Data Science / Data Analyst Syllabus

Course Duration : 3 Months

1. Introduction
   - Compilation v/s Interpretation
   - Script mode and Interactive mode
   - Command Line Arguments
2. Data Types
   - Basic Data types
   a. Numbers (int, float, complex)
   b. Strings
   c. Bool
   - Advance Data types (List, tuple, set, dictionary)
   - Type casting
   a. Implicit
   b. Explicit
3. Functions
   - Types of Functions
   a. User Defined Functions
   b. Built-in Functions
   c. Lambda Function
   i. Filter
   ii. Reduce
   iii. Map
   d. Recursive Function
   - Doc String
   - Types of Arguments
   a. Positional arguments
   b. Default arguments
   c. Keyword arguments
   d. Variable length arguments
   e. Variable length Keyword argument

Python for Data Science

4. Modules in Python
   - Importing a Module using alias
   - Importing using from keyword
   - Input()
   - name () and main ()
   - Turtle Module
   - Math module
5. List
   - Creation of lists
   - Accessing list elements
   - List slicing
   - List replication

- Appending two list
  a. append()
  b. extend()
  c. using '+' operator
- Removing an element from a list
  a. pop()
  b. del keyword
- Reference Type Assignment
- Copy Operation using memory map
  a. Shallow copy
  b. Deep copy
- List Comprehension
  a. Using for loop
  b. Using list comprehension
  c. List comprehension using single if condition
  d. List comprehension with multiple if conditions
  e. List comprehension using else condition
- Accessing list
  a. Using for loop
  b. Using range()
  c. Accessing elements present within nested list
- Reversing a list
- List Comparison
- List Sorting

  a. Ascending order
  b. Descending order
- Membership Check of List

6. Tuples
   - Membership Check of List
     a. Creation of tuple
     b. Creation of singleton tuple
     c. Packing and Unpacking
     d. Unpacking using disposable variable
     e. Accessing elements within a tuple
     f. Tuple Slicing
     g. Copy operation in tuple
   - List and tuple Comparison
7. Set
   - Creation of set
   - Set operations
     a. Union
     b. Intersection
     c. Difference
     d. Symmetric Difference
     e. Subset
     f. Super set
     g. Disjoint set
   - Set methods
     a. add()

b. discard()
c. remove()
· Frozen set
· Set Comprehension
a. Using for loop
b. Using set comprehension
c. set comprehension using single if condition
d. set comprehension with multiple if conditions
e. set comprehension using else condition
· All and Any function
· Internal Implementation of List

· List performance analysis
· When to use a List.
· Internal Implementation of tuple
· Performance Analysis
· Difference between list and tuple
· Internal Implements of set
· Performance analysis of set
· Difference between list and set
· Difference between tuple and set

8. Dictionary
· Internal Implementation of Dictionary
· Creation of Dictionary
· Adding elements to a dictionary
· Accessing elements from a dictionary
· Accessing values from a dictionary using get()
· Different ways of deleting elements from a dictionary
a. pop()
b. popitem()
c. del keyword
d. clear()
· Different ways of accessing a dictionary
e. keys()
f. values()
g. items()
· Different ways of iterating over a dictionary
a. keys()
b. values()
c. items()
· Membership check in a dictionary
· Merging of dictionaries
a. Using update()
b. Using **
· Dictionary Comprehensions
a. Using for loop
b. Using dictionary
c. Dictionary comprehension using single if condition

d. Dictionary comprehension with multiple if conditions

e. Dictionary comprehension using else condition
· When to use a Dictionary
· Zip()
a. Zip() function on list of varying length
· Difference between List, tuple, set and dictionary

9. Collections Module
· Dequeue
· Named Tuple
· Ordered Dictionary
· Default Dictionary
· Chain map
· Counter

10. String
· Different ways of creating a string
· Internal Implementation of String
· String Formatting
a. Default formatting
b. Positional formatting
c. Keyword formatting
d. Binary formatting
e. % Formatting Specifier
· Built-in functions in String
a. lower()
b. upper()
c. title()
d. capitalize()
e. swapcase()
f. maketrans()
g. translate()
h. split()
i. startswith()
j. endswith()
· Accessing individual character of a String
a. Forward direction
b. Reverse direction

· String Comparison
a. Using values
b. Using Reference
c. Ignoring case
d. Difference between casefold() and lower()
· String Concatenation
a. Using '+' Operator
b. Using join()
c. Using format()
d. Using 'f' string literal

