*Suggested Teaching Guidelines for*

**Analytics & Statistics using Python and**

**Numerical Methods in Science & Engineering (100 Hrs)**

***PG-DHPCAP***

**Duration:** 50 classroom hours + 50 lab hours

**Objective:** To introduce the student to Analytics and Statistics using Python and Numerical methods concepts.

**Prerequisites:** Knowledge of programming in any language like C, C++ ,basic Mathematics, statistics and computer fundamentals.

**Evaluation method:** Theory exam– 40% weightage

Lab exam – 40 weightage

Internal exam – 20% weightage

**List of Books / Other training material**

## Reference Book:

1. Learn Python the Hard Way, [Zed A.Shaw,](https://www.flipkart.com/books/zed-ashaw~contributor/pr?sid=bks) Pearson
2. Introduction to Computer Science using Python, Charles/ Wiley
3. Python Power!: The Comprehensive Guide
4. Python Crash Course: A Hands-on, Project-Based Introduction to Programming
5. Beginning Programming with Python For DummiesLearning Python by: Fabrizio Romano
6. Python Cookbook by David B. Brain K. Jones / Shroff / O'reilly Publisher
7. Python Data Analytics by Fabio Nelli
8. Python for Data Analysis Data Wrangling with Pandas, NumPy, and IPython by Wes McKinney
9. Numerical Methods for Engineers, 5th Ed., McGraw Hill, Steven.S.Chapra and Raymond.P.Canale
10. The Monte Carlo Method Yu.A.Shreider
11. Input/Output In Parallel and Distributed Computer Systems by Ravi Jain , John Werth , James C. Browne
12. Numerical Methods for Engineers, New Age International , Santhosh.K.Gupta
13. Exploring Monte Carlo Methods , William Dunn,J.Kenneth shultis

## *Note: Each session mentioned is for theory and of 2 hours duration.*

## *Lab assignments are indicatives, faculty need to assign more assignments for better practice.*

## *Analytics & Statistics using Python*

**Session 1 :**

**Lecture**

* Working with Pandas
* Data wrangling with Pandas
* Working with Numpy, Scipy
* Data cleaning with Python

**Session 2:**

* Working with beautiful soup
* Working with scrappy

**Session 3 :**

* Data analytics Life Cycle:
  + Discovery,
  + Data Preparation
  + Model planning
  + Model building implementation
  + Quality assurance
  + Documentation
  + Management approval
  + Installation
  + Acceptance and operation

**Session 4:**

* Intelligent data analysis,
* Nature of Data,
* Analytic Processes and Tools,
* Analysis vs. Reporting
* Modern Data Analytic Tools

**Session 5:**

* Visualization and Exploring Data

## Session 6 :

* Basics of Statistics
* Statistical Analytics
* Descriptive Statistical Measures
* Statistics - Central Tendency & Dispersion (Mean, Median, Mode, Quartiles, Percentiles, Range, Interquartile Range, Standard Deviation, Variance, and Coefficient of Variation)

**Session 7:**

* Random Variable
* Concepts of Correlation
* Covariance
* Outliers
  + Producing graphs like bar chart, pie chart, histogram, boxplot, density plot, scatter plot with different options in pandas, matplotlib and seaborn libraries
  + Detecting Outliers using Boxplot

**Session 8:**

* + Sample Spaces and Events
  + Concept of Probability: Addition, Multiplicative, Complement Rules
  + Joint, Conditional and Marginal Probability
  + Bayes’ Theorem
  + Usage of sklearn.BernoulliNB function to predict probabilities ( predict\_proba ( ) method )

**Session 9 :**

* Probability Distribution
* Discrete distribution – (Binomial, Poisson) Probability Mass Functions, Distribution Functions
* Continuous distribution – (Normal) Probability Density Function, Distribution Function, Inverse of Distribution Function

**Session 10:**

* + Descriptive Statistical Measures
  + Summary Statistics - Central Tendency & Dispersion (Mean, Median, Mode, Quartiles, Percentiles, Range, Interquartile Range, Standard Deviation, Variance, and Coefficient of Variation)

