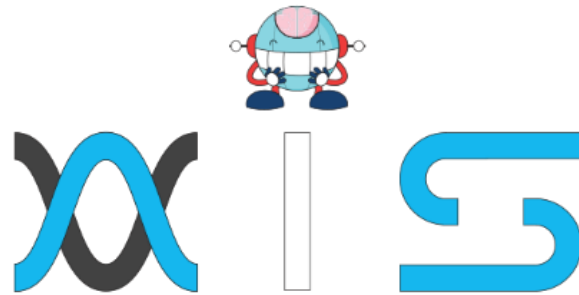




**SPARTAN**  
Allied Services  
*Deliver Results - Create Value*

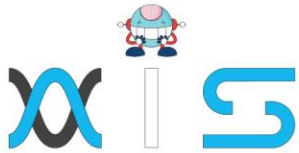
simplilearn



# DATA SCIENCE WITH PYTHON

## Course Outline





### Program Overview:

Establish your mastery of data science and analytics techniques using Python by enrolling in this Data Science with Python course. You'll learn the essential concepts of Python programming and gain in-depth knowledge of data analytics, machine learning, data visualization, web scraping, and natural language processing. Python is a required skill for many data science positions, so jumpstart your career with this interactive, hands-on, Data Science with Python course.

### Program Features:

- 24 hours of Online self-paced learning
- 44 hours of instructor-led training
- 4 industry-based course-end projects
- Interactive learning with Jupyter notebooks integrated labs
- Dedicated mentoring session from faculty of industry experts

### Delivery Mode:

Online Bootcamp - Online self-paced learning and live virtual classroom

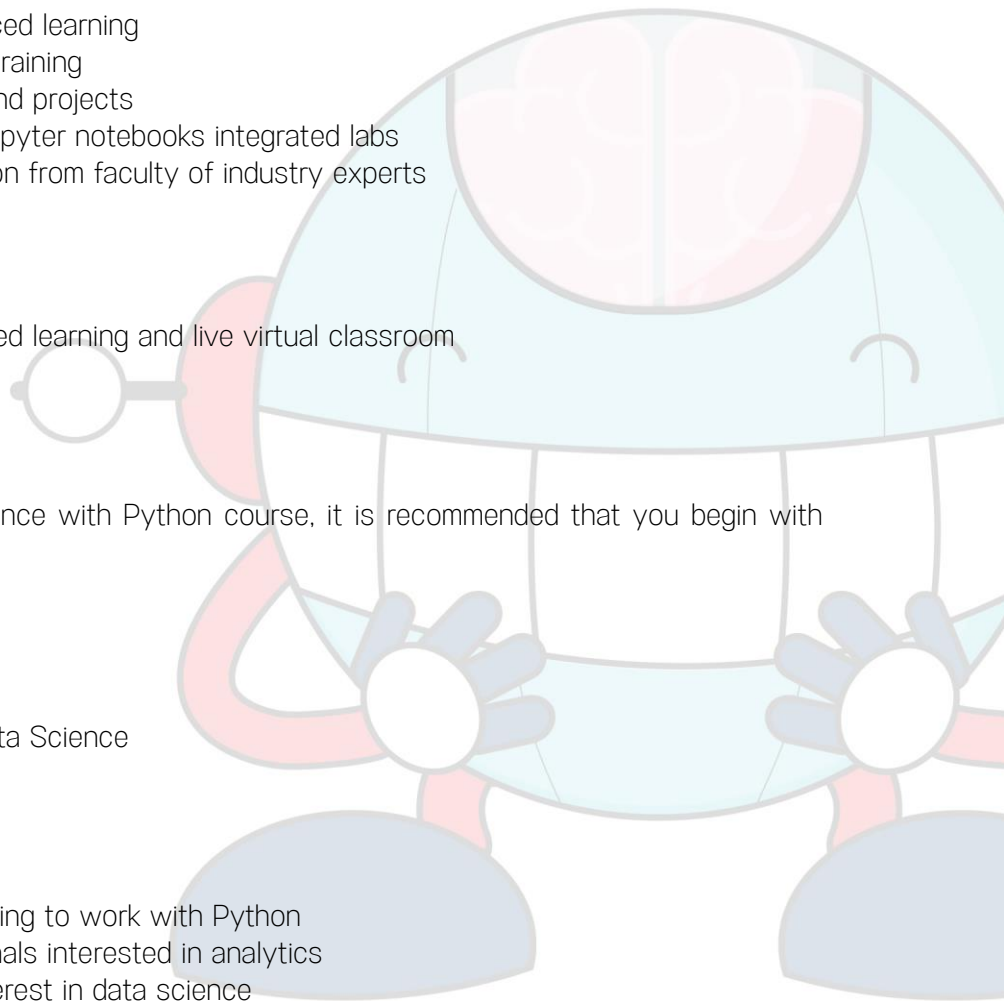
### Prerequisites:

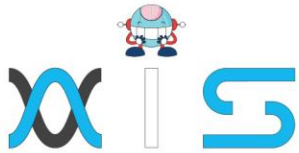
To best understand the Data Science with Python course, it is recommended that you begin with these courses:

- Python Basics
- Math Refresher
- Data Science in Real Life
- Statistics Essentials for Data Science

### Target Audience:

- Analytics professionals willing to work with Python
- Software and IT professionals interested in analytics
- Anyone with a genuine interest in data science





