MIT Academy of Engineering (An autonomous Institute Affilated to SPPU)	COURSE SYLLABUS	
SCHOOL OF COMPUTER ENGINEERING	W.E.F	2022 - 2023 (Rev.2022)
FIRST YEAR BACHELOR OF TECHNOLOGY	COURSE NAME	ESSENTIALS OF DATA SCIENCE
	COURSE CODE	CS 104
	COURSE CREDITS	4
RELEASED DATE : 01/11/2022	REVISION NO	0.0

TEACHING SCHEME (HOURS/WEEK)		EXAMINATION SCHEME AND MARKS					
		THEORY			PRACTICAL	CONTINUOUS	TOTAL
LECTURE	PRACTICAL	IA	MSE	ESE	AND PROJECT	ASSESSMENT	
2	4	15	20	40	50	25	150

PRE-REQUISITE :

COURSE OBJECTIVES :

CS104.CEO.1: To get familiar with the basics of Python programming.
CS104.CEO.2: To learn different data structures in Data Science.
CS104.CEO.3: To use data computation methods in Data Science.
CS104.CEO.4: To introduce data manipulation methods in Data Science.
CS104.CEO.5: To use data visualization methods in Data Science.
CS104.CEO.6: To understand the importance and basic techniques of data science.

COURSE OUTCOMES :

The students after completion of the course will be able to,

CS104.CO.1: Build a python program for handling syntax and semantics.[Level 3]
CS104.CO.2: Demonstrate proficiency in handling data structures useful in Data Science.[Level 3]
CS104.CO.3: Apply the different methods of data computations on real time data.[Level 3]
CS104.CO.4: Interpret the different methods of data manipulation on real time data.[Level 3]
CS104.CO.5: Apply data visualization for real time data.[Level 3]
CS104.CO.6: Apply basic techniques of data science.[Level 3]

COURSE ABSTRACT

Essentials of Data Science (EDS) is the First Year Semester-II course. This course aims at equipping learners to be able to provide the essence of Data Science by using python programming. In this course, firstly, the learners will learn python fundamentals required for Data Science. The learners are able to use different data organizational structures to store the real time data. Data computations are required to maintain the quality of data. So, EDS helps to learn different data computations. Data manipulations help in handling missing and noisy values in Data Processing. The learner will learn data storing, loading, cleaning, preparation, wrangling, transformation etc. Data visualization is the practice of translating information into a visual context, such as a map or graph which tends to identify the patterns, trends from the data. EDS helps the learners to understand the prediction of the likelihood of a future outcome by using some basic machine learning algorithm. This course will increase the learner's interest in Data Science. The activity and project based learning is adapted to learn this course.

THEORY COURSE CONTENT

UNIT 1	**Python Fundamentals for Data Science**	**5 HOURS**

App/System/Case study: Weather Analysis

Contents:

Introduction, Data Types- Mutable and immutable types, Data Conditioning using Decision Statements and Iterative Statements.

Functions: Basics of functions, variable scope and lifetime, Lambda or anonymous function.

Modules: Introduction to modules, packages and standard library modules.

Data Handling using Files.

UNIT 2	**Data Organizational Structures**	**5 HOURS**

App/System/Case study: Retail-Industry/Twitter Reviews

Contents:

Categories of Data- Unstructured, Structured, Categorical and Time Series

List- Creating List, List operations

Tuple- List of Tuples, Immutability,

Dictionaries- Creating Dictionary, Adding to Dictionary with SetDefault, Loading JSON to Dictionary

Set- Removing Data from Sequences, Performing common data set operation

UNIT 3	**Data Computation**	**5 HOURS**

App/System/Case study: Stock Market

Contents:

Data Operations- Arithmetic and Statistical, Bitwise Operators, Linear Algebra, Copying and viewing data in arrays, Stacking, Data Sorting, Data Searching and Indexing, Data counting, Mathematical Operations, Broadcasting, Matrix Operations, Structured Data.

UNIT 4	Data Manipulation	5 HOURS

App/System/Case study: Salary Analysis

Contents:

Data Loading, Data Storage, Summarizing and Computing Descriptive Statistics, Data Cleaning, Data Preparation, And Data Wrangling: Join, Combine, and Reshape, Data Transformation, Data Aggregation and Group Operations.

UNIT 5	Data Distribution and Visualization	5 HOURS

App/System/Case study: Sales

Contents:

Population and Samples

Data Distribution- Statistical Analysis of Data, Normal Distribution Significance of Data Visualization in Data Science, Plots- Line Chart, Bar Plot, Histogram, Scatter Plot, Pie Chart, Density Plot, Facet Grids for categorical data, Group Plots

UNIT 6	Problem Analysis in Data Science	5 HOURS

App/System/Case study: Weather Forecasting

Contents::

Data Analytics Techniques; Descriptive, Predictive and Prescriptive Learning Methods: Supervised, Semi Supervised, Unsupervised and Reinforcement Predictive Techniques- Simple Linear Regression, K-NN Classification, K-Means Clustering

PRACTICAL:Perform following experiments using python programming language.

