Data Science using Python (Beginners Course)

‘’’We live in a world of connected things where tons of data are generated and it is humanly impossible to analyze all the incoming data and make decisions. Human decisions are increasingly replaced by decisions made by computers, thanks to the field of Data Science’’’ (*Python Data Science Cookbook, Packt Publishing 2015*).

Data Science is an interdisciplinary field about Scientific methods, processes and systems to extract knowledge or insights from data. Data science comprises knowledge about statistics/mathematics, analysis, information/computer science processing techniques and theories from Artificial Intelligence, Machine Learning, Data Mining along Visualization.

Python as a programming language has evolved over the years and today, it is the number one choice for a data scientist. It’s ability to act as a scripting language for quick prototype building and its sophisticated language constructs for full-fledged software development combined with its fantastic library support for numeric computations has led to its current popularity among data scientists and the general scientific programming community.

Course contents and resources comprises concepts of Data collection, Data processing and Data analysis using File management, Web Scraping and analytical processing techniques with Python Programming using ‘pandas’ and related modules.

Conceptual Steps for Data Science:

1. Data Collection (Raw Data), Data Cleaning, Formatting, Ordering
2. Exploring Data using Python Programming Language
3. Achieving sense/insights from Data thru Visualization/Analysis.

Course Contents:

1. Introduction
    a. Data Science, Data Analysis, Web Scraping, Machine Learning
    b. Python Programming
2. Python Programming
    a. Overview, Requirement Installation
    b. Operators, Decision Making, Loops, Functions
    c. Strings, Lists, Tuples, Dictionary (Comprehensions)
    d. Lambda, zip, map, filter, reduce
    e. Modules (re, os, datetime, numpy..)
    f. Jupyter Notebook
3. Data Collection
    a. Python File Handling (CSV, JSON, XML, XSL..)
    b. Web Scraping (lxml, requests, urllib)
4. Data Processing & Visualization
    a. Pandas (Data loading, Data Wrangling, Data Aggregation & Operations)
    b. Visualization using matplotlib, seaborn
5. Reporting
    a. Jupyter Notebook , pdfPages

Course Duration: 5 Weeks

TimeLine:

1. Week 1: Introduction & Python Programming
2. Week 2: Python Programming & Data Collection
3. Week 3: Data Collection and Data Processing
4. Week 4: Data Processing and Reporting
5. Week 5: Conducting Real-Time-Data-Driven Analysis, Project perspective