### Key Learning Outcomes:

This Python for Data Science training course will enable you to:

- Gain an in-depth understanding of data science processes, data wrangling, data exploration, data visualization, hypothesis building, and testing; and the basics of statistics
- Understand the essential concepts of Python programming such as datatypes, tuples, lists, dicts, basic operators, and functions
- Perform high-level mathematical computations using the NumPy and SciPy packages and their large library of mathematical functions
- Perform data analysis and manipulation using data structures and tools provided in the Pandas package
- Gain an in-depth understanding of supervised learning and unsupervised learning models such as linear regression, logistic regression, clustering, dimensionality reduction, K-NN, and pipeline
- Use the Scikit-Learn package for natural language processing and matplotlib library of Python for data visualization

### Key Learning Outcomes:

- 85 percent of online self-paced completion or attendance of one live virtual classroom
- A score of at least 75 percent in course-end assessment
- Successful evaluation in at least one project

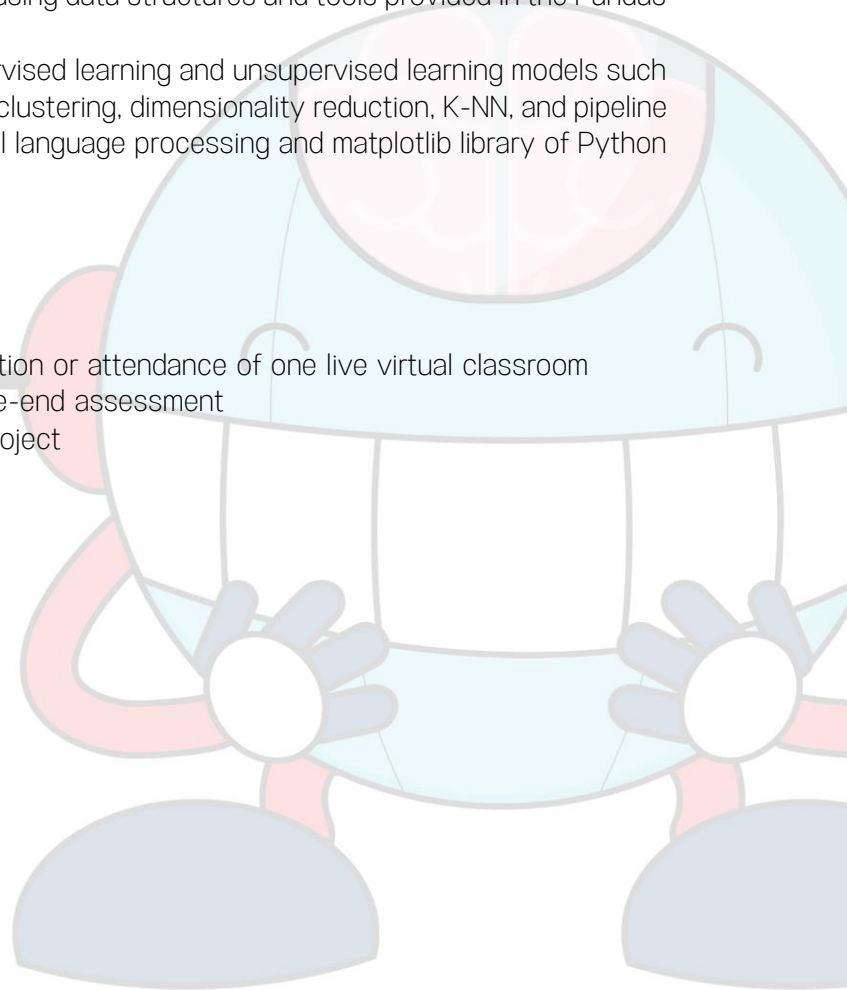
### Table of Contents

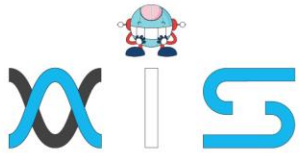
#### Lesson 00: Course Overview

- Course Overview

#### Lesson 1: Data Science Overview

- Introduction to Data Science
- Different Sectors Using Data Science
- Purpose and Components of Python
- Quiz





