Subject Name: Data Science **Subject code: BEIT605T.1**

Load	Lecture	Tutorial	Credits	College Assessment Marks	University Evaluation	Total Marks
3Hrs (Theory)	3	-	3	30	70	100

Aim: To study Data Science and Data Engineering

Prerequisite: - Mathematics, Statistic, Artificial Intelligence, Database Systems

Course Objectives:

1	To acquire skills in data preparatory and prepossessing steps
2	To learn the tools and packages in Python for data science
3	To acquire knowledge in data interpretation and visualization techniques

Course Outcome:

At the end of this course students are able to:

CO1	Apply quantitative modeling and data analysis techniques to the solution of real world business problems, communicate findings, and effectively present results using data science tools
CO2	Apply principles of Data Science to the analysis of business problems.
CO3	Apply ethical practices in everyday business activities and make well -reasoned ethical business and data management decisions
CO4	Demonstrate knowledge of statistical data analysis techniques utilized in business decision making
CO5	Apply algorithms to build machine intelligence

Unit 1	**Introduction to Data Science** Need for data science benefits and uses facets of data data science process setting the research goal retrieving data cleansing integrating and transforming data exploratory data analysis build the models presenting and building applications

Unit 2	**Mathematical Foundations for Data Science** Basics of Data Science: Introduction; Typology of problems; Importance of linear algebra, statistics and optimization from a data science perspective; structured thinking for solving data science problems. Linear Algebra: Matrices and their properties (determinants, traces, rank, nullity, etc.); Eigenvalues and eigenvectors; Matrix factorizations; Inner

	products; Distance measures; Projections; Notion of hyperplanes; half -planes. Probability, Statistics and Random Processes: Probability theory and axioms; Random variables; Probability distributions and density functions (univariate and multivariate); Expectations and moments; Covariance and correlation; Statistics and sampling distributions; Hypothesis testing of means, proportions, variances and correlations; Confidence (statistical) intervals; Correlation functions; White-noise process.Optimization: Unconstrained optimization; Necessary and sufficient conditions for optima; Gradient descent methods; Constrained optimization, KKT conditions; Introduction to non -gradient techniques; Introduction to least squares optimization; Optimization view of machine learning.Introduction to Data Science Methods: Linear regression as an exemplar function approximation problem; Linear classification problems.
Unit 3	**PYTHON FOR DATA HANDLING** Basics of Numpy arrays aggregations computations on arrays comparisons masks boolean logic fancy indexing structured arrays Data manipulation with Pandas data indexing and selection operating on data missing data hierarchical indexing combining datasets – aggregation and grouping – pivot tables
Unit 4	**PYTHON FOR DATA VISUALIZATION** Visualization with matplotlib – line plots – scatter plots – visualizing errors – density and contour plots – histograms, binnings, and density – three dimensional plotting – geographic data – data analysis using statsmodels and seaborn – graph plotting using Plotly – interactive data visualization using Bokeh
Unit 5	**Advanced Data Analysis** Decision Trees: What Is a Decision Tree? Entropy, The Entropy of a Partition, Creating a Decision Tree, Random Forests Neural Networks : Perceptrons, Feed-Forward Neural Networks, Backpropagation, Example: Defeating a CAPTCHA MapReduce : Why MapReduce? Examples like word count and matrix multiplication