5 Days (Instructor-Led)

Introduction

Our lives are flooded by large amount of information, but not all of them is useful data. Therefore, it is essential for us to learn how to applying data science to every aspect of our daily life from personal finances, reading, lifestyle to making business decisions. Leveraging on this data to make our life easier, or unlock new economic value for a business, is what you are going to learn in this course.

This course is a hands-on guided course for you to learn the concepts, tools, and techniques that you need to begin learning data science. We will cover the key topics from data science to big data, and the processes of gathering, cleaning and handling data. This course is well balanced between theory and practical, and key concepts are taught using case studies references. Upon completion, participants will be able to perform the basic data handling tasks, collect and analyze data, and present them using industry standard tools.

Target audience

This workshop is intended for individuals who are interested in learning data science, or who want to begin their career as a data scientist.

Prerequisites

All participants should have basic understanding of data, relations, and basic knowledge of mathematics.

Objectives

Upon completion of this course, you will be able to:

- Identify appropriate model for different data types.
- Create your own data process and analysis workflow.
- Define and explain the key concepts and models relevant to data science.
- Differentiate key data ETL process, from cleaning, processing to visualization.
- Implement algorithms to extract information from dataset.
- Apply best practices in data science, and familiar with standard tools.

Course Outline

Day 1

Introduction to Data Science

- What is Data?
- Types of Data
- What is Data Science?
- Statistical thinking
- Knowledge Check
- Lab Activity

Data Processes

- Extract, Transform and Load (ETL)
- Data Cleansing
- Aggregation, Filtering, Sorting, Joining
- Data Workflow
- Knowledge Check
- Lab Activity

Data Quality

- Raw vs Tidy Data
- Key features of data quality
- Maintenance of data quality
- Data profiling
- Data completeness and consistency

Life of a data scientist

- Identify problem
- Define question
- Define ideal dataset
- Obtain data
- Analyze data
- Interpret results
- Distribute results
- Knowledge Check

Day 2

Beginning Databases

- Types of Databases
- Relational Databases

- NoSQL
- Hybrid database
- Knowledge check
- Lab activity

Structured Query Language (SQL)

- Performing CRUD (Create, Retrieve, Update, Delete)
- Designing a Real-world database
- Normalizing a table
- Knowledge check
- Lab Activity

Introduction to Python

- Basics of Python language
- Functions and packages
- Python lists
- Functional programming in Python
- Numpy and Scipy
- iPython
- Knowledge check
- Lab Activity

Lab: Exploring data using Python

Day 3

Data Gathering

- Obtain data from online repositories
- Import data from local file formats (json, xml)
- Import data using Web API
- Scrape website for data
- Knowledge check
- Lab Activity

Instructor-led case study

Exploratory Data Analysis

- What is EDA?
- Goals of EDA
- The role of graphics
- Handling outliers
- Dimension reduction

Introduction to R

- Features of R
- Vectors
- Matrices and Arrays
- Data Frame
- Input / Output

Lab: Exploring data using R

Day 4

Introduction Text Mining

- What is Text Mining?
- Natural Language Processing
- Pre-processing text data
- Extracting features from documents
- Using BeautifulSoup
- Measuring document similarity
- Knowledge check
- Lab activity

Supervised Learning

- What is prediction?
- Sampling, training set, testing set.
- Constructing a decision tree.
- Knowledge check
- Lab Activity

Day 5

Presenting Data

- Choosing the right visualization
- Plotting data using Python libraries
- Plotting data using R
- Using Jupyter Notebook to validate scripts
- Knowledge check
- Lab activity

Data Analysis Presentation

- Using Markdown language
- Convert your data into slides
- Data presentation techniques
- The pitfall of data analysis

- Knowledge check
- Lab activity
- Group presentation

Lab: Mini Project

Big Data Landscape

- What is small data?
- What is big data?
- Big data analytics vs Data Science
- Key elements in Big Data (3Vs)
- Extracting values from big data
- Challenges in Big data

Big data Tools and Applications

- Introducing Hadoop Ecosystem
- Cloudera vs Hortonworks
- Real world big data applications
- Knowledge check
- Group discussion

What's next?

- Preview of Data Science Specialist
- Showing advanced data analysis techniques
- Demo: Interactive visualizations