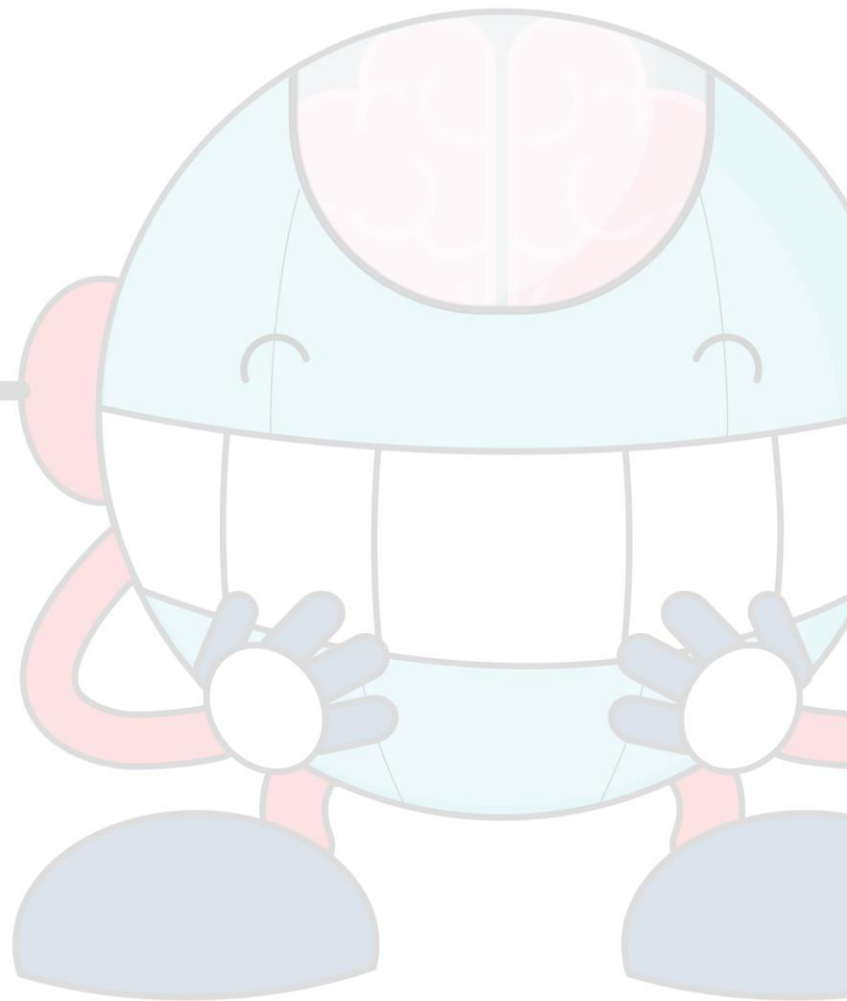
- Key Takeaways

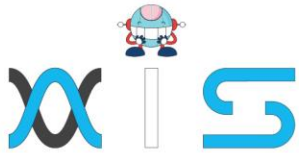
## Lesson 2: Data Analytics Overview

- Data Analytics Process
- Knowledge Check
- Exploratory Data Analysis (EDA)
- Quiz
- EDA-Quantitative Technique
- EDA - Graphical Technique
- Data Analytics Conclusion or Predictions
- Data Analytics Communication
- Data Types for Plotting
- Data Types and Plotting
- Quiz
- Key Takeaways
- Knowledge Check

## Lesson 03: Data Analytics Overview

- Introduction to Statistics
- Statistical and Non-statistical Analysis
- Major Categories of Statistics
- Statistical Analysis Considerations
- Population and Sample
- Statistical Analysis Process
- Data Distribution
- Dispersion
- Knowledge Check
- Histogram
- Knowledge Check
- Testing
- Knowledge Check
- Correlation and Inferential Statistics
- Quiz
- Key Takeaways





## Lesson 04: Python Environment Setup and Essentials

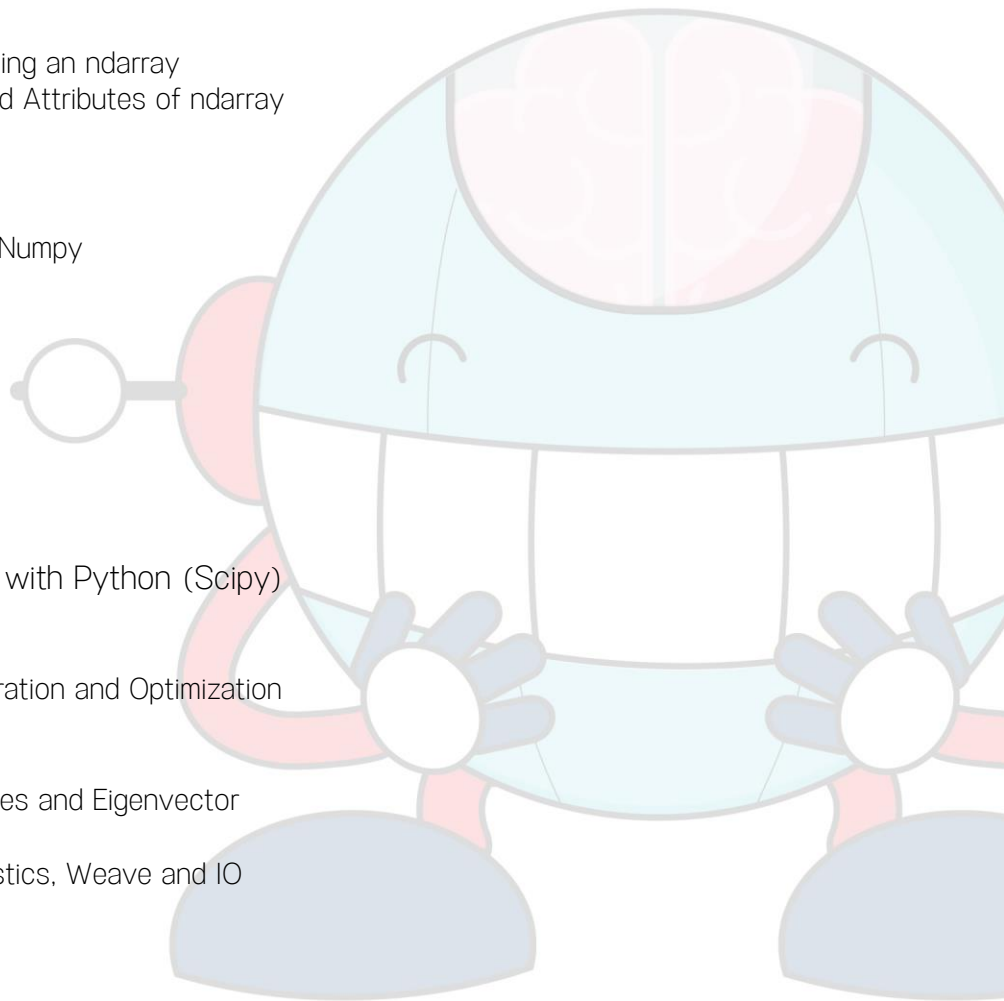
- Anaconda
- Installation of Anaconda Python Distribution (contd.)
- Data Types with Python
- Basic Operators and Functions
- Quiz
- Key Takeaways

