

# Data Science with Python Syllabus

## SECTION I --- Python Basics

### Lesson 1: Overview

- Why do we need Python?
- Program structure

### Environment Setup

- Python Installation
- Execution Types
  - What is an interpreter?
  - Interpreters vs Compilers
  - Using the Python Interpreter
  - Interactive Mode
  - Running python files
  - Working with Python shell
  - Integrated Development Environments (IDES)
  - Interactive Mode Programming
  - Script Mode Programming

### Lesson 2 : Basic Concepts

- **Basic Operators**
  - Types of Operator
  - Python Arithmetic Operators
  - Python Comparison Operators
  - Python Assignment Operators
  - Python Bitwise Operators
  - Python Logical Operators
  - Python Membership Operators (in, not in)
  - Python Identity Operators (is, is not)
  - Python Operators Precedence
- **Data Types**
  - Variables
  - Assigning Values to Variables
  - Multiple Assignment
  - Python Numbers
  - Python Strings
    - Accessing Values in Strings
    - String Special Operators
    - String Formatting Operator
    - Triple Quotes
    - Built-in String Operations
  - Python Lists

- Accessing Values in Lists
  - Updating Lists
  - Delete List Elements
  - Basic List Operations
  - Indexing, Slicing, and Matrixes
  - Built-in List Functions & Methods
- Python Tuples
  - Accessing Values in Tuples
  - Updating Tuples
  - Delete Tuple Elements
  - Basic Tuples Operations
  - Indexing, Slicing, and Matrixes
  - No Enclosing Delimiters
  - Built-in Tuple Functions
- Python Dictionary
  - Accessing Values in Dictionary
  - Updating Dictionary
  - Delete Dictionary Elements
  - Properties of Dictionary Keys
  - Built-in Dictionary Functions & Methods

### **Lesson 3:Loops and Decision Making**

- if statements
- if...else statements
- nested if statements
- while loop
- for loop
- nested loops
- Loop Control Statements
- 1) break statement
- 2) continue statement
- 3) pass statement

### **Lesson 4 :Functions**

- Defining a Function
- Syntax
- Calling a Function
- Pass by reference vs value
- Function Arguments
- Required arguments
- Keyword arguments
- Default arguments
- Variable-length arguments
- The return Statement
- Scope of Variables
- Global vs. Local variables

### **Lesson 5: Basic OOPs Concept**

- Creating class in Python

- Documented String
- Private Identifier
- Constructor
- Inheritance
- Polymorphism

### **Lesson 6 : Python Modules and Packages**

- Framework vs Packages
- Folium Introduction
- Why are modules used?
- Creating modules
- The import Statement
- The from...import Statement
- The from...import \* Statement
- Locating Modules
- The PYTHONPATH Variable
- Namespaces and Scoping
- The dir( ) Function
- The globals() and locals() Functions
- The reload() Function
- Packages in Python

### **Lesson 7: Advance Python**

- **Decorator, Iterator and Generator**
- **Anonymous Function**
  - Lambda
  - Map
  - Filter
  - Reduce

### **SQL and Python**

- Overview of SQLite
- Integrating Python with SQLite
- **Errors and Exception Handling**
  - Standard exceptions
  - Assertions in Python
  - The assert Statement
  - What is Exception?
  - Handling an exception
  - Syntax
  - The except Clause with No Exceptions
  - The except Clause with Multiple Exceptions
  - The try-finally Clause
  - Argument of an Exception
  - Example with Tkinter Application

# Section II -- Statistics and Data Science Overview

## Lesson 8 : Data Science Overview

### ● Data Science Disciplines

- Data Science and Business Buzzwords Why are there so many
- What is the difference between Analysis and Analytics
- An Introduction--Business Analytics, Data Analytics, and Data Science
- Data Science Diagram
- Introduction -- BI, ML and AI
- Careers in Data Science Fields

### ● Data Overview

- What is Data
- Measuring Data
- Measurement of Central Tendency
- Measurements Dispersion
- Measurement Quartile
- Bi-variate Data and Co-variance
- Pearson Correlation Coefficient

### ● Lesson 9 : Probability

- What is Probability
- Permutations
- Combinations
- Intersections Unions and Complements
- Independent and Dependent Events
- Conditional Probability
- Addition and Multiplication Rules

- Bayes Theorem

- **Lesson 10: Distributions**

- Introduction to Distributions
- Uniform Distribution
- Binomial Distribution
- Poisson Distribution
- Normal Distribution

- **Lesson 11: Statistics**

- What is Statistics
- Sampling
- Central Limit Theorem
- Standard Error
- Hypothesis Testing
- Hypothesis Testing Example Exercise
- Type 1 and Type 2 Errors
- Students T Distribution
- Practical Example Descriptive Statistics Exercise
- What are Confidence Intervals
- Correlation Matrix

- **Lesson 12: Anova**

- Introduction to ANOVA
- Two Way ANOVA Overview
- F- Distribution

- **Lesson 13: Chi Square Analysis**

- Chi-Square Analysis
- Chi Squared Analysis - Exercise Example

## **Section III -- Python for Data Analysis**

- **Lesson 14 : Python: Environment Setup and Essentials**
  - Introduction to Anaconda
  - Installation of Anaconda Python Distribution – For Windows, Mac OS, and Linux
  - Jupyter Notebook Installation
  - Jupyter Notebook Introduction
  
