



PYTHON FOR DATA SCIENCE

PROF. RAGHUNATHAN RENGASAMY

Department of Chemical Engineering
IIT Madras

PRE-REQUISITES : Knowledge of basic data science algorithms

INTENDED AUDIENCE : Final Year Undergraduates

COURSE OUTLINE :

The course aims at equipping participants to be able to use python programming for solving data science problems

ABOUT INSTRUCTOR :

Prof. Rengaswamy was a professor of Chemical Engineering before joining at IIT Madras as a professor and Co-Director of the Process Control and Optimization Consortium at Texas Tech University, Lubbock, USA. He was also a professor and associate professor at Clarkson University, USA and an assistant professor at IIT Bombay. His major research interests are in the areas of fault detection and diagnosis and development of data science algorithms for manufacturing industries.

COURSE PLAN :

Week 1:

BASICS OF PYTHON SPYDER (TOOL)

- Introduction Spyder
- Setting working Directory
- Creating and saving a script file
- File execution, clearing console, removing variables from environment, clearing environment
- Commenting script files
- Variable creation
- Arithmetic and logical operators
- Data types and associated operations

Week 2:

Sequence data types and associated operations

- Strings
- Lists
- Arrays
- Tuples
- Dictionary
- Sets
- Range

NumPy

- ndarray

Week 3:

Pandas dataframe and dataframe related operations on Toyota Corolla dataset

1. Reading files
2. Exploratory data analysis
3. Data preparation and preprocessing

Data visualization on Toyota Corolla dataset using matplotlib and seaborn libraries

1. Scatter plot
2. Line plot
3. Bar plot
4. Histogram
5. Box plot
6. Pair plot

Control structures using Toyota Corolla dataset

1. if-else family
2. for loop
3. for loop with if break
4. while loop
5. Functions

Week 4: CASE STUDY

Regression

1. Predicting price of pre-owned cars

Classification

1. Classifying personal income