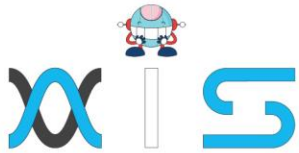
## Lesson 05: Mathematical Computing with Python (NumPy)

- Introduction to Numpy
- Activity-Sequence it Right
- Demo 01-Creating and Printing an ndarray
- Knowledge Check Class and Attributes of ndarray
- Basic Operations
- Activity-Slice It
- Copy and Views
- Mathematical Functions of Numpy
- Assignment 01
- Assignment 01 Demo
- Assignment 02
- Assignment 02 Demo
- Quiz
- Key Takeaways

## Lesson 06: Scientific computing with Python (SciPy)

- Introduction to SciPy
- SciPy Sub Package - Integration and Optimization
- Knowledge Check
- SciPy sub package
- Demo - Calculate Eigenvalues and Eigenvector
- Knowledge Check
- SciPy Sub Package - Statistics, Weave and IO
- Assignment 01
- Assignment 01 Demo
- Demo Assignment 02
- Assignment 02 Demo





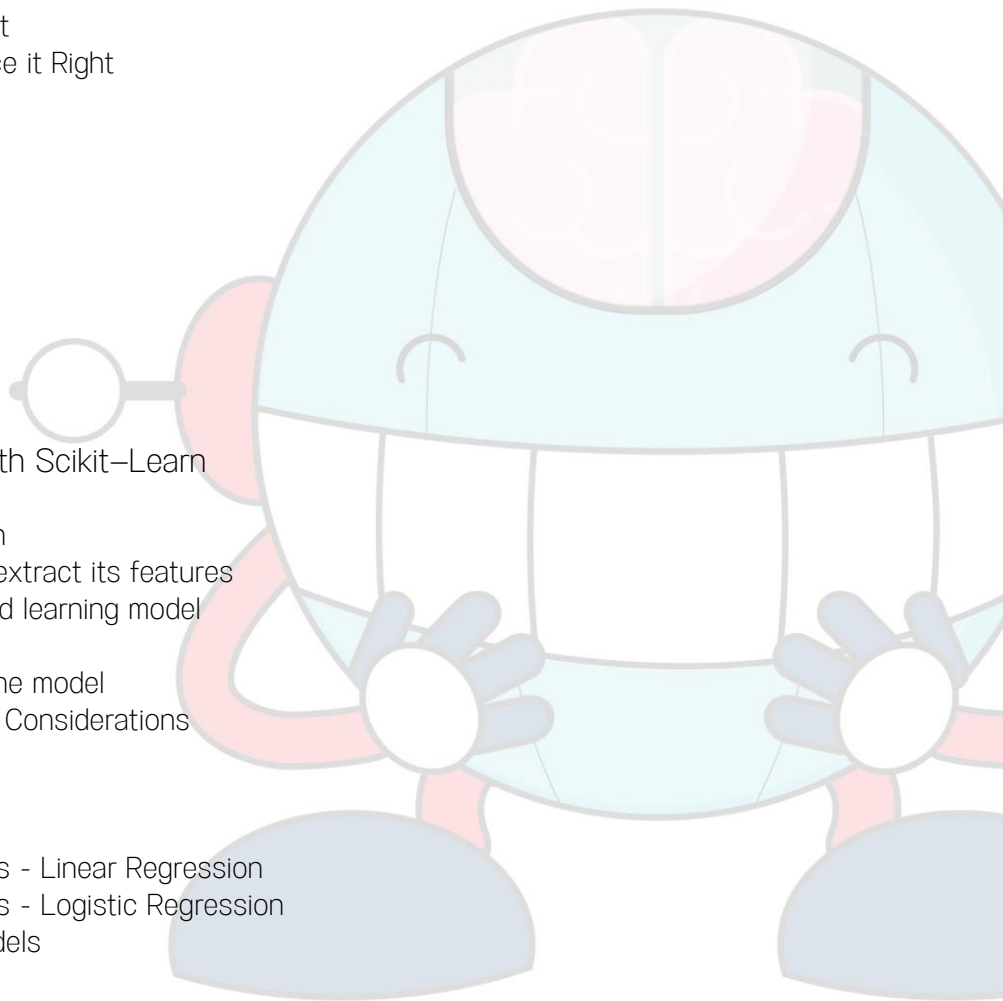
- Quiz
- Key Takeaways

## Lesson 07: Data Manipulation with Pandas

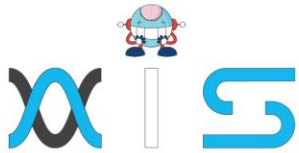
- Introduction to Pandas
- Knowledge Check
- Understanding DataFrame
- View and Select Data Demo
- Missing Values
- Data Operations
- Knowledge Check
- File Read and Write Support
- Knowledge Check-Sequence it Right
- Pandas Sql Operation
- Assignment 01
- Assignment 01 Demo
- Assignment 02
- Assignment 02 Demo
- Quiz
- Key Takeaways

