Marwadi University

FACULTY OF COMPUTER APPLICATIONS

Master of Computer Applications

- **Sem.** : 2
- **Subject Code** : 05MC0207
- **Subject** : Data Analytics and Visualization
- **Course Objectives**:

1. To develop proficiency in data science using python.
2. To understand how different types of object can be used in python.
3. To understand the Data cleaning & importing data from different file format.
4. To explore knowledge of Data frame in python data-analysis.
5. To learn different techniques of data visualization in order to provide new insights in data analysis.

- **Prerequisites**: Basic Knowledge of Programming Languages

Marwadi University

FACULTY OF COMPUTER APPLICATIONS

Master of Computer Applications

Unit No	Topics Covered	No of lectures required
1	**Introduction to Data Science with Python :** The Stages of Data Science, Basic features of Python, Basics of Python Programming, Fundamental Python Programming techniques, Data cleaning and manipulation techniques, Abstraction of the series and Data Frame, Running basic inferential Analyses	10
2	**Data Collection Structures :** **Lists :** Creating Lists, Accessing values in Lists, Adding and updating Lists, Deleting List Elements, Basic List operations, Indexing-Slicing and matrices, Built-in List Functions and Methods, List sorting and traversing, List and strings, Parsing Lines and aliasing **Dictionaries :** Creating Dictionaries, Updating and accessing values in Dictionaries, Deleting Dictionary elements, Built-in Dictionary Functions, Built-in Dictionary Methods **Tuples:** Creating Tuples, Concatenating Tuples, Accessing values in Tuples, Basic Tuple Operations **Series:** Creating a Series with index, Creating a Series from a Dictionary, Creating a Series from a	15

	Scalar value, Vectorized operations and Label Alignment with Series **Data Frames:** Creating Data Frames from a Dict of Series or Dicts, Creating Data Frames from a Dict of Ndarrays/Lists, Creating Data Frames from a structured or Record Array, Creating Data Frames from a List of Dicts, Creating Data Frames from a Dict of Tuples, Selecting- Adding-and Deleting Data Frame Columns, Assigning new columns in Method Chains, Indexing and Selecting Data Frames, Transposing a Data Frame, Data Frame Interoperability with Numpy Function	
3	**Data Gathering and Cleaning :** Cleaning Data – Checking of Missing value, Handling the Missing values, Reading and cleaning CSV Data, Merging and integrating Data	**08**
4	**Data Analysis :** Statistical Analysis, Data Grouping, Iterating through Groups, Aggregation, Transformations, Filtration	**07**
5	**Data Visualization :** **Direct Plotting** – Line Plot, Bar Plot, Pie Chart, Box	**10**

	Plot, Histogram Plot, Scatter Plot **Seaborn Plotting System** - Strip Plot, Box Plot, Swarm Plot, Joint Plot **Matplotlib Plot –** Line Plot, Bar Chart, Histogram Plot, Scatter Plot, Stack Plot, Pie Chart	

Course Outcomes:

1. Understanding of basic concepts of python programming.
2. Interpretation of various type of object of python.
3. Handling of Data import & data cleaning before analysis.
4. Understanding of Various data Exploring & data frame usage in data analysis.
5. Visualization of various graphs to get insight of the data.

Course Outcomes – Program Outcomes Mapping Table:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11
CO1		H			L						M
CO2	H				M					L	
CO3		L	M							M	
CO4		H	L						M		
CO5	M	L									H

Marwadi University

FACULTY OF COMPUTER APPLICATIONS

Master of Computer Applications

Text Book:

1. **"Data Analysis and visualization using Python", by Dr. OssamaEmbarak**

Reference Books:

1. **"Introduction to Computation and Programming Using Python", John V Guttag. ,2nd Edition, Prentice Hall of India**
2. **Core Python Programming, R Nageswara Rao, 2nd Edition, Dreamtech Press**
3. **Core Python Applications Programming, Wesley J Chun, 3rd Edition. Pearson**

Web References:

1. **"Python Programming", http://en.wikibooks.org/wiki/Python_Programming**
2. **"The Python Tutorial", http://docs.python.org/release/3.0.1/tutorial/**

App References:

1. **https://play.google.com/store/apps/details?id=com.androfrenzy.datascience&hl=en_IN&gl=US**
2. **https://play.google.com/store/apps/details?id=com.admob9931.python_panda&hl=en_IN&gl=US**

Marwadi University

FACULTY OF COMPUTER APPLICATIONS

Master of Computer Applications

Syllabus Coverage from text /reference book & web/app reference:

Unit #	Chapter Numbers
1	Chapter-1,2 from Text Book
2	Chapter-3 from Text Book
3	Chapter-5 from Text Book
4	Chapter-6 from Text Book
5	Chapter-7 from Text Book

Marwadi University

FACULTY OF COMPUTER APPLICATIONS

Master of Computer Applications

PRACTICALS

Unit No	List of Practicals
1	1. Write a Python script to prompt users to enter two values; then perform the basic arithmetical operations of addition, subtraction, multiplication and division on the values. 2. Write a Python script to prompt users to enter the first and last values and generate some random values between the two entered values. 3. Write a Python program to prompt users to enter a distance in kilometers; then convert kilometers to miles, where 1 kilometer is equal to 0.62137 miles. Display the result. 4. Write a Python program to prompt users to enter a Celsius value; then convert Celsius to Fahrenheit, where T(°F) = T(°C) x 1.8 + 32. Display the result. 5. Write a program to prompt users to enter their working hours and rate per hour to calculate gross pay. The program should give the employee 1.5 times the hours worked above 30 hours. If Enter Hours is 50 and Enter Rate is 10, then the calculated payment is Pay: 550.0.

