- Deploying Endpoints with Gunicorn and Heroku

Pyspark for Batch pipelines

- Introduction to Pyspark
- Resilient distributed data sets
- Structured data processing with SparkSQL and Python
- Machine learning with Pyspark

Cloud Computing

- Fundamentals of Google Cloud Platform for Machine learning

- Cloud Data warehouses with Google cloud platform

- Batch Model pipelines in cloud

- Introduction to streaming model workflows

- Overview of Apache Kafka

- Sklearn Streaming

- Streaming analytics systems

Contact us

✉ support@nxtwave.tech

🌐 www.ccbp.in