## Lesson 08: Machine Learning with Scikit-Learn

- Machine Learning Approach
- Understand data sets and extract its features
- Identifying problem type and learning model
- How it Works
- Train, test and optimizing the model
- Supervised Learning Model Considerations
- Knowledge Check
- Scikit-Learn
- Knowledge Check
- Supervised Learning Models - Linear Regression
- Supervised Learning Models - Logistic Regression
- Unsupervised Learning Models
- Pipeline
- Model Persistence and Evaluation







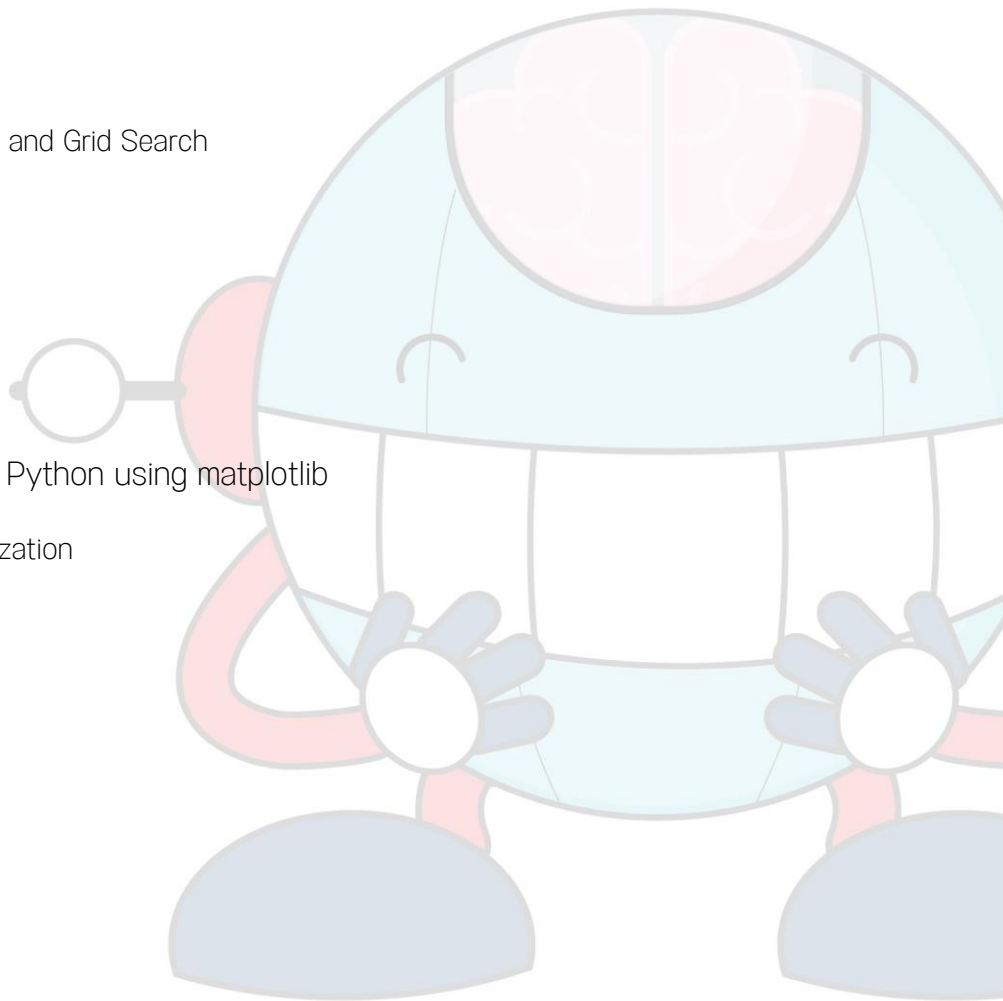
- Assignment 01
- Knowledge Check
- Assignment 01
- Assignment 02
- Assignment 02
- Quiz
- Key Takeaways

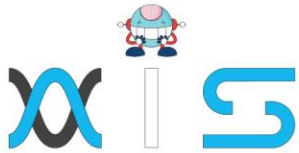
## Lesson 09: Natural Language Processing with Scikit Learn

- NLP Overview
- NLP Applications
- Knowledge Check
- NLP Libraries-Scikit
- Extraction Considerations
- Scikit Learn-Model Training and Grid Search
- Assignment 01
- Demo
- Assignment 01
- Assignment 02
- Demo Assignment 02
- Quiz
- Key Takeaway

## Lesson 10: Data Visualization in Python using matplotlib

- Introduction to Data Visualization
- Knowledge Check
- Line Properties
- (x,y) Plot and Subplots
- Knowledge Check
- Types of Plots
- Assignment 01
- Assignment 01 Demo
- Assignment 02
- Assignment 02 Demo
- Quiz
- Key Takeaways





## Lesson 11: Web Scraping with BeautifulSoup

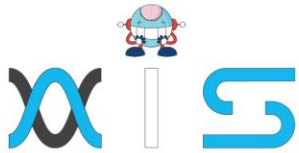
- Web Scraping and Parsing
- Knowledge Check
- Understanding and Searching the Tree
- Navigating options
- Demo3 Navigating a Tree
- Knowledge Check
- Modifying the Tree
- Parsing and Printing the Document
- Assignment 01
- Assignment 01 Demo
- Assignment 02
- Assignment 02 demo
- Quiz
- Key takeaways

