SEMESTER	**VI**					
YEAR	**III**					
COURSE CODE						
TITLE OF THE COURSE	**DATA SCIENCE LAB**					
SCHEME OF INSTRUCTION	Lecture Hours	Tutorial Hours	Practical Hours	Seminar/Projects Hours	Total Hours	Credits
	-	**-**	**2**	**-**	**26**	**01**

Prerequisite Courses (if any)			
#	**Sem/Year**	**Course Code**	**Title of the Course**
***	4th /2nd	20DS2401	DATA SCIENCE-1

COURSE OBJECTIVES:

- To use the statistical and computational techniques to Discover, Analyze, Visualize and Present the Data.

COURSE OUTCOMES:

CO No.	Outcomes	Bloom's Taxonomy Level
CO1	To Summarize the data using visual & summary analytics and common probability distributions	L2
CO2	To make inference about a sample & population using hypothesis test.	L2
CO3	To fit, interpret, and assess regression models and classification with one or more predictors.	L6
CO4	To assess the data integrity and data relevancy to a specific application	L3
CO5	To understand the significance of clustering and classification	L3

List of Experiments activities to be conducted

1. R AS CALCULATOR APPLICATION a. Using with and without Python/R objects on console b. Using Python/R mathematical functions on console
2. DESCRIPTIVE STATISTICS IN R a. Write an Python/R script to find basic descriptive statistics using summary, str, quartile function on mtcars & cars datasets.

b. Write an Python/R script to find subset of dataset by using subset (), aggregate () functions on iris dataset.
3. READING AND WRITING DIFFERENT TYPES OF DATASETS a. Reading types of data sets (.txt, .csv) from web and disk and writing in file in specific disk location. different b. Reading Excel data sheet in Python/R. c. Reading XML dataset in Python/R.
4. VISUALIZATIONS USING PYTHON/R a. Find the data distributions using box and scatter plot. b. Find the outliers using plot. c. Plot the histogram, bar chart and pie chart on sample data
5. CORRELATION AND COVARIANCE USING Python/R a. Find the correlation matrix. b. Plot the correlation plot on the dataset and visualize giving an overview of relationships among data on iris data. c. Analysis of covariance: variance (ANOVA), if data have categorical variables on iris data.
6. REGRESSION MODEL using Python/R Import data from web storage. Name the dataset and now do Logistic Regression to find out relation between variables that are affecting the admission of a student in an institute based on his or her GRE score, GPA obtained and rank of the student. Also check if the model is fit or not.
7. MULTIPLE REGRESSION MODEL Apply multiple regressions, if data have a continuous independent variable. Apply on above dataset
8. REGRESSION MODEL FOR PREDICTION Apply regression Model techniques to predict the data on above dataset.
9. CLASSIFICATION MODEL a. Install relevant package for classification. b. Choose classifier for a classification problem. c. Evaluate the performance of classifier.
10. CLUSTERING MODEL a. Clustering algorithms for unsupervised classification. b. d. Plot the cluster data using R visualizations.