

Information Booklet cum Syllabus

Of

Data Science using Python

Revision-I



May2022

National Institute of Electronics and Information Technology

An Autonomous Scientific Society under
Ministry of Electronics and Information Technology, Government of India

NIELIT Gorakhpur
Deoria Road
Gorakhpur (U.P.) -273010

NIELIT Extension Centre Lucknow
NIELIT Lucknow
Sumit Complex, A-1/9, VibhutiKhand,
Gomti Nagar, Lucknow – 226010

CONTENTS

Sl. No.	Title	Page No.
1.	About the course	
2.	NIELIT	
3.	Objective of Course	
4.	Job Role of Course	
5.	Eligibility	
6.	Duration of Course	
7.	Practical/ Tutorial	
8.	Credits of Course	
9.	Examination Pattern	
10.	Hardware Requirement	
11.	Software Requirement	
12.	Course Details	
13.	Reference Books / Study Material	
14.	Practical Assignments	
15.	Sample Questions	

1. **About Course**

This course teaches to master the concepts of Python programming, data analysis and machine learning. In this course, student will gain knowledge in data analysis, Machine Learning and Data Visualization. After the course, student will master the essential tools of Data Science with Python.

2. **NIELIT**

National Institute of Electronics and Information Technology, NIELIT, (Erstwhile DOEACC Society) is an autonomous scientific society of the Ministry of Electronics & Information Technology, Government of India. The Society is registered under the Societies Registration Act, 1860. NIELIT was set up to carry out Human Resource Development and related activities in the area of Information, Electronics & Communications Technology (IECT). NIELIT is engaged both in Formal & Non-Formal Education in the areas of IECT besides development of industry oriented quality education and training programmes in the state-of-the-art areas. NIELIT has endeavored to establish standards to be the country's premier institution for Examination and Certification in the field of IECT. It is also one of the National Examination Body, which accredits institutes/organizations for conducting courses in IT and Electronics in the non-formal sector.

3. **Objective of Course**

Through this training, student will gain knowledge in data analysis, Machine Learning and data visualization.

After completing the module, the learner will be able to:

- Introduction to Python Programming and statements.
- Understand and master the concept of Modular Programming, using Function, Module and Packages.
- Understand and master the concept of Sequence Data Types – List, Set, Dictionaries.
- Understand and master the concept of NumPy and Matplotlib.
- Understand the concept of Statistical functions.
- Understand the concepts of Data Science and Analysis tools- Pandas
- Understand the concept of Machine Learning
 - Supervised
 - Regression Algorithms
 - Classification Algorithms
 - Unsupervised
 - Clustering Algorithms

4. **Job Roles of Course**

This Course is designed to equip a learner with necessary skills as per following job roles:

- Junior Data Analyst
- Python Programmer
- Python Associate

5. **Eligibility**

Pursing Diploma/Graduation.

6. Total duration of the Course

80 Hours(Theory: 30 Hrs, Practical/Tutorial: 50 Hrs)

7. Course Details

7.1.Course Outline and Objective of Each Unit

S. No.	Unit Name	Duration (Theory) in Hours	Duration (Practical) in Hours	Total Learning Hrs.	Learning Objectives
1	Module 1 : Python	10	15	25	After completion of this module, learner will be able to <ul style="list-style-type: none">• Understand the concept of Programming.• Use the various constructs of Python viz. conditional, iteration• Write programs making judicious use of Lists, Strings, Tuples, Dictionaries wherever required• Handle files and create Modules in Python
2	Module 2: Data Science	10	20	30	After completing this module the participant will be able to: <ul style="list-style-type: none">• Know about the packages like NumPy and Pandas which are essential to solve machine learning problems and perform data manipulation and analysis before machine learning is applied.• Grab raw data, clean it and make it ready for building machine learning models• Visualization of the data using Matplotlib

3	Module 3: Machine Learning	10	15	25	<p>After completion of this unit of module, the Learner will be able to</p> <ul style="list-style-type: none"> • Understand regression and classification problem • Will be able to solve real life regression problem like admission prediction, house price prediction problem • Learn classification problem like loan prediction. • Will be able to develop machine learning models for regression and classification problems • Know about the performance metrics for evaluation of a Machine Learning Model
---	---------------------------------------	----	----	----	---