- **Lesson 15:Data Analysis- Numpy**
  - Introduction to Numpy
  - Numpy Array
  - Numpy Indexing
  - Numpy Operations
  - Broadcasting Numpy Array
  
- **Lesson 16:Data Analysis -- Pandas**
  - Introduction to Pandas
  - Series
  - Data Frames
  - Missing Data
  - Groupby
  - Operations
  - Merging, Joining and concatenating
  - Missing Data
  - Data Input and Output
  
- **Lesson 17:Pandas Exercise**
  - Salaries Exercise
  - Ecommerce Purchases Exercise
  
- **Lesson 18:Numpy Exercise**

- Solving Linear System
- Problem Set

## **Section IV -- Python for Data Visualization**

### **● Lesson 19:Matplotlib**

- Introduction
- Matplotlib Drawing Graph -- Histogram, Plotting, Box Plot etc
- Exercise

### **● Lesson 20:Seaborn**

- Introduction
- Distribution
- Categorical Plots
- Matrix Plots
- Regression Plots
- Grids
- Style and Colors
- Exercise

### **● Lesson 21:Data Visualization with Pandas**

- Pandas Built-in Data Visualization
- Pandas Data Visualization Exercise

### **● Lesson 22:Data Visualization - Geographical Plotting**

- Introduction to Geographical Plotting
- Choropleth Maps - Part 1 - USA
- Choropleth Maps - Part 2 - World
- Choropleth Exercises

- **Capstone Project I**
  - Calls Data Capstone Project
  - Finance Project
  
- **Lesson 23: Time Series Analysis**
  - Pandas for Time Series
  - Introduction to Time Series with Pandas
  - Date time Index
  - Time Re-sampling
  - Time Shifts
  - Pandas Rolling and Expanding
  - Time Series Analysis
  - Introduction to Time Series
  - Time Series Basics
  - Introduction to Statsmodel
  - ETS Theory
  - EWMA Theory
  - ARIMA Theory
  - ACF and PACF
  - ARIMA with Statsmodel
  
- **Capstone Project II**
  - Stock Market Analysis Project
  
- **Lesson 24: Scientific computing with Python (Scipy)**
  - SciPy and its Characteristics
  - SciPy sub-packages
  - SciPy sub-packages – Integration
  - SciPy sub-packages – Optimize
  - Linear Algebra



- SciPy sub-packages – Statistics
- SciPy sub-packages – Weave
- SciPy sub-packages – I O

- **Lesson 25: Data Science with Python Web Scraping**

- Web Scraping
- Common Data/Page Formats on The Web
- The Parser
- Importance of Objects
- Understanding the Tree
- Searching the Tree
- Navigating options
- Modifying the Tree
- Parsing Only Part of the Document
- Printing and Formatting
- Encoding

## **Section V -- Machine Learning**

- **Lesson 26: Machine Learning with Python (Scikit-Learn)**

- Introduction to Machine Learning
- Machine Learning Approach
- How Supervised and Unsupervised Learning Models Work
- Scikit-Learn
- Supervised Learning Models – Linear Regression
- Supervised Learning Models: Logistic Regression
- K Nearest Neighbors (K-NN) Model
- K Means Algorithm
- SVMs
- Unsupervised Learning Models: Clustering
- Unsupervised Learning Models: Dimensionality Reduction
- Pipeline
- Model Persistence
- Model Evaluation – Metric Functions

- **Lesson 27: Natural Language Processing with Scikit-Learn**

- NLP Overview
- NLP Approach for Text Data
- NLP Environment Setup
- NLP Sentence analysis
- NLP Applications
- Major NLP Libraries
- Scikit-Learn Approach
- Scikit – Learn Approach Built – in Modules
- Scikit – Learn Approach Feature Extraction

- Bag of Words
  - Extraction Considerations
  - Scikit – Learn Approach Model Training
  - Scikit – Learn Grid Search and Multiple Parameters
  - Pipeline
- **Lesson 28: Python integration with Hadoop, MapReduce and Spark**
    - Need for Integrating Python with Hadoop
    - Big Data Hadoop Architecture
    - MapReduce
    - Cloudera QuickStart VM Set Up
    - Apache Spark
    - Resilient Distributed Systems (RDD)
    - PySpark
    - Spark Tools
    - PySpark Integration with Jupyter Notebook

## **Section VI: Project Works**

**Project 1-- Board Game Review Prediction** -- To perform a Linear Regression Analysis by predicting the average reviews in a board game

**Project 2 -- Credit Card Fraud Detection** -- To focus on Anomaly Detection by using probability densities to detect credit card fraud

**Project 3 – Stock Market Clustering** – Learn how to use the K-means clustering algorithm to find related companies by finding correlations among stock market movements over a given time span

**Project 4 – Getting Started with Natural Language Processing in Python** – This project will focus on Natural Language Processing (NLP) methodology, such as tokenizing words and sentences, part of speech identification and tagging, and phrase chunking.