## Lesson 12: Python integration with Hadoop MapReduce and Spark

- Why Big Data Solutions are Provided for PythonO
- Hadoop Core Components
- Python Integration with HDFS using Hadoop Streaming
- Demo 01 - Using Hadoop Streaming for Calculating Word Count
- Knowledge Check
- Python Integration with Spark using PySpark
- Demo 02 - Using PySpark to Determine Word Count
- Knowledge Check
- Assignment 01
- Assignment 01 Demo
- Assignment 02
- Assignment 02 Demo
- Quiz
- Key takeaways







### Course End Projects:

The course includes four real-world, industry-based projects. Successful evaluation of one of the following projects is a part of the certification eligibility criteria:

#### Project 1: Products rating prediction for Amazon

Amazon, one of the leading US-based e-commerce companies, recommends products within the same category to customers based on their activity and reviews of similar products. Amazon would like to improve this recommendation engine by predicting ratings for the nonrated products and add them to recommendations accordingly.

Domain: E-commerce

#### Project 2: Demand Forecasting for Walmart

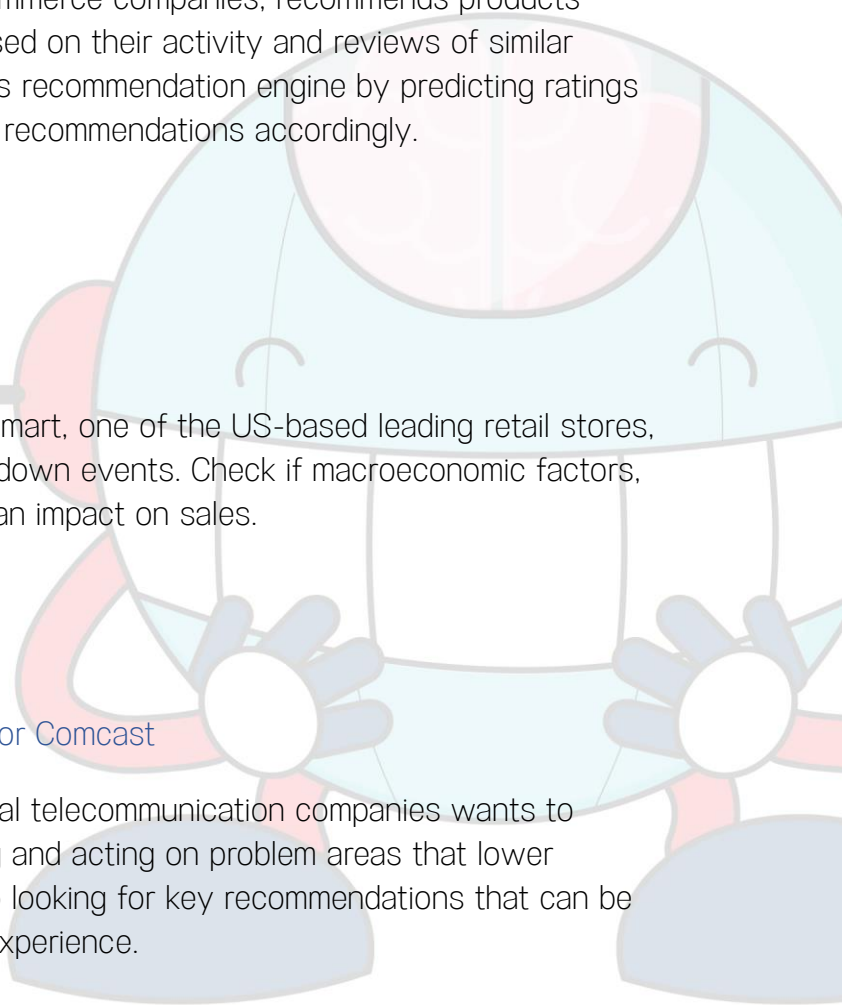
Predict accurate sales for 45 stores of Walmart, one of the US-based leading retail stores, considering the impact of promotional markdown events. Check if macroeconomic factors, such as CPI and unemployment rate, have an impact on sales.

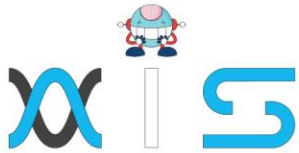
Domain: Retail

#### Project 3: Improving Customer Experience for Comcast

Comcast, one of the largest US-based global telecommunication companies wants to improve customer experience by identifying and acting on problem areas that lower customer satisfaction. The company is also looking for key recommendations that can be implemented to deliver the best customer experience.

Domain: Telecom





#### Project 4: Attrition Analysis for IBM

IBM, one of the leading US-based IT companies, would like to identify the factors that influence attrition of employees. Based on the parameters identified, the company would also like to build a logistics regression model that can help predict if an employee will churn or not.