7.2. Detailed Syllabus

Unit Name	Contents	Hrs.
Module 1: Python	<ul style="list-style-type: none"> • Python Introduction, Program execution, Using Comments, Literals, Constants, Python's Built-in Data types, Numbers (Integers, Floats, Complex Numbers, Real, Sets), Strings(Slicing, Indexing, Concatenation, other operations on Strings), Accepting input from Console, printing statements • Assignment statement, expressions, Conditional statements: if, if-else, if-elif-else; While Statement, For loop, break statement, Continue Statement, Pass statement. • Lists, tuples and dictionary, (Slicing, Indexing, Concatenation, other operations on Sequence datatype), concept of mutability, Examples of finding the maximum, minimum, mean; linear search on list/tuple of numbers, and counting the frequency of elements in a list using adictionary. • Basic Functions, Module Basics, Module Files as Namespaces, Import Model, Reloading Modules. 	25
Module 2 : Data Science	<ul style="list-style-type: none"> • NumPy Library, Numpy functions, Numpy Indexing, Broadcasting, Storage of images in Numpy, reading csv files in Numpy, Analysis. • Pandas Library, working with dataframes and Series, loading csv, manipulating dataframes, Applying custom functions on dataframes, Aggregation functions, Filtering Data, Analysis. 	30

	<ul style="list-style-type: none"> • Visualization using matplotlib and Seaborn. 	
Module 3 : Machine Learning	<ul style="list-style-type: none"> • Data Preparation, Feature Selection and Engineering, Training Data, Testing Data, Scikit Learn library, Predictive Machine Learning Algorithms - Regression , classification & clustering, • Data Science Concepts • Framing the problem, collecting , processing, cleaning and munging the data • Data analysis and Visualizing results • Machine Learning Basics and its Types • Implementation of Regression, clustering and classification algorithms on data. 	25

8. Reference Books/Study Material

- 1) Python Programming- A modular Approach (with Graphics, database, Mobile and Web Applications by Sheetal Taneja and Naveen Kumar) Pearson.
- 2) Machine Learning using Python by ShriManaranjanPradhan& U Dinesh Kumar
- 3) Head First Python by Paul Berry, O'Reilly
- 4) Beginning Programming with Python Dummies by John Paul Meuller.
- 5) Python for Data Analysis by O'Reilly
- 6) Python Data Science Handbook: Essential Tools for Working with Data By O'Reilly

9. Practical Assignments

Assignment 1:

Create a numpy array with following columns: hindi, English, science, math and commerce with data type int32.

- Insert at least 10 rows in the above array.
- Display size and shape of the array.
- Print sum of each column.
- Print maximum element from each column.
- Print sum of 1,4,5 row.

Assignment 2:

Create an array for a given range of elements between 2 to 40 step by 3.

Assignment 3:

Create an array of size (3, 4) and reshape to (4, 3).

Assignment 4:

- Create two array of size (3, 3) and print their sum and multiplication.
- Create an array of size 10 and calculate square root and standard deviation.
- Print size and dimension of above arrays.

Assignment 5:

Write a Python program to create and display a series of data using Pandas module:

- Create a pandas series of 10 elements and specify their index as 101 to 110.
- Print bottom 5 elements of the series created in question 2.
- Insert 3 new elements in above series on index 111, 112 and 113.
- Delete the elements at index- 103, 104,107,111 in above list.

Assignment 6:

Write a Pandas program to create and display a DataFrame from a specified dictionary data which has the index labels. Sample Python dictionary data and list labels:

- exam_data = {'name': ['Ankita', 'Dia', 'Kapil', 'Jayesh', 'Esha', 'Mayank', 'Ravi', 'Lata', 'Kamal', 'Jatin'],
- 'score': [12.5, 9, 16.5, 15, 9, 20, 14.5, 17.5, 8, 19],
- 'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],
- 'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']}

Assignment 7:

Create a data frame using dictionary.

- Dictionary ('id':[P101,P102,P103,P104,P105], 'Price':[256, 340, 540, 260, 470])
- Print the price of product id – p102.
- Print values of Price column.
- Rename the column id to Product_Id and Price to Base_Price.

Assignment 8:

Create a new data frame with three columns – Product_Name, Cost, Sales.

- Add 10 values in data frame.
- Add a new column named quantity with 10 values.
- Add a new column named: Profit and total_profit and fill values.
- Insert a new column named location after Product_Name column with 10 cities.

(New Delhi, Lucknow, Kolkata, Lucknow, New Delhi, Bengaluru, Chennai, Chennai, Kolkata, Bengaluru)

Assignment 9:

Create a data frame using given data:

'Name': ['Ankit', 'Amit', 'Aishwarya', 'Priyanka', 'Priya', 'Shaurya'],
'Age': [21, 19, 20, 16, 17, 21],
'Stream': ['Math', 'Commerce', 'Science', 'Math', 'Commerce', 'Science'],
'Percentage': [88, 92, 95, 70, 65, 78]}

Perform the following operations on the above dataframe.

- Insert a new row – Name: Sahil, Age-23, Stream- Commerce, Percentage-88.
- Selecting rows where percentage is >80
- Selecting all the rows from the given dataframe in which ‘Stream’ is Commerce and science.
- Selecting all the rows from the given dataframe in which ‘Age’ is greater than 18.
- Print sum of age and percentage only.

Assignment 10:

Download the house price prediction data from:

<https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data?select=test.csv>

Train the data with appropriate model and predict the price of new house.