Course Code	Course Name	Credits
20MCADC304	Data Analytics and Visualization	03

❖ Aim of the Course:

1	The main goal of this course is to help students learn, understand, practice and to introduce various data analytics techniques and presenting result graphically by means of various types of appropriate chart.

❖ Course Overview and Context:

1	The course provides various hands-on techniques and examples to pursue career in data analytics and visualization area. Learners will be able to derive meaningful insight by learning fundamentals of python data structure, numpy and pandas library, data cleaning and organizing operations and plotting them visually using various library functions of matplotlib library.

❖ Course Outcomes:

Sr #	Course Outcome	Cognitive Level
1	To **recognize** need of data analytics and visualization, and able to **distinguish** basic data structure of python and	Remember, Understand
2	To **solve** real-life examples using numpy and pandas library.	Understand, Apply
3	To **determine** and **select** appropriate data cleaning and organizing strategy.	Understand, Apply, Analyze
4	To **recognize** and **distinguish** importance of each chart, to **judge** real-life situation and present result as appropriate chart.	Remember, Understand, Apply, Analyze, Evaluate
5	To **outline** various subplots in one plot.	Analyze

❖ Content of the Course:

Unit-1 Introduction to Data Analysis and basic python data structure

- **Introduction to Data Analysis**: Data Analysis, Knowledge domain of data analysis, nature and types of data, data analysis process.
- **List**: Creating list, accessing values in list, adding, updating and deleting values from list, indexing and slicing from list, basic list operations, list sorting and traversing, aliasing, parsing lines.
- **Tuple**: Creating tuple, concatenating tuples, accessing values in tuple, basic tuple operations.
- **Dictionary**: Creating dictionary, updating and accessing values in dictionary, deleting dictionary elements, built-in dictionary methods and functions.

Unit-2 Data gathering, NumPy and Pandas Library

- **Numpy library:** basic operations, indexing, slicing, iterating ndarray, shape manipulation, array manipulation, structured arrays.
- **Pandas library:** Pandas data structure – series, dataframe and index objects, reindexing, dropping, arithmetic and data alignment, operations between data structure, function application and mapping, sorting and ranking, correlation and covariance, not a number data, hierarchical indexing and levelling.
- **Reading and Writing data:** Reading and writing data in CSV/ text file, Reading and Writing HTML files, Reading and writing data from excel files, Reading data from XML, JSON data, Pickle -Python

object serialization, Interacting with database.

Unit-3 Cleaning and Organizing Data

- **Cleaning data:** Data preparation – merging, Concatenating – combining, pivoting, removing, Discretization and Binning – detecting and filtering/removing outliers, data transformation - missing data, filtering inappropriate values, finding duplicate rows, removing punctuation and whitespace from column content, standardizing dates, data aggregation – group by, hierarchical grouping, Chain of transformations, Function on groups, advanced data aggregation, Permutation – random sampling.
- **Organizing data:** removing and adding columns, selecting columns, change column name, finding matching rows, filter rows based on condition, selecting rows-based condition.

Unit-4 Basic Data Visualization with Matplotlib - 1

- **Introduction to Matplotlib :** installation, matplotlib architecture, pyplot, plotting window, using kwargs, adding text, grid and legend to chart, saving code, saving chart as image,
- **Bar chart –** horizontal bar chart, multi serial bar chart, multi series stacked bar chart.
- **Pie chart –** piechart with pandas dataframe

Unit-5 Advanced Data Visualization with Matplotlib - 1

- **Line Chart, Histogram, contour plots, polar charts**
- **3-D Chart :** mplot3d toolkit
- **Multi-panel plots** – display subplot within other plot, grid of subplots.

❖ Learning Resources:

Sr #	Textbook | References | Internet Links
1	"Data Analysis and Visualization using Python" by Dr. Ossama Embarak – APress – 2018.
2	"Python Data Analytics – with Pandas, numpy and matplotlib" by Fabio Nelli – APress – 2nd Edition - 2018.
3	"Learn Data Analysis With Python", A.J.Henly, Dave Wolf, Apress, 2018.
4	"Python for Data Analysis, Data Wrangling with Pandas, NumPy and IPython", Wes McKinney, 2nd Edition, O'Reilly

❖ Assignments (Optional):

Sr #	Description	Available From (Date)	Submission Date
1	Numpy, Pandas library, Data Cleaning operations	After 3 Weeks	Within 10 Days
2	Data organizing operations, Data Visualization	After 6 Weeks	Within 7 Days