Data Science

Python

Duration: 6 Months

Introduction To Python

- Why Python
- Application areas of python
- Python implementations
  - Cpython
  - Jython
  - Ironpython
  - Pypy
- Python versions
- Installing python
- Python interpreter architecture
  - Python byte code compiler
  - Python virtual machine(pvm)

Writing and Executing First Python Program

- Using interactive mode
- Using script mode
  - General text editor and command window
  - Idle editor and idle shell
- Understanding print() function
- How to compile python program explicitly

Python Language Fundamentals

- Character set
- Keywords
- Comments
- Variables
- Literals
- Operators
- Reading input from console
- Parsing string to int, float

Python Conditional Statements

- If statement
- If else statement
- If elif statement
- If elif else statement
- Nested if statement

Looping Statements

- While loop
- For loop
- Nested loops
- Pass, break and continue keywords

Standard Data Types

- Int, float, complex, bool, nonetype
- Str, list, tuple, range
- Dict, set, frozenset

String Handling

- What is string
- String representations
- Unicode string
- String functions, methods
- String indexing and slicing
- String formatting

Python List

- Creating and accessing lists
- Indexing and slicing lists
- List methods
- Nested lists
- List comprehension

Python Tuple

- Creating tuple
- Accessing tuple
- Immutability of tuple

Python Set

- How to create a set
- Iteration over sets
- Python set methods
- Python frozenset

Python Dictionary

- Creating a dictionary
- Dictionary methods
- Accessing values from dictionary
- Updating dictionary
- Iterating dictionary
- Dictionary comprehension

Python Functions

- Defining a function
- Calling a function
- Types of functions
- Function arguments
  - Positional arguments, keyword arguments
  - Default arguments, non-default arguments
  - Arbitrary arguments, keyword arbitrary arguments
- Function return statement
- Nested function
- Function as argument
- Function as return statement
- Decorator function
- Closure
- Map(), filter(), reduce(), any() functions
- Anonymous or lambda function

Modules & Packages

- Why modules
- Script v/s module
- Importing module
- Standard v/s third party modules
- Why packages
- Understanding pip utility

File I/O

- Introduction to file handling
- File modes
- Functions and methods related to file handling
- Understanding with block

Regular Expressions(Regex)

- Need of regular expressions
- Re module
- Functions /methods related to regex
- Meta characters & special sequences

Object Oriented Programming

- Procedural v/s Object Oriented Programming
- OOP Principles
- Defining a Class & Object Creation
- Inheritance
- Encapsulation
- Polymorphism
- Abstraction
- Garbage Collection
- Iterator & Generator

Exception Handling

- Difference Between Syntax Errors and Exceptions
- Keywords used in Exception Handling
  - try , except , finally , raise , assert
- Types of Except Blocks
- User-defined Exceptions

GUI Programming

- Introduction to Tkinter Programming
- Tkinter Widgets
- Layout Managers
- Event handling
- Displaying image

Multi-Threading Programming

- Multi-processing v/s Multi-threading
- Need of threads
- Creating child threads
- Functions /methods related to threads
- Thread synchronization and locking

SQL Using MySQL

Introduction to RDBMS

- What is Relational Database Package
- Difference between SQL & Database
- Installing MySQL Server database

SQL Basic

- DDL: Create, Alter, Drop, etc.
- DML: Insert, Update, Delete ,etc.
- DQL : Select
- Autoincrement field
- SQL Comments
- SQL Aliases
- Savepoint & rollback

SQL Constraints

- Not NULL, Unique key
- Primary key, Check

SQL Operators

- Arithmetic operators
- Logical operators
- Conditional operators
- Like, between, in operators

SQL Clauses

- Order by
- Where
- Limit/top
- Group by
- having

SQL Joins

- Inner Join
- Left Join
- Right Join
- Full Join

SQL View

- creating view
- updating view
- fetching data from view

SQL Functions

- String functions
- Aggregate functions
- Date & time functions

Stored Procedures & Functions

- Understanding stored procedures and their key benefits
- Working with stored procedures
- Studying user-defined functions

Working with CSV Files:

- How to write result to csv file
- How to read csv file

Python Database Connectivity

- Database Drivers and connectors
- Creating connection object
- Understanding cursor object
- Executing SQL statements using cursor
- Fetching records from cursor
- Storing and retrieving Date and Time

MONGODB

Introduction To MongoDB

- Understanding NoSQL DB
- NoSQL vs. SQL DB
- Understanding Mongo DB
- Downloading & Installation
- Introduction of MongoDB shell and Compass
- Understanding database, collection & document

Crud Operations

- Insert Document
- Delete Document

- Update Document
- Query Document

Operators In MongoDB

- Query and Projection operators
- Update operator
- Aggregation Pipeline operators

Methods In MongoDB

- limit and sort
- bulk methods
- other methods

Indexing And Relationships

- Types of Indexes
- Creating an Indexes
- Dropping an Indexes
- Defining Relationships between Documents

Python Connectivity With MongoDB

- Introduction to pymongo
- Installing pymongo module
- MongoClient
- Getting database and collection
- CRUD operations
- Range Queries

Statistics & Analytics

Introduction to Statistics

- Sample or Population
- Measures of Central Tendency
- Arithmetic Mean
- Harmonic Mean
- Geometric Mean
- Mode
- Quartiles
- Variance & Standard Deviation
- Outliers

Data Distributions

- Normal Distribution
- Uniform Distribution
- Exponential Distribution
- Right & Left Skewed Distribution
- Random Distribution
- Central Limit Theoremndard Deviation

Numpy Package

- Difference between list and numpy array
- Vector and Matrix operations
- Array indexing and slicing