11. Regular Expressions
· Quantifiers
· Word Character
· Character class

- Grouping
- Raw String
- Re Module

a. match()
b. search()
c. Difference between match() and search()
d. findall()
e. sub()
f. split()
g. start()
h. end()
i. group()
j. compile()

- Flags

12. Exception Handling
    - User Defined Exception Handler
    - Disadvantage of having single Except block
    - Grouping multiple exceptions in a single except block
    - Printing the default message of exception by aliasing exception object
    - Optional else block
    - Propagation of exception object
    - Use of 'raise' keyword in exception handling
    - Handling the exception using try-except blocks
    - Re-throwing an exception
    - Valid and in-valid syntax of try-except block
    - Difference between exception handling in python and java
    - Customized Exception in python
    - Exception Hierarchy
    - Scope of variable in python

    a. Accessing global variable within the function
    b. Difference between globals() and locals() methods
    c. Nested function scope in python

13. Loggers
    - Levels of Loggers

    a. INFO
    b. DEBUG
    c. ERROR
    d. WARNING
    e. CRITICAL

    - Formatters in loggers
    - Traceback error log file
    - Levels for respective file handling

    a. ERROR
    b. INFO
    c. DEBUG

14. File Handling in Python
    - Reading a file
    - Modes of file
    - Closing a file
    - Context Managers in python

- Reading file contents line by line
- Reading a single line in a file
- Reading multiple lines in a file
- Reading a character in a line
- readline()
- Cursor Position

a. tell()
b. seek()

- Writing a file
- File modes in python
- Read
- Write
- Create
- Erase
- Position
- Exclusive Creation
- Exclusive Mode operation
- File to File transfer

15. Os Modules
16. Object Orientation in Python
    - Object creation in python
    - Creation of instance variable in python
    - new ()
    - init ()
    - self keyword in python
    - static and non-static methods in python
17. Iterators and Generators
    - Iterators

    a. Difference between containers and non-containers
    b. iter () and next ()
    c. Iter tools module

    i. Islice
    ii. cycle

    - Examples
    - Generators

    a. Difference between normal function and iterator function
    b. Use of 'yield' keyword in python
    c. Control flow diagram of iterators and generators
18. First class functions
19. Closures
20. Decorators

21. Oops Concept

- Encapsulation
- Polymorphism
- Inheritance
- Abstraction

Introduction to Pandas

1. Need For DataFrame
2. Creating DataFrame
3. Series
4. Inspecting Dataframes and Must know operations
5. Inspecting Dataframes
6. Renaming Columns
7. Pandas Summary
8. Display Options
9. Conditional filtering and sorting
10. Filtering Data That Satisfy Conditions
11. Membership Filtering
12. Query and Eval
13. Removing Duplicates
14. Sorting
15. Map and Applymap
16. Data preparation and transformation
17. Apply a function rowwise or columnwise
18. Scaling and Standardization
19. Make Index as a Dataframe Column
20. Discretization and Binning
21. Useful tips and tricks
22. Random Sampling
23. Dummy Variables
24. Categorical Data Part-1
25. Categorical Data Part-2
26. Efficiently Read Data From Multiple Files
27. Data grouping and aggregation
28. Group by Mechanism
29. Iterating Between Groups
30. Transform
31. Reshaping and pivoting data
32. Cross Tabulation
33. Pivoting
34. Combining dataframes
35. Joining Dataframes
36. Types of Joins
37. Concatenating Dataframes
38. Data cleaning and transformations
39. Representing Missing Values
40. Threshold Based Dropping
41. Approaches To Filling Missing Data
42. Interpolation
43. Compressed File Formats
44. Sparse Datatype

45. Combining Categories
46. Split Contents of a Column
47. Insert Column at Specific Location
48. Select using both Position and Lab
49. Interactive Data Analysis
50. Optimizing dataframes
51. Optimizing Dataframes
52. Handling Large Data
53. Sampling On Load
54. Efficient File Formats
55. HDF5
56. Chunking
57. Load to Database
58. Data Visualization
59. Matplotlib Part 1 - Getting Started
60. Matplotlib Part 2 - Plot Components
61. Matplotlib Part 3 - Subplots
62. Matplotlib Part 4 - Annotations
63. Dual Axis Line Plots
64. Bar Charts
65. Histogram and Density Plots
66. Regression Plots
67. Pair Plots
68. Need For DataFrame
69. Creating DataFrame
70. Series
71. Inspecting Dataframes and Must know operations
72. Inspecting Dataframes
73. Renaming Columns
74. Pandas Summary
75. Display Options
76. Conditional filtering and sorting
77. Filtering Data That Satisfy Conditions
78. Membership Filtering
79. Query and Eval
80. Removing Duplicates
81. Sorting
82. Map and Applymap
83. Data preparation and transformation
84. Apply a function rowwise or columnwise
85. Scaling and Standardization
86. Make Index as a Dataframe Column
87. Discretization and Binning
88. Useful tips and tricks
89. Random Sampling
90. Dummy Variables
91. Categorical Data Part-1
92. Categorical Data Part-2
93. Efficiently Read Data From Multiple Files
94. Data grouping and aggregation

