



CVVM
UNIVERSITY

Aegis: Charutar Vidya Mandal (Estd.1945)

FACULTY OF ENGINEERING & TECHNOLOGY

Effective from Academic Batch: 2022-23

Programme: Bachelor of Technology (Computer Engineering)

Semester: V

Course Code: 202045603

Course Title: Python for Data Science

Course Group: Professional Core Course

Course Objectives: The main objective of the course is to make students aware about the basic concepts of data science and Python programming. At the end of the course, students should be able to develop various application to draw useful insights for enhancing decision-making capabilities by performing various operations on data through implementation in Python and able to write short scripts to perform data pre-processing, exploratory data analysis and data visualization.

Teaching & Examination Scheme:

Contact hours per week			Course Credits	Examination Marks (Maximum / Passing)				
Lecture	Tutorial	Practical		Theory		J/V/P*		Total
				Internal	External	Internal	External	
3	0	2	4	50 / 18	50 / 17	25 / 9	25 / 9	150 / 53

* J: Jury; V: Viva; P: Practical

Detailed Syllabus:

Sr.	Contents	Hours
1	Getting started with Python: Basics of Python including data types, operators, variables, expressions, control structures, conditional statements, looping structure, functions, OOP concepts. Python sequence data structures including String, Array, List, Tuple, Set, and Dictionary and associated operations	05
2	Data Science and Descriptive Statistics: Understanding data, its type and data science, Considering the emergence and fundamentals of data science, Data science life cycle, Data analysis (Univariate, bivariate and Multivariate), Data analytics, Data measurement scale, Data descriptive statistics (Measures of central tendency, dispersion/variation, measure of location, Shape and symmetry), Understanding Python's role in data science.	05



CVVM
UNIVERSITY

Aegis: Charutar Vidya Mandal (Estd.1945)

3	Exploring Python libraries: Reading from a text file, CSV file, excel file, Streaming, Sampling and uploading data, Managing data with relational database (MySQL), Exploring statistics of data using stats package from SciPy, Numerical computing and working with multidimensional array through NumPy, Data manipulation through Pandas, Basic functionalities of machine learning using Scikit-learn, Extracting components from HTML document through beautiful soup.	08
4	Probabilistic and Inferential Statistics: Random variable, basics of probability, Probability mass function, Probability density function, Cumulative distribution function, Discrete probability distribution (Binomial and Poisson), Continuous probability distributions (Normal and Exponential). Sampling and its various techniques, Estimation, point estimation and interval estimation, Standard normal distribution, Central limit theorem, Hypothesis testing through one sample test.	10
5	Data preparation: Data loading, Data cleaning, dealing with missing data, removing duplicates, Slicing and Dicing, Filtering, and selecting data, Concatenating and transforming, adding new cases and variables, removing data, Sorting and shuffling, Aggregating data, Handling outliers, Data wrangling, Data Normalization.	08
6	Data Visualization: Visualizing data through figure, subplot and its properties, graphs, plots in matplotlib, basics of advanced visualizations with Seaborn.	04
	Total	40

List of Practicals / Tutorials:

1	Understanding Python environment Setup: Installing python with anaconda, Introduction to various Python IDEs like IDLE, Jupyter notebook, Pycharm , spyder.
----------	---



CVVM
UNIVERSITY

Aegis: Charutar Vidya Mandal (Estd.1945)