Pandas Package

Introduction To Pandas

- Labeled and structured data
- Series and DataFrame Objects

How To Load Datasets

- From excel
- From csv
- From html table
- From database table

Accessing Data From Data Frame

- at & iat
- loc & iloc
- head() & tail()

Exploratory Data Analysis (EDA)

- Describe()
- Groupby()
- Crosstab()
- boolean slicing / query()

Data Manipulation & Cleaning

- Map(),apply()
- Combining data frames
- Adding/removing rows & columns
- Sorting data
- Handling missing values
- Handling duplicacy
- Handling Data Error

Categorical Data Encoding

- Label Encoding
- One Hot Encoding

Categorical Data Encoding

- Label Encoding
- One Hot Encoding

Handling Date and Time

Data Visualization Using Matplotlib And Seaborn Packages

- Scatter plot, lineplot, bar plot
- Histogram, pie chart,
- Jointplot, pairplot ,Heatmap
- Outlier detection using boxplot

Power BI

INTRODUCTION TO POWER BI

- Introduction to Business Intelligence (BI)
- Various BI tools
- Introduction to Power BI
- Why Power BI
- Power BI Components
- Introduction of Power BI Desktop
- Installation of Power BI Desktop

DATA VISUALIZATION

Understanding Power View and Power Map
Data visualization techniques
Page layout & Formatting

- Power BI Desktop visualization
- Formatting and customizing visuals
- Column chart, Pie chart, Donut chart,
- Scatter chart, Funnel chart
- Include & exclude
- Geographical data visualization using Maps
- Drill down
- Drill through
- Page navigations
- Bookmarks
- Selection pane to show/hide visuals
- Comparing volume and value-based analytics
- Combinations charts (dual axis charts)
- Filter pane
- Slicers
- Use of Hierarchies in drill down analysis
- Sync slicers
- Tooltips & custom tooltips
- Tables & matrix
- Conditional formatting on visuals

POWER BI SERVICE, PUBLISING & SHARING

- Introduction to Power BI Service
- Introduction of workspaces
- Dashboard
- Creating & Configuring Dashboards
- Dashboard theme
- Reports vs Dashboards
- Sharing reports & dashboards

DATA TRANSFORMATION – SHAPING & COMBINING DATA

- Shaping data using Power Query Editor
- Formatting data
- Transformation of data
- Understanding of Data types
- Naming conventions & best practices to consider
- Working with Parameters
- Merge Query
- Append Query
- Group by of data (aggregation of data)
- Duplicate & Reference tables
- Fill
- Pivot & Un-pivot of data
- Custom columns
- Conditional columns
- Replace data from the tables
- Split columns values
- Move columns & sorting of data
- Detect data type, count rows & reverse rows
- Promote rows as column headers
- Hierarchies in Power BI

DATA MODELING & DAX

- Introduction of relationships
- Creating relationships
- Cardinality
- Cross filter direction
- Use of inactive relationships
- Introduction of DAX
- Why DAX is used
- DAX syntax
- DAX functions
- Context in DAX
- Calculated columns using DAX
- Measures using DAX
- Calculated tables using DAX
- Learning about table, information, logical, text, iterator,
- Time intelligence functions (YTD, QTD, MTD)
- Cumulative values, calculated tables, and ranking and rank over groups
- Date and time functions

Machine Learning

Introduction To Machine Learning

- Traditional v/s Machine Learning Programming
- Real life examples based on ML
- Steps of ML Programming
- Data Preprocessing revised
- Terminology related to ML

Supervised Learning

- Classification
- Regression

Unsupervised Learning

- clustering

KNN Classification

- Math behind KNN
- KNN implementation
- Understanding hyper parameters

Performance metrics

- Confusion Matrix
- Accuracy Score
- Recall & Precision
- F-1 Score
- R2 Score

Regression

- Math behind Regression
- Simple Linear Regression
- Multiple Linear Regression
- Polynomial Regression
- Boston Price Prediction
- Cost or Loss Functions

- Mean absolute error
- Mean squared error
- Root mean squared error
- Least Square Error

- Regularization

Logistic Regression for classification

- Theory of Logistic Regression
- Binary and Multiclass classification
- Implementing titanic dataset
- Implementing iris dataset
- Sigmoid and softmax functions

Support Vector Machines

- Theory of SVM
- SVM Implementation
- kernel,gamma,alpha

Decision Tree Classification

- Theory of Decision Tree
- Node Splitting
- Implementation with iris dataset
- Visualizing Tree

Ensemble Learning

- Random Forest
- Bagging and Boosting
- Voting Classifier

Model Selection Techniques

- Cross Validation
- Grid and Random Search for hyper parameter tuning

Recommendation System

- Content based technique
- Collaborative filtering technique
- Evaluating similarity based on correlation
- Classification-based recommendations

Clustering

- K-means Clustering
- Hierarchical Clustering
- Elbow technique
- Silhouette coefficient
- Dendogram

Text Analysis

- Install NLTK
- Tokenize words
- Tokenizing sentences
- Stop words customization
- Stemming and Lemmatization
- Feature Extraction
- Sentiment Analysis
- Count Vectorizer
- TfidfVectorizer
- Naive Bayes Algorithms

Dimensionality Reduction

- Principal Component Analysis(PCA)

Open CV

- Reading images
- Understanding Gray Scale Image
- Resizing image
- Understanding Haar Classifiers
- Face , eyes classification
- How to use webcam in open cv
- Building image data set
- Capturing video
- Face classification in video

Projects

- One project using Python & SQL
- One project using Python & ML
- One Dashboard using Power bi