	6. Write a program to prompt users to enter a value;then check whether the entered value is positive or negative value and display a proper message. 7. Write a program to prompt users to enter a value; then check whether the entered value is odd or even and display a proper message. 8. Write a program to prompt users to enter an age; then check whether each person is a child, a teenager, an adult, or a senior. Display a proper message. 9. Write a program to prompt users to enter a year; then find whether it's a leap year. A year is considered a leap year if it's divisible by 4 and 100 and 400. If it's divisible by 4 and 100 but not by 400, it's not a leap year. Display a proper message. 10. Write a program to prompt users to enter a Fibonacci sequence.
2	1. Write a program to create a list of names; then define a function to display all the elements in the received list. Call the function to execute its statements and display all names in the list. 2. Write a program to read text file data and create a dictionary of all keywords in the text file. The program should count how many times each word is repeated inside the text file and then find the keyword with a highest repeated number.The program should display both the keywords dictionary and the most repeated word.

	3. Write a program to compare tuples of integers and tuples of strings. 4. Write a program to create a series to maintain three students' names and SPI values. 5. Write a program to create a data frame to maintain three students' names associated with their grades in three courses and then add a new column named Mean to maintain the calculated mean mark per course. Display the final data frame. 6.Write a program in python to swap two variables without using temporary variable 7.Write a program which will allow user to enter 10 numbers and display largest odd number from them. It will display appropriate message in case if no odd number is found. 8. Write a Python program to perform following operation on given string input: a) Count Number of Vowel in given string b) Count Length of string (do not use Len ()) c) Reverse string d) Find and replace operation e) check whether string entered is a palindrome or not 9. Write a program in python to find factorial of user entered number.

Marwadi University

FACULTY OF COMPUTER APPLICATIONS

Master of Computer Applications

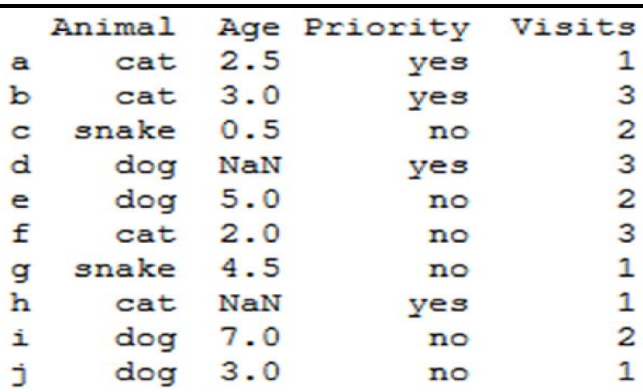
	10. Write a program in python to find factorial of user entered number. (Use recursive Function)
3	Write a Python script to read the data in an Excel file named movies.xlsx and save this data in a data frame called mov. Perform the following steps: 1. Read the contents of the second sheet that is named 2000s in the Excel file (movies.xlsx)and store this content in a data frame called Second_sheet. 2. Write the code needed to show the first seven rows from the data frame Second_sheet usingan appropriate method. 3. Write the code needed to show the last five rows using an appropriate method. 4. Use a suitable command to show only one column that is named Budget. 5. Use a suitable command to show the total rows in the first sheet that is called 2000s.

	6. Use a suitable command to show the maximum value stored in the Budget column. 7. Use a suitable command to show the minimum value stored in the Budget column. 8. Write a single command to show the details (count, min, max, mean, std, 25%, 50%, 75%) about the column User Votes. 9. Use a suitable conditional statement that stores the rows in which the country name isUSA and the Duration value is less than 50 in a data frame named USA50. Show the values indata frame USA50. 10. Using a suitable command, create a calculated column named Avg Reviews in Second_sheetby adding Reviews by Users and Reviews by Critics and divide it by 2. Display the first fiverows of the Second_sheet after creating the previous calculated column. 11. Using a suitable command, sort the Country values in ascending order (smallest to largest)and Avg_reviews in descending order (largest to smallest).
4	1. Create a data frame called df from the following tabular data dictionary that has these index labels: ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h','i', 'j'].

Marwadi University

FACULTY OF COMPUTER APPLICATIONS

Master of Computer Applications

	 2. Display a summary of the data frame's basic information. 3. Return the first three rows of the data frame df. 4. Select just the animal and age columns from the data frame df. 5. Count the visit priority per animal. 6. Find the mean of the animals' ages. 7. Display a summary of the data set. 8. Return the first three rows of the data frame df. 9. Extract first and last column in one table. 10. Observe output of ndim(), shape()
5	1. Create 500 random temperature readings for sixcities over a season and then plot the generated datausing Matplotlib.

Marwadi University

FACULTY OF COMPUTER APPLICATIONS

Master of Computer Applications

	2. Load the well-known Iris data set, which lists measurements of petals and sepals of three iris species. Then plot the correlations between each pair using the .pairplot() method. 3. Load the well-known Tips data set, which shows the number of tips received by restaurant staff based on various indicator data; then plot the percentage of tips per bill according to staff gender. 4. Load the well-known Tips data set, which shows the number of tips received by restaurant staff based on various indicator data; then implement the factor plots to visualize the total bill per day according to staff gender. 5. Reimplement the above exercise using the Seaborn joint plot distributions. 6. Python program of Barplot with all parameters of a sample data. 7. Python program of Pie-chart with all parameters of a sample data. 8. Python program of Histogram with all parameters of a sample data. 9. Python program of Line Plot with all parameters of a sample data. 10. Python program of Scatter plot with all parameters of a sample data.