<p>2</p>	<p>Create a list named “Subjects” by inserting 10 subjects into it through any loop and create a list “Elective Subjects” with 5 subjects through direct initialization. Extend list “Subject” by another list “Elective Subjects”. Append 3 duplicate subjects into “Subject” list. Find the index of first occurrence of that duplicate value and then remove all the occurrences of that specific subject through loop. Define function remove range(i1,i2) to remove range of element from i1 to i2 through del keyword and return the resultant list. Pop 5th element after reversing and sorting your list. Count total elements in your list and finally clear the list. Which of the above operations can be performed directly? Which of the above operations cannot be performed directly on Tuple and why? Update and remove specific item from the tuple by converting it into list.</p> <p>Create a dictionary named “Students Data” with 5 students and id_no, name and marks as the key values. Provide the separate list of all the keys and values. Add details of one more student. Retrieve value corresponding to specific key through get method. Define a function update detail(k) by looping over keys to search for specific key ‘k’ whose details to be updated and then update it with new details and return updated dictionary. If specific detail is not available in list print appropriate message. Convert dictionary’s keys into a list by looping through keys and appending it to the other list. Convert dictionary values into list through list constructor. Count and display total number of students in the dictionary. Remove all the details from the dictionary. Define a dictionary named “exam_data_array” with 4 keys, namely ‘name’, ‘score’, ‘attempts’ and ‘qualify’. Values for each of these 4 keys will be an 1Darray with 5 elements. by creating a dictionary named “exam_data_list” with 5 list and each list stores all 4 key-value pairs for single student.</p>
<p>3</p>	<p>Do the slicing of a given String to generate various substring by passing different index (like positive index, negative index, end index > string length, entire string), split this string into chunks of length 3 using list comprehension, split the string with specific character, iterate over the words of string. Apply trim, toupper, tolower, replace string and character, title, join and other operations on String.</p> <p>Perform add, union, intersection, difference, symmetric_difference, union, intersection_update, symmetric_difference_update, difference_update, discard, issubset, issuperset, isdisjoint, remove, pop and clear operations on Set.</p> <p>Import array module in python and perform all operations available in the module</p>
<p>4</p>	<p>Download “heart_2020_cleaned.csv” dataset from “https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease” and perform all the descriptive statistics on above dataset using statistics module of python and scipy.stats package (Measures of central tendency, measure of dispersion/variation, measure of location, measure of shape and symmetry).</p>



CVVM
UNIVERSITY

Aegis: Charutar Vidya Mandal (Estd.1945)

<p>5</p>	<p>Write following program on Pandas DataFrame:</p> <ul style="list-style-type: none">● Create an array “rank” with 5 element(rank1,rank2,...,rank5). Create and display a DataFrame “exam” from a specified dictionary “exam_data_array” with “rank” as label. <p>Also display a summary of basic information and its data.</p> <p>Perform following operations on DataFrame “exam” :</p> <ul style="list-style-type: none">● Select the rows where the score is between 15 and 20 (inclusive).● Sort the data first by “score” in ascending order , then by “name” in descending order.● Replace the 'yes' and 'no' values from column “qualify” with True and False.● Display specified columns(columns: 2 and 4) and rows(row: 1,3 and 5).● Select the rows where number of attempts in the examination is less than 2 and score greater than 15.● Change the name 'James' to 'Suresh' in “name” column of the data frame.● Calculate the sum of the examination attempts by the students● Append one row.● Insert a new column “exam_name” and then Delete the ”exam_name” column.● Convert a NumPy array, dictionary and first column of a DataFrame to a series. <p>Write following program on NumPy Array:</p> <ul style="list-style-type: none">● Create an array of all the even integers from 30 to 70.● Create an array of 10 zeros, other with 10 ones, and one more with10 fives.● Create a vector of length 10 with values evenly distributed between 5 and 50.● Create a 3x4 matrix filled with values from 10 to 21 and compute sum of all elements, sum of each column and sum of each row of a given array.● Create a 3x4 array and find the missing data in the array.● Calculate round, floor, ceiling, truncated and round (to the given number of decimals) of the input, element-wise of an array.● Find the maximum and minimum value, median , Weighted average, mean, standard deviation, variance, covariance matrix, of a given flattened array, minimum and maximum value along the second axis. <p>Create a structured array from given student name, height, class and their data types. Now sort by class, then height if class are equal.</p>
<p>6</p>	<p>Write following python programs on Beautiful Soup</p> <p>Perform following operations on a HTML document.</p> <ul style="list-style-type: none">● Find the title tag from a given html document.● Count and retrieve all the paragraph tags and extract the text in the first paragraph tag.● Find the text of the first <a> tag and length of the text of the first <h2> tag.● Find the href of the first <a> tag.● Find the first tag with a given attribute value in an html document. <p>Write a python program for MySQL Database connectivity(import sqlite3 module)</p> <p>Establish the connection with Education database named “Education” in SQLite , create a table named Student(with id_no, name, department, gander, total_marks) in Education database. Perform insert, update,select and delete operation on Student table.</p>