Domain: Workforce Analytics

#### Domain: Workforce Analytics

Perform a service request data analysis of New York City 311 calls. You will focus on data wrangling techniques to understand patterns in the data and visualize the major complaint types

Domain: Telecommunication

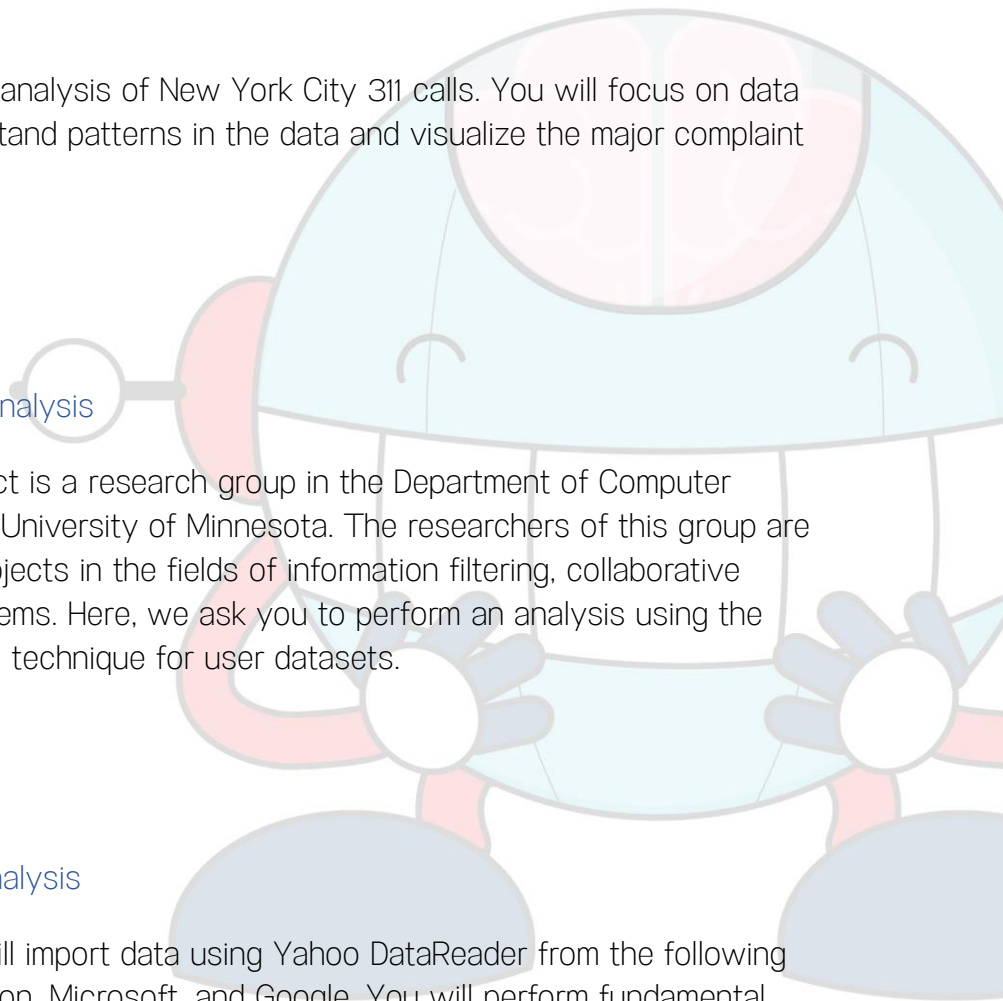
#### Project 6: MovieLens Dataset Analysis

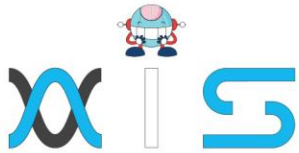
The GroupLens Research Project is a research group in the Department of Computer Science and Engineering at the University of Minnesota. The researchers of this group are involved in several research projects in the fields of information filtering, collaborative filtering, and recommender systems. Here, we ask you to perform an analysis using the exploratory data analysis (EDA) technique for user datasets.

Domain: Engineering

#### Project 7: Stock Market Data Analysis

As a part of this project, you will import data using Yahoo DataReader from the following companies: Yahoo, Apple, Amazon, Microsoft, and Google. You will perform fundamental