**Session 11:**

* + Sample & population, Uni-variate and bi-variate sampling, re-sampling
  + Sampling and Estimation: Sampling Distribution
  + Concept of Confidence Interval
  + Central Limit Theorem

**Session 12 :**

* + Statistical Inference Terminology (types of errors, tails of test, confidence intervals etc.)
  + Hypothesis Testing
  + Parametric Tests: One sample t-test, paired t-test, 2 independent samples t-test, 1-Way ANOVA
  + Non-parametric Tests- chi-Square, U-Test
  + Inferential Statistics

**Session 13:**

* + Predictive Modelling (From Correlation to Supervised Segmentation):
  + Identifying Informative Attributes,
  + Segmenting Data by Progressive Attributive
  + Induction and Prediction,
  + Supervised Segmentation,
  + Visualizing Segmentations,
  + Trees as Set of Rules,
  + Probability Estimation

**Session 14:**

* Decision Analytics:
* Evaluating Classifiers,
* Analytical Framework,
* Evaluation,
* Baseline,
* Performance and Implications for Investments in Data

**Session 15:**

* Evidence and Probabilities:
* Explicit Evidence Combination with Bayes Rule,
* Probabilistic Reasoning

**Session 16:**

* Factor Analysis,
* Directional Data Analytics

## Session 17 & 18:

* Regression- Linear & Logistics
* Random Processes

**Numerical Methods in Science & Engineering**

**NUMERICAL METHODS**

**Session 19 :**

**Lecture**

* Introduction and Approximations
* Accuracy and Precision
* Truncation Errors
* Round-Off Errors
* Error Propagation
* Roots of equations
* Bracketing Methods
* Open Methods
* Roots of Polynomials
* Conventional Methods
* Muller's Method
* Bairstow’s Method
* Linear Algebraic Equations
* Gauss Elimination
* Matrix Inversion
* LU Decomposition

**Session 20:**

**Lecture**

* Regression
* Linear Regression
* Linear Least Squares
* Nonlinear Regression
* Interpolation
* Newton’s Divided-Difference Interpolating Polynomials
* Inverse Interpolation
* Numerical Integration
* The Trapezoidal Rule
* Simpson’s Rules
* Gauss Quadrature
* Numerical Differentiation
* High-Accuracy Differentiation Formulas
* Richardson Extrapolation

**Session 21:**

**Lecture**

* Numerical solution of ordinary differential equations
* Euler’s Method
* Improvements of Euler’s Method
* Runge-Kutta Methods
* Adaptive Runge-Kutta Methods
* Numerical solution of partial differential equations
* Finite Difference: Elliptic Equations
* Finite Difference: Parabolic Equations
* Finite-Element Method
* Case studies include Sequential and Parallel Algorithms

**Monte Carlo Methods**

**Session 22:**

**Lecture**

* Introduction to Monte Carlo Methods
* Basis of Monte Carlo
* Generating Random Variables
* Pseudo-random number generators
* Parallel random number generation
* Monte Carlo methods for numerical integration
* Monte Carlo methods for simulation
* Applications (e.g. radiation transport and network simulation)

**Data Management**

**Session 23:**

**Lecture**

* Introduction to data management
* Parallel file systems
* Parallel I/O
* Scientific data formats and libraries (e.g. NetCDF, and HDF5)
* Input/output using MPI
* Overview of MPI
* Data partition in MPI I/O
* Data Access Function in MPI-I/O
* File layout in MPI-I/O

**Session 24:**

**Lecture**

* Data-intensive computing
* Distributed file systems
* HDFS
* Hadoop Cluster Architecture
* GFS
* Map Reduce Algorithm

**Session 25:**

**Lecture**

* Web interface
* Introduction to data exchange
* Protocols for data exchange
* OPeNDAP
* Workflows for data processing
* Benchmarking
* Data intensive applications
* Types of Data intensive applications
* Profiling
* Modeling
* Management