



Youth Employability Skill Training Co-operative Educational Society Ltd.
AN AUTONOMOUS VOCATIONAL TRAINING INSTITUTION
(NATIONAL YOUTH PROGRAMME)

COURSE NAME:

DIPLOMA IN DATA ANALYSIS IN PYTHON

DURATION:6 Months

SUBJECTS:

<u>SL. NO</u>	<u>SUBJECTS</u>	<u>TOTAL MARKS</u>
1	DATA ANALYSIS WITH PYTHON	50
2	DATA VISUALIZATION WITH PYTHON	50
3	PRACTICAL / VIVA	100

SYLLABUS:

DATA ANALYSIS WITH PYTHON

- Learn how to work with basics of python
- Learn how to work with NumPy datatypes
- Be proficient in pandas Series
- Be proficient in pandas DataFrames

Unit I Introduction to Python

Introduction to Python programming language -Basics of Python: variables, data types, operators - Control structures: if-else, loops - Functions and modules - Lists-Tuples – Dictionaries- Sets

Unit II Introduction to NumPy

Numpy – Data Types Numpy – Array Attributes- Numpy – Array Creation Routines - Numpy – Array From Existing Data - Numpy – Array From Numerical Ranges- Numpy – Indexing and Slicing Numpy – Advanced Indexing

Unit III Numpy Fundamentals and Usage

Numpy – Iterating Over Array- Numpy – Array Manipulation- Numpy – String Functions - Numpy – Mathematical Functions - Numpy – Arithmetic Operations - - Operations on Arrays - Concatenating Arrays - Reshaping Arrays - Splitting Arrays - Statistical Operations on Arrays

Unit IV Pandas

Introduction- Environment Setup- Introduction to Data Structures - Panel - Basic Functionality -Descriptive Statistics- Function Application -Reindexing - Sorting - Working with Text Data -Options & Customization - Window Functions

Unit V Data Wrangling with Pandas

Working with Series and DataFrames - Indexing and selecting data - Data manipulation techniques: merging, concatenating, reshaping - Handling missing data - Data aggregation and summarization

Reference Books

1. “Python for Data Analysis”, DATA WRANGLING WITH PANDAS, NUMPY, AND IPYTHON, II Edition, O’Reilly Media, Author Wes McKinney.
2. “Data Analysis Using Python: A Complete Beginner Guide”, by Liam Foster.
3. “Data Science using Python: A Step-by-Step Practical Approach for Beginners Paperback” by Dr. Vishal Goyal Dr.Monika Bansal, Dr.Munish Jindal, Dr.Harmandeep Kaur

DATA VISUALIZATION WITH PYTHON

- Understand how to use data visualization
- Know how to import and clean data
- Introduce statistical tools for working with data sets
- An introduction to the problems of working with PDF data sources
- Introduce machine learning tools for working with data sets

Unit – I Matplotlib Introduction

Creating Different Types of Plot - Line Graph in Matplotlib - Stem Plot in Matplotlib - Bar chart in Matplotlib - Plotting Histogram in Matplotlib - Scatter Plot in Matplotlib - Stack Plot in Matplotlib -Box Plot in Matplotlib -Pie Chart in Matplotlib- Error Plot in Matplotlib-Violin Plot in Matplotlib -3D Plots in Matplotlib

Unit – II Data Visualization with Matplotlib

Introduction to data visualization - Basic plotting techniques with Matplotlib - Advanced plots and customization

Unit – III Introduction to Seaborn

Using Seaborn with Matplotlib- Customizing Seaborn Plots -Changing Figure Aesthetic - Removal of Spines-Changing the figure Size-Scaling the plots-Setting the Style Temporarily

Unit – IV Color Palette and Plot

Diverging Color Palette - Sequential Color Palette - Setting the default Color Palette - Multiple plots with Seaborn - Using Matplotlib - Using Seaborn - Creating Different Types of Plots-Relational Plots - Categorical Plots - Distribution Plots - Regression Plots

Unit – V Introduction to Machine Learning with Python

Overview of machine learning concepts - Supervised vs unsupervised learning - Introduction to scikit-learn library - Building and evaluating machine learning models - Model selection and hyperparameter tuning

