

Data Science with Python Course Equipment and Material

Course Equipment: Required computer hardware and software.

COURSE SYLLABUS

COURSE NAME: Data Science with Python

OVERVIEW This course will teach you Data Science through Python, the simplest programming language. You will learn and practice problem solving techniques like pandas, scikit, and others. This will also give you an insight into how to manipulate, analyze, and perform various actions on complex datasets.

ELIGIBILITY Candidates from various quantitative backgrounds, like Engineering, Finance, Mathematics, Statistics, Business Management who want to start their career in analytics.

DURATION 80 hours

LESSON 1:	Introduction to Data Science. What is Data Science? Why Python for data science? Relevance in industry and need of the hour.
LESSON 2:	How leading companies are harnessing the power of Data Science with Python? Different phases of a typical Analytics/Data Science projects and role of python. Anaconda vs. Python
LESSON 3:	Overview of Python- Starting with Python. Introduction to installation of Python. Introduction to Python Editors & IDE's(Canopy, pycharm, Jupyter, Rodeo, lpython etc...). Understand Jupyter notebook & Customize Settings.
LESSON 4:	Concept of Packages/Libraries - Important packages(NumPy, SciPy, scikit-learn, Pandas, Matplotlib, etc) Installing & loading Packages & Name Spaces. Data Types & Data objects/structures (strings, Tuples, Lists, Dictionaries). List and Dictionary Comprehensions
LESSON 5:	Variable & Value Labels – Date & Time Values. Basic Operations - Mathematical - string - date Reading and writing data. Simple plotting
LESSON 6:	Control flow & conditional statements. Debugging & Code profiling. How to create class and modules and how to call them?
LESSON 7:	SCIENTIFIC DISTRIBUTIONS USED IN PYTHON FOR DATA SCIENCE - Numpy, scify, pandas, scikitlearn, statmodels, nltk etc.
LESSON 8:	ACCESSING/IMPORTING AND EXPORTING DATA USING PYTHON MODULES - Importing Data from various sources (Csv, txt, excel, access etc). Database Input (Connecting to database)
LESSON 9:	Viewing Data objects - subsetting, methods. Exporting Data to various formats. Important python modules: Pandas, beautifulsoup
LESSON 10:	DATA MANIPULATION – CLEANSING – MUNGING USING PYTHON MODULE - Cleansing Data with Python. Data Manipulation steps (Sorting, filtering, duplicates, merging, appending, subsetting, derived variables, sampling, Data type conversions, renaming, formatting etc).
LESSON 11:	Data manipulation tools(Operators, Functions, Packages, control structures, Loops, arrays etc). Python Built-in Functions (Text, numeric, date, utility functions).
LESSON 12:	Python User Defined Functions. Stripping out extraneous information. Normalizing data. Formatting data. Important Python modules for data manipulation (Pandas, Numpy, re, math, string, datetime etc)
LESSON 13:	DATA ANALYSIS – VISUALIZATION USING PYTHON- Introduction exploratory data analysis. Descriptive statistics, Frequency Tables and summarization.
LESSON 14:	Univariate Analysis (Distribution of data & Graphical Analysis). Bivariate Analysis (Cross Tabs,

	Distributions & Relationships, Graphical Analysis).
LESSON 15:	Creating Graphs- Bar/pie/line chart/histogram/ boxplot/ scatter/ density etc). Important Packages for Exploratory Analysis(NumPy Arrays, Matplotlib, seaborn, Pandas and scipy. stats etc)
LESSON 16:	BASIC STATISTICS & IMPLEMENTATION OF STATS METHODS IN PYTHON - Basic Statistics - Measures of Central Tendencies and Variance.
LESSON 17:	Building blocks - Probability Distributions - Normal distribution - Central Limit Theorem. Inferential Statistics -Sampling - Concept of Hypothesis Testing.
LESSON 18:	Statistical Methods - Z/t-tests (One sample, independent, paired), Anova, Correlation and Chi-square. Important modules for statistical methods: Numpy, Scipy, Pandas
LESSON 19:	PYTHON: MACHINE LEARNING -PREDICTIVE MODELING – BASICS. Introduction to Machine Learning & Predictive Modeling. Types of Business problems - Mapping of Techniques - Regression vs. classification vs. segmentation vs. Forecasting
LESSON 20:	Major Classes of Learning Algorithms -Supervised vs Unsupervised Learning. Different Phases of Predictive Modeling (Data Pre-processing, Sampling, Model Building, Validation).
LESSON 21:	Overfitting (Bias-Variance Trade off) & Performance Metrics. Feature engineering & dimension reduction.
LESSON 22:	Concept of optimization & cost function. Concept of gradient descent algorithm. Concept of Cross validation (Bootstrapping, K-Fold validation etc).
LESSON 23:	Model performance metrics (R-square, RMSE, MAPE, AUC, ROC curve, recall, precision, sensitivity, specificity, confusion metrics)
LESSON 24:	MACHINE LEARNING ALGORITHMS & APPLICATIONS – IMPLEMENTATION IN PYTHON. Linear & Logistic Regression.
LESSON 25:	Segmentation - Cluster Analysis (K-Means). Decision Trees (CART/CD 5.0). Ensemble Learning (Random Forest, Bagging & boosting).
LESSON 26:	Artificial Neural Networks(ANN). Support Vector Machines(SVM). Other Techniques (KNN, Naïve Bayes, PCA). Introduction to Text Mining using NLTK.
LESSON 27:	Introduction to Time Series Forecasting (Decomposition & ARIMA). Important python modules for Machine Learning (SciKit Learn, stats models, scipy, nltk etc). Fine tuning the models using Hyper parameters, grid search, piping etc.

TAKE AWAYS After completing this course, the student will understand and learn to apply Data science concepts using Python. This course is an HOTTEST Trending Technology tool that is highly in-demand in IT industry right now. Learning these concepts will increase your options of getting better job opportunities in market.