CVVM
UNIVERSITY

Aegis: Charutar Vidya Mandal (Estd.1945)

7	Write a python program to download appropriate dataset and explore random variable, Probability mass function, Probability density function, Cumulative distribution function, Discrete probability distribution and continuous probability distribution using scipy.stats , rv_discrete class and rv_continuous class .
8	Write a python program to compute and explore normal distribution, central limit theorem, point estimate, interval estimation and hypothesis testing.
9	<ul style="list-style-type: none">Identify the column(s) of a given DataFrame which have at least one missing value, count the number of missing values in each column and drop the rows and columns with missing values. Check for the null values. Also remove the duplicate values from the DataFrame. Handle outliers in the Data Frame.Access subset of data through indexing(Select data using labels(column headings)), Slicing(Extract range based subset, subset of rows, subset of columns, select a subset of rows and columns from our DataFrame using iloc method). Perform other data processing on a given dataset.
10	Write a python program to perform data visualization through Matplotlib.
11	Write python program for advanced data visualization through Seaborn.
12	Exploring Google Data Analytics tool.

Reference Books:

1	Python for data science for dummies, John Paul Mueller, Luca Massaron, and Wiley
2	Programming through Python, M. T. Savaliya, R. K. Maurya, G. M. Magar, STAREDU Solutions
3	Pandas for everyone: Python Data Analysis, Daniel Y. Chen, Pearson
4	Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools, Davy Cielen, Arno D.B. Meysman, et al., Minning
5	Applied Data Science with Python and Jupyter, Alex Galea,
6	Business Analytics: The Science of Data - Driven Decision Making, U Dinesh Kumar, Wiley India.

Supplementary learning Material:

1	https://www.udemy.com/course/python-for-data-science-and-machine-learning-bootcamp/
2	https://www.coursera.org/specializations/data-science-python
3	https://www.w3resource.com/python-exercises/
4	Python Tutorials – Real Python

Pedagogy:

- Direct classroom teaching
- Audio Visual presentations/demonstrations
- Assignments/Quiz
- Continuous assessment
- Interactive methods
- Seminar/Poster Presentation
- Industrial/ Field visits
- Course Projects



CVVM
UNIVERSITY

Aegis: Charutar Vidya Mandal (Estd.1945)

Suggested Specification table with Marks (Theory) (Revised Bloom's Taxonomy):

Distribution of Theory Marks in %						R: Remembering; U: Understanding; A: Applying; N: Analyzing; E: Evaluating; C: Creating
R	U	A	N	E	C	
15%	20%	30%	15%	10%	10%	

Note: This specification table shall be treated as a general guideline for students and teachers. The actual distribution of marks in the question paper may vary slightly from above table.

Course Outcomes (CO):

Sr.	Course Outcome Statements	%weightage
CO-1	To identify the need for data science and solve basic problems using python basic concepts.	20
CO-2	To learn the fundamentals of some of the widely used python packages and apply them into data analytics and visualization.	50
CO-3	To prepare the data for design applications through various data preprocessing operations	15
CO-4	To understand the concept of probabilistic and inferential concepts in data science.	15

Curriculum Revision:

Version:	2.0
Drafted on (Month-Year):	June-2022
Last Reviewed on (Month-Year):	-
Next Review on (Month-Year):	June-2025