PRACTICAL NO.01		10 HOURS

Practice Lab Assignments:

1. Find statistical analysis of Employee Records
2. If two datasets i.e. employee dataset and salary dataset are available. Then merge both dataset. Sort the records in ascending order of salary. Perform Data analysis on the merge dataset, Write a function to find the first 5 employees who have the highest salary from the records.
3. Read the birth date of employees from the Employee Record. Perform data transformation for birthdate to age and also salary which is in rupees to salary in dollars.

Lab Assignment:

Take/Prepare any text files for any real life application. For Ex. “Stud.txt”, “Placement.csv” and “Result.csv” files for result Analysis. Combine into “StudentDetails.csv”. Perform all statistical analysis (Average, Max, Min, Count, Sum, Percentage) on it

Self Study Assignment: Perform all statistical analysis (Average, Max, Min, Count, Sum, and Percentage) on F.Y.BTech students’ results where data will be in a separate data sheet.

PRACTICAL NO.02		10 HOURS

Practice Lab Assignment:

1. Perform all List Operations, Tuple Operations, Dictionary Operations.

Lab Assignment:

Prepare/Take dataset for any real life application. For Ex. Sales of the company. Read the data from Sales.csv/.xls/.txt. Store Product details in List data structure. Store Supplier Details in Dictionary Data Structure. Store Customer Details in Tuple Data Structure. Now perform the following operations:

- Find the most popular product for sales.
- Find the best supplier for sales.
- Find the customer who buys most of the products.
- Find the number of customer who are 'Female'

Self Study Assignment: Consider the student result dataset. Find 10 grains for the student result. For 10 grains find the solution using list, tuple and dictionary.

PRACTICAL NO.03		10 HOURS

Practice Lab Assignment:

1. Perform all the Numpy operations in python.

Lab Assignment:

Prepare/Take dataset for any real life application. Read a dataset into an array. Perform following operations on it as:

- Perform all matrix operations
- Horizontal and vertical stacking of Numpy Arrays
- Custom sequence generation
- Arithmetic and Statistical Operations, Mathematical Operations, Bitwise Operators
- Copying and viewing arrays
- Data Stacking, Searching, Sorting, Counting, Broadcasting

Self Study Assignment: For any real life application, perform advanced data operations such as image as array and image manipulations.

PRACTICAL NO.04		10 HOURS

Practice Lab Assignment:

1. Perform all the pandas operations in python.

Lab Assignment:

Read any real life dataset. Store the data into Data Frames. Identify 10 grains for the given dataset. Implement all 20 grains using Pandas methods. The Sample Grains for Sales Dataset as:

- Which was the best month for sales? How much was earned that month?
- Which product sold the most? Why do you think it did?
- Which city sold the most products?
- What Products are most often sold together?

Self Study Assignment: For any real life application, perform advanced Data Manipulation operations such as combining data, handling time series data.

PRACTICAL NO.05		10 HOURS

Practice Lab Assignment:

1. Install MatPlotLib/Seaborn library. Draw basic graphs for sales dataset using MatPlotLib/Seaborn.

Lab Assignment:

Select any one real-life dataset. Perform data analysis. Identify 10 grains for a given dataset. Develop an interactive dashboard using the MatPlotLib/Seaborn library. (Use any 10 different graphs with proper title, legends, axis names, etc.to map identified grains)

Self Study Assignment: For any real life application, perform advanced graphs for data visualization such as Span Selector, Broken Barh-Broken Horizontal Bar plot, Watermarking Images with Matplotlib.

PRACTICAL NO.06		10 HOURS

Practice Lab Assignment:

1. Install sklearn library. Perform Simple Linear Regression, K-NN Classification, and K-Means Clustering in python.

Minor Project:

Select a real life dataset. Perform 5 data computation, manipulation and data visualization operations. Implement a mini project based on classification (Linear Regression / KNN Classification) or Clustering (K-Means) on a selected dataset.Evaluate the performance of classification and clustering.

Guidelines:

- Project is for a period of 2 weeks.
- Group of two or three has to choose a project topic from the list designed by concerned faculty of a particular division.
- Each group has to collect requirements for the project and get approved by concerned teachers in the first week.
- Implementation and testing need to be performed in the second week.
- Demonstration along with presentation need to be given as final project submission.
- Project carries 20 Marks

TEXT BOOK

1. VanderPlas, J. (2016). Python data science handbook: Essential tools for working with data. " O'Reilly Media, Inc.".
2. McKinney, W. (2012). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. " O'Reilly Media, Inc.".
3. McKinney, W. (2017). Python for data analysis, " O'Reilly Media, Inc".
4. Brownley, C. W. (2016). Foundations for Analytics with Python: From Non-Programmer to Hacker. " O'Reilly Media, Inc.".

REFERENCE BOOK

1. Mueller, J. P., & Massaron, L. (2019). Python for data science for dummies. John Wiley & Sons.
2. Grus, J. (2019). Data science from scratch: first principles with python. O'Reilly Media.
3. Kane, F. (2017). Hands-on data science and python machine learning. Packt Publishing Ltd.
4. Madhavan, S. (2015). Mastering python for data science. Packt Publishing Ltd.