R.V.R & J.C. College of Engineering (A)
Department of Computer Science and Engineering (Data Science)
Minor Degree Course
CDMR2: ANALYSING, VISUALIZING, AND APPLYING DATA SCIENCE

Prerequisite:
Knowledge is also assumed on basic concepts of Data Science and mathematics.
Introduction:

- To learn how to use python for data science.
- To understand and use all the tools and libraries of python for data science.

Course Contents:

Unit 1:
Pandas: Series, data frames, read csv, read json, data types, missing values, analyzing data.
SciPy: Constants, Optimisers, Sparse data, graphs, spatial data, interpolation.
Scikit-learn: Modelling Process, Data Representation, Conventions, Linear Modeling, implement algorithms of machine learning.

Unit 2:
Matplotlib: Visualizing Data: bar plots, grouped and stacked bars, dot plots and heat maps Visualizing Distributions: Histograms and density plots, Empirical cumulative distribution functions and q-q plots, scatterplots, scatterplot matrix, ggplots, correlograms.

Unit 3:
R: fundamentals of R, Control statements, functions, strings, lists, arrays, data frames in R, Data visualization in R.

Unit 4:
GIT: Version Control and GitHub, Version Control, Github and Git, Linking Github and R Studio, Projects under Version Control.

Lab Work:

1. Explore New York City - 311 Complaints and Housing datasets.
2. Analyze and Visualize data using Python.
3. Perform feature engineering exercise using Python.
4. Build and validate predictive machine learning model using R.
5. Create and share Actionable Insights to real life data problems in python and R.

Text Books:

1. Data Visualization with Python and JavaScript, Kyran Dale, Shroff Publisher/O'Reilly Publishers.

2. Data Science Using Python and R by Chantal D. Larose and Daniel T. Larose, Wiley.
3. Domain-Specific Languages in R, Advanced Statistical Programming, Thomas Mailund
4. Hands-On Programming with R, Garrett Grolemund, 1st edition, O'Reilly Publishers.

Reference Books:

1. Data Science & Analytics (with Python, R, SPSS Programming), V.K. Jain, Khanna Publishing House.
2. Python for Data Science and Visualization -Beginners to Pro, Udemy.