



Python Programming Language and Statistical Modelling

Course Overview

Python is one of the most popular open source programming language used in statistical modelling and deep learning. In this course you will learn basics of Python, built in functions, how to write your own function. You will also learn how to visualize data. Finally, you will learn use of Python in descriptive statistics, anova and building models like linear and logistic regression, principal component analysis, cluster analysis and time series.

Course Expert

Satyaki Dasgupta has more than 30 years of experience in IT industries. He has executed more than 35 projects in India and abroad in different domains and technologies. He is a certified PMP. He handled many big clients like Income Tax of India, General Insurance Company, Peerless Hospital, CIGNA Insurance, Bankers Trust, St George Bank, Diary Farm, Universal Music, Southern California Edison, Wolters Kluwer, DHL, ITC, Centrica, Danske Bank, Outokumpu, TESCO, Channel4, Thermo Fisher. He worked for leading IT companies of India like Hinditron, Digital India, HCL, ITC, NIIT Technologies.

Satyaki worked extensively on CMM process and was instrumental in CMM level 5 certification for HCL Infosystems.

For last 5 years he has been working on data analytics using open source tools like R, Python, Azure ML Studio, Power BI, Tableau. He used various statistical models for stock forecast, viewership prediction of TV channels, telecom churning, sales forecast, product recommendation, face recognition.

He teaches above open source tools and conducts workshops at University of Petroleum and Energy Studies, Dehradun.



Python Programming Language and Statistical Modelling

Day One

Time 9:30 – 10:00
Module 1 Getting Started
Session Course Objective
Content History of Python

Time 10:00 – 12:00
Module 2 Python Basics
Session Python Environment Setup
Content Variables
Lists, tuples, sets
Strings
Dictionary
Dataframe

Time 12:00 – 13:00
Module 3 Functions
Session Common Python Utility Functions
Content Writing Functions
Exceptions

Lunch Break – 13:00 – 14:00

Time 15:00 – 17:00
Module 4 Data Manipulation and transformation
Session Loops
Content Sorting, merging, appending
Aggregation
Reading/writing files
Data file processing



Python Programming Language and Statistical Modelling

Day Two

Time 09:30 – 12:00

Module 5 Graphics

Session Basics

Content

Time 12:00 – 13:00

Module 6 Class

Session Object
Content Abstraction
Inheritance
Modules

Lunch Break – 13:00 – 14:00

Time 14:00 – 15:00

Module 6 Class

Session Object
Content Abstraction
Inheritance
Modules

Time 15:00 – 17:00

Module 7 Python Libraries

Session Numpy
Content SciPy
Mathplotlib



Python Programming Language and Statistical Modelling

Day Three

Time	09:30 – 13:00
Module 8	Statistics
Session	Missing Data
Content	Descriptive Statistics Frequency and Contingency Tables Tests of Independence Measure of Association Variance, Covariance, Correlation t-Tests Nonparametric Tests of Group Differences Outlier and influential observation Detection Cook's Distance and Leverage

Lunch Break – 13:00 – 14:00

Time	14:00 – 17:00
Module 9	Anova
Session	One way Anova
Content	Two way Anova Multivariate Anova



Python Programming Language and Statistical Modelling

Day Four

Time	09:30 – 13:00
Module 10	Linear Regression
Session Content	Supervised and Unsupervised Model Simple Linear Regression Polynomial Regression Multiple Linear Regression Regression Diagnostics Model Assumptions Stepwise Regression Relative Importance of Predictors Cross Validation Prediction

Lunch Break – 13:00 – 14:00

Time	14:00 – 17:00
Module 11	Logistic regression
Session Content	Logistic Regression Interpreting the Model Parameters Overdispersion Regression Diagnostics Prediction Survival Analysis



Python Programming Language and Statistical Modelling

Day Five

Time	09:30 – 11:00
Module 12	Principal Component Analysis
Session Content	Number of Components to Extract Extracting Principal Components Interpret Result Correlation - Variable and Component Correlation - Rows and Component Graphs

Time	11:00 – 13:00
Module 13	Cluster Analysis
Session Content	Steps in Cluster Analysis K-means Clustering PAM - Partitioning Around Medoids Hierarchical Cluster Analysis K-means Vs Hierarchical Clustering Visualization of Clusters

Lunch Break – 13:00 – 14:00

Time	14:00 – 16:00
Module 14	Time Series
Session Content	Smoothing and Seasonal Decomposition Exponential Forecasting Models ARIMA Forecasting Models Graphs Prediction

Q & A – 16:00 – 17:00



STRIATUM
TECHNOLOGIES

YOUR GATEWAY TO CUTTING EDGE