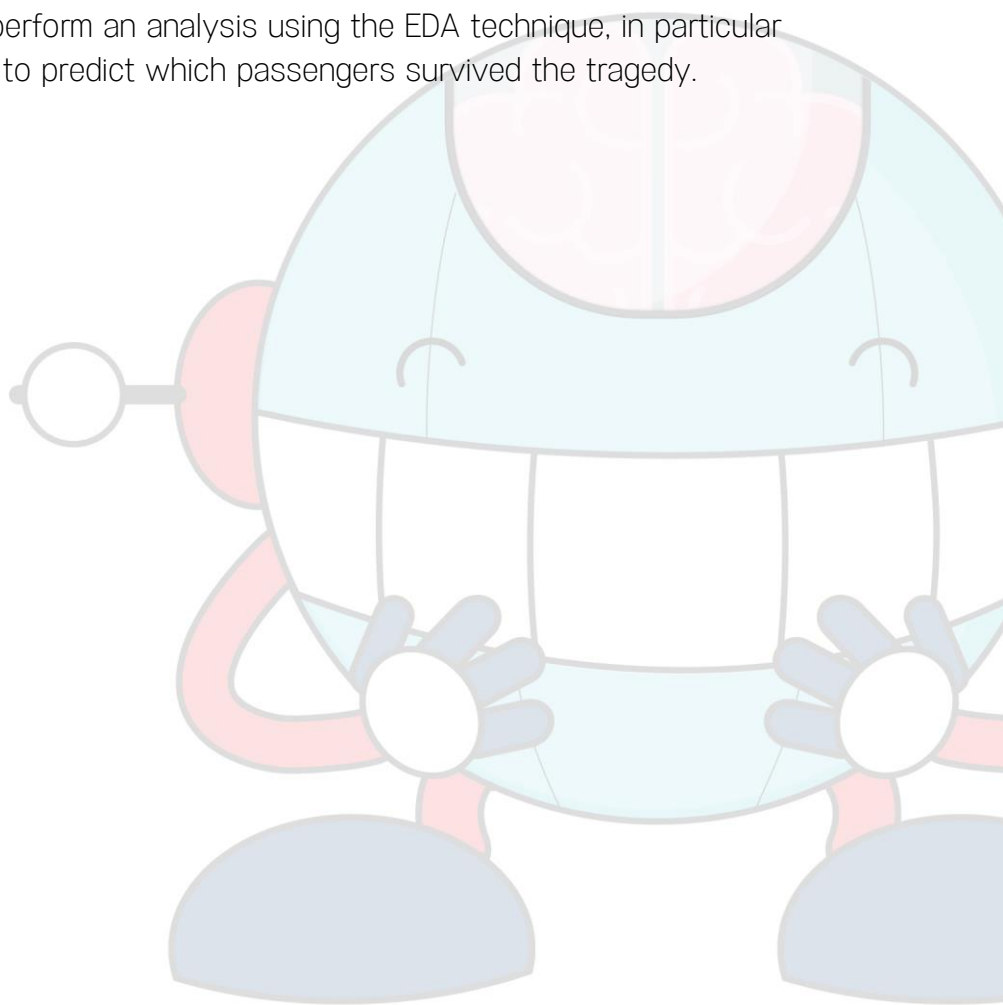
analytics, including plotting, closing price, plotting stock trade by volume, performing daily return analysis, and using pair plot to show the correlation between all of the stocks

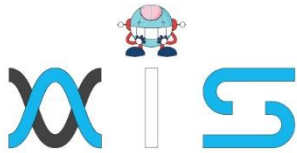
Domain: Stock Market

### Project 8: Titanic Dataset Analysis

On April 15, 1912, the Titanic sank after colliding with an iceberg, killing 1502 out of 2224 passengers and crew. This tragedy shocked the world and led to better safety regulations for ships. Here, we ask you to perform an analysis using the EDA technique, in particular applying machine learning tools to predict which passengers survived the tragedy.

Domain: Hazard





## AIS

With the rising demand in scalable technology, AIS provides tailored goal-setting based on your organization's needs and expectations. With products geared towards the growing needs of your organization's customers and employees, AIS has innovative and value-driven solutions.

### Learning Partners



#### SAS Management, Inc.

SAS Management, Inc. has been a leader in the training, certification and consulting services industry in the Philippines. With offices in Makati, Ortigas and Cebu, Philippines, the company has been the top training and consulting company for over 9 years. In its years of existence, it has provided certificates to over 100,000 professionals including 10,000 participants in the 160+ courses that have been conducted just during the 2020-2021 pandemic alone. The most sought-after courses being offered are ITIL, Project Management Professional (PMP)®, COBIT 5, Six Sigma Yellow Belt, Six Sigma Green Belt, PRINCE2, and organizational skills, among others.



#### Simplilearn

Simplilearn, based in California and India, is the world's #1 online bootcamp and one of the world's the leading certification and training providers. Simplilearn trained over 2,000,000 professionals with 2,000+ trainers and provides over 400 different courses.

### Technology Partner



#### Freshworks

With over 40,000 customers, Freshworks is a business solutions software provider with products ranging from IT helpdesk and service desk, customer relationship management, live chat, marketing automation, phone system, and HR.

### Partners & Affiliates

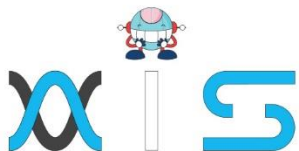


PECB

SDI Service Desk Institute

EC-COUNCIL





*The program presented, as well as a wide range of programs are products of the partnership of AIS, SAS Management, Inc., and Simplilearn. With a common goal of providing the best practice of facilitating learning with an array of choices available in the online platform in this time of digitization.*

## Office Addresses

### NXT AIS Inc.

Juan Antonio III Building  
315 Maysilo Circle, Mandaluyong, Philippines

### SAS Management, Inc.

- **Head Office**

Unit 1602 Cityland 10 Tower  
2 6815 HV dela Costa cor Ayala Avenue  
Makati City, Philippines

- **Branch Office**

Unit 204 Villa Building  
Jupiter St. corner Makati Avenue  
Makati City, Philippines

- **Branch Office**

Room 201 2nd floor OLS Building  
443 Gorordo Avenue., Brgy. Kamputhaw  
Cebu City, Philippines

- **Branch Office**

Malaysia - SAS Management  
Sdn Bhd 6f-16, IOI Business Park,  
Persiaran Puchong Jaya  
47170, Selangor

### Simplilearn

- **Simplilearn Americas, Inc Pvt Ltd.**

201 Spear Street, Suite 1100,  
San Francisco, CA 94105 United States

- **Simplilearn Solutions Pte Ltd.**

NALANDA 53/1 C, Manoj Arcade, 24th Main Rd,  
Sector 2, HSR Layout, Bengaluru  
560102, Karnataka, India.

- **Simplilearn Singapore**

#14-302, The Plaza, 7500A Beach Rd,  
Singapore 199591