GUJARAT TECHNOLOGICAL UNIVERSITY

Master of Engineering Syllabus

Subject Code : 3720319

Subject Name : Data Science for Engineering

WEF Academic Year :	2018-19
Semester :	2
Category of the Course :	Program Elective IV

Prerequisite :	Basic knowledge of Python programming.
Rationale :	The main objective of the course is to make students learn about the basic concepts of Python in data manipulation and analysis. Students will be able to write and execute Python code. At the end of the course, students should be able to write short scripts to perform data preprocessing, exploratory data analysis and data visualization.

Course Scheme :

Teaching Scheme			Total Credits	Assessment Pattern and Marks				Total Marks
				Theory		Practical		
L	T	PR	C	ESE (E)	PA(M)	ESE (V)	PA (I)	
3	0	2	4	70	30	30	20	150

Course Content :

Sr. No.	Course Content	No. of Hours	% of Weightage
1	**Introduction of Data Science :** What is data, What is data science, Fundamentals of data science, Data science life cycle, Why data science is important, Applications of data science, Why Python is necessary for data science	4	10
2	**Getting Started with Python :** Jupyter/pycharm/spyder or any other python tool set up and installation. Basics of Python including data types, operators, variables, expressions, control structures using sample dataset, objects and functions. Python sequence data structures including String, Array, List, Tuple, Set, and Dictionary. Introduction to various python libraries for data science	10	22
3	**Basic data Processing :** Introduction to Data Frames in Pandas, Learning to access elements with indexes, Re-indexing in pandas Series and Data frames, Data preparation. Numerical Computing with NumPy,	10	22

	Going from Python list to Numpyarrays, working with multi-dimensional array, array operations. Several scientific numerical routines through SciPy.		
4	**Exploratory data analysis :** Data preprocessing (data loading, dealing with missing values and outliers, data wrangling, filtering data, Data Normalization , Data Formatting ,data cleaning), Web scraping with beautiful soup.	10	24
5	**Data Visualization :** Basic visualizations with Matplotlib, Advanced visualizations with Seaborn, Plotting images, graphs and grids of charts.	8	22
	Total	**42**	**100**

Reference Book :

1. Python for data science for dummies 2nd Edition, John Paul Mueller, Luca Massaron, and Wiley.
2. Pandas for everyone: Python Data Analysis, Daniel Y. Chen, Pearson.
3. Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools, Davy Cielen, Arno D.B. Meysman, et al., Mining.
4. Applied Data Science with Python and Jupyter: Use powerful industry-standard tools. to unlock new, actionable insights from your data.

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

Distribution of Theory Marks					
R Level	U Level	A Level	N Level	E Level	C Level
12 %	20 %	20 %	20 %	16 %	12 %

Legends: R: Remembrance; U: Understanding; A: Application, N: Analyze and E: EvaluateC: Create

Course Outcome :
After Completion of the Course, Student will able to :

Sr. No	Course Outcomes	RBT Level*
01	Apply various Python data structures to effectively manage various types of data.	AP
02	Learn the fundamentals of some of the widely used python packages and apply them into data analytics.	UN AP
03	Design applications applying various operations for data cleansing and transformation.	AN EL CR
04	Describe the various areas where data science is applied.	RM UN

*RM: Remember, UN: Understand, AP: Apply, AN: Analyze, EL: Evaluate, CR: Create

Suggested Course Practical List :

1. Write a program to create a list, insert elements into the list and sort it in ascending order.
2. Write a program to create a dictionary of 10 elements, change/delete the values off ewkeys anddisplay the dictionary before and after the updates.
3. Write a program to create a tuple and a list. Convert the list totuple and display the elementsof both. Write the program to remove the duplicate element of the list.
4. Write a program to perform all basic data pre processing steps on the given data set.
5. Write a program to perform exploratory data analysis on the given dataset.
6. Develop programs to learn the concept of Modules and packages.
7. Develop a program to learn concept of array and numpy module.
8. Write a NumPy program to convert a list of numeric value into a one-dimensional NumPy array. And perform all operations on that array.
9. Write a NumPy program to find the union of two arrays. Union will return the unique, sortedarray of values that are in either of the two input arrays.
10. Write a Pandas program to convert a NumPy array to a Pandas series. Also write a Pandas program to calculate the frequency counts of each unique value of a given series.
11. Write a Pandas program to read a dataset from diamonds Data Frame and modify the default columns values and print the first 6 rows. Also find the number of rows and columns and data typeof each column of diamonds Dataframe.

12. Consider dataset with student name, gender, Enrollment no, 4 semester result with marks of eachsubject, his mobile number, and city. Implement following in Python. Plot various graphs and chart to visualize students‘ data.

List of Laboratory/Learning Resources Required :

- www.anaconda.com
- www.python.org
- www.w3schools.com
- https://www.learnpython.org/

****