Reference Books

1. Python Data Analytics with Pandas, NumPy, and Matplotlib , Second Edition ,A Press, Fabio Nelli
2. “Data Science using Python: A Step-by-Step Practical Approach for Beginners Paperback” by Dr. Vishal Goyal Dr.Monika Bansal, Dr.Munish Jindal, Dr.Harmandeep Kaur
3. “Python Data Visualization Essentials Guide”, by Kalilur Rahman

PRACTICAL

EXERCISE:

1. Write a Python program to perform the below operations
 - Create a 1D array
 - Extract items from 1D array
 - Replace items with another value in numpy array
2. Write a Python program to compute the Mean, Median, standard deviation of a numpy array.
3. Write a Python program to replace all odd numbers in the given array with -1.
4. Write a Python program to convert a 1-D array into a 2-D array with 3 rows.
5. Write a python program to create an 1D array and add 202 to all the values in given array.
6. Write a python program to normalize an array so the values range exactly between 0 and 1 and Replace all odd numbers in the given array with -1
7. Generate a 1-D array of 10 random integers. Each integer should be a number between 30 and 40 (inclusive)
8. Write a Pandas program to create and display a one-dimensional array-like object containing an array of data using Pandas module.
9. Write a Pandas program to convert Series of lists to one Series.
10. Write a Pandas program to create a dataframe from a dictionary and display it.
Sample data: {'X':[78,85,96,80,86], 'Y':[84,94,89,83,86], 'Z':[86,97,96,72,83]}
11. Write a Pandas program to remove whitespaces, left sided whitespaces and right sided whitespaces of the string values of a given pandas series.
12. Write a Pandas program to convert all the string values to upper, lower cases in a given pandas series. Also find the length of the string values.
13. Write a Pandas program to split the following dataframe into groups based on school code. Also check the type of GroupBy object.

Test Data:

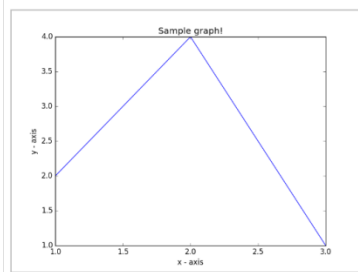
	school	class	name	date_Of_Birth	age	height	weight	address
S1	s001	V	Alberto Franco	15/05/2002	12	173	35	street1
S2	s002	V	Gino Mcneill	17/05/2002	12	192	32	street2
S3	s003	VI	Ryan Parkes	16/02/1999	13	186	33	street3
S4	s001	VI	Eesha Hinton	25/09/1998	13	167	30	street1
S5	s002	V	Gino Mcneill	11/05/2002	14	151	31	street2
S6	s004	VI	David Parkes	15/09/1997	12	159	32	street4

14. Write a Pandas program to detect missing values of a given DataFrame. Display True or False.

Test Data:

	ord_no	purch_amt	ord_date	customer_id	salesman_id
0	70001.0	150.50	2012-10-05	3002	5002.0
1	NaN	270.65	2012-09-10	3001	5003.0
2	70002.0	65.26	NaN	3001	5001.0
3	70004.0	110.50	2012-08-17	3003	NaN
4	NaN	948.50	2012-09-10	3002	5002.0
5	70005.0	2400.60	2012-07-27	3001	5001.0
6	NaN	5760.00	2012-09-10	3001	5001.0
7	70010.0	1983.43	2012-10-10	3004	NaN
8	70003.0	2480.40	2012-10-10	3003	5003.0
9	70012.0	250.45	2012-06-27	3002	5002.0
10	NaN	75.29	2012-08-17	3001	5003.0
11	70013.0	3045.60	2012-04-25	3001	NaN

15. Write a Python program to draw a line using given axis values with suitable label in the x axis , y axis and a title. The code snippet gives the output shown in the following screenshot:



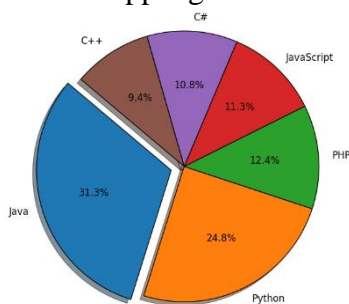
16. Write a Python programming to create a pie chart of the popularity of programming Languages.

Sample data:

Programming languages: Java, Python, PHP, JavaScript, C#, C++

Popularity: 22.2, 17.6, 8.8, 8, 7.7, 6.7

The code snippet gives the output shown in the following screenshot:



17. Write a Python program to load the iris data from a given csv file into a dataframe and print the shape of the data, type of the data and first 3 rows.