95. Group by Mechanism
96. Iterating Between Groups
97. Transform
98. Reshaping and pivoting data
99. Cross Tabulation
100. Pivoting
101. Combining dataframes
102. Joining Dataframes
103. Types of Joins
104. Concatenating Dataframes
105. Data cleaning and transformations
106. Representing Missing Values
107. Threshold Based Dropping
108. Approaches To Filling Missing Data
109. Interpolation
110. Compressed File Formats
111. Sparse Datatype
112. Combining Categories
113. Split Contents of a Column
114. Insert Column at Specific Location
115. Select using both Position and Lab
116. Interactive Data Analysis
117. Optimizing dataframes
118. Optimizing Dataframes
119. Handling Large Data
120. Sampling On Load
121. Efficient File Formats
122. HDF5
123. Chunking
124. Load to Database
125. Data Visualization
126. Matplotlib Part 1 - Getting Started
127. Matplotlib Part 2 - Plot Components
128. Matplotlib Part 3 - Subplots
129. Matplotlib Part 4 - Annotations
130. Dual Axis Line Plots
131. Bar Charts
132. Histogram and Density Plots
133. Regression Plots
134. Pair Plots

Statistics Basics •

Central Tendency

o Mean
o Median
o Mode
o Skewness

o Normal Distribution
• **Probability Basics**
o What does it mean by probability?
o Types of Probability
o ODDS Ratio?
• Standard Deviation
o Data deviation & distribution
o Variance • Bias variance Tradeoff
o Underfitting
o Overfitting • Distance metrics
o Euclidean Distance
o Manhattan Distance • Outlier analysis
o What is an Outlier?
o Inter Quartile Range
o Box & whisker plot
o Upper Whisker
o Lower Whisker
o Scatter plot
o Cook's Distance
• Missing Value treatment
o What is NA?
o Central Imputation
o KNN imputation
o Dummification
• Correlatio
Pearson correlation
o positive & Negative correlation

Module 1:

Supervised Learning •
Linear Regression o
Linear Equation
o Slope
o Intercept
o R square value
• Logistic regression
o ODDS ratio
o Probability of success
o Probability of failure Bias Variance Tradeoff
o ROC curve o Bias Variance Tradeoff

Module 2: Unsupervised Learning

• K-Means
• K-Means ++
• Hierarchical Clustering

Module 3: SVM

- Support Vectors
- Hyperplanes
- 2-D Case
- Linear Hyperplane

Module 4: SVM Kernal

- Linear
- Radial
- polynomial

Module 5: Other Machine Learning Algorithms

- K – Nearest Neighbour
- Naïve Bayes Classifier
- Decision Tree – CART
- Decision Tree – C50
- Random Forest

Deep Learning Module

1: Deep Learning Algorithms

- CNN – Convolutional Neural Network
- RNN – Recurrent Neural Network
- ANN – Artificial Neural Network

Module 2: Introduction to NLP

- Text Pre-processing
- Noise Removal
- Lexicon Normalization
- Lemmatization
- Stemming
- Object Standardization

Module 3: Text to Features (Feature Engineering)

- Syntactical Parsing
- Dependency Grammar
- Part of Speech Tagging
- Entity Parsing
- Named Entity Recognition
- Topic Modelling
- N-Grams
- TF – IDF
- Frequency / Density Features
- Word Embedding’s

Module 4: Tasks of NLP

- Text Classification
- Text Matching
- Levenshtein Distance
- Phonetic Matching
- Flexible String Matching

NEWGEN CORPORATE TRAINING CENTER

◆◆OFFLINE/ONLINE BATCH◆◆

Address: NEWGEN CORPORATE TRAINING CENTER, Above Radhika Hotel,
. Near Parihar Chowk, Aundh, Pune – 411007

URL: http://www.newgensoftech.com

Mail Id: balkrishna8588@gmail.com

Insta Id: @Newgen_Softech