

CSE313: DATA ANALYTICS TOOLS

Hours/Week: 3

Credits: 3

I.A. Marks: 30

Exam. Marks: 70

Course Learning Objectives: Students will try to learn

1. To learn the probability distributions and density estimations to perform analysis of various kinds of data
 2. To explore the statistical analysis techniques using Python and R programming languages.
 3. To expand the knowledge in R and Python to use it for further research.
 4. The students will be able to carry out data analysis/statistical analysis Effectively visualize the data.
-

Course Outcomes: After completing the course, the students will be able to,

CO1: Course Outcomes: After completing the course, the students will be able to

CO1: Understand the fundamentals of data analytics.

CO2: Study the basic concepts of Excel spreadsheet Functions

CO3: Realize the importance of filtering functions, charts and tables.

CO4: Identify the importance and usage of R package and its features

CO5: Learn the fundamentals of python programming

CO6: understand the various search methods and visualization techniques.

CO7: learn to use various techniques for mining data stream.

CO8: understand the applications using Map Reduce Concepts.

CO9: introduce programming tools PIG & HIVE in Hadoop ecosystem.

UNIT-I

9 Hrs.

Introduction to data analytics (DA), data preparation, and data cleaning, Data types and measures of similarity, Data Pre-processing and numerosity reduction, Introduction to **data analysis techniques**: Basic analysis techniques, Statistical hypothesis generation and testing, Correlation analysis, Maximum likelihood test, **Regression analysis**, **Classification techniques**, clustering.

UNIT-II

9 Hrs.

Introduction to Spreadsheets: Reading data into Excel using various formats, Basic functions in Excel, arithmetic as well as various logical functions, Formatting rows and columns, Using formulas in Excel and their copy and paste using absolute and relative referencing. Spreadsheet Functions to Organize Data: IF and the nested IF functions, **VLOOKUP** and HLOOKUP, The RANDBETWEEN function. **Introduction to Filtering, Pivot Tables, and Charts**: VLOOKUP across worksheets, Data filtering in Excel, Use of Pivot tables with categorical as well as numerical data, Introduction to the charting capability of Excel. Advanced Graphing and Charting: Line, Bar and Pie charts, Pivot charts, Scatter plots, Histograms

UNIT-III

9 Hrs.

Getting Started and Basics: An introductory R session, R as a calculator, Vectors and matrices, Getting help and loading packages, Data entry and exporting data. Exploratory Data Analysis with R: Summary statistics, **Probability and Distribution**- Generate numbers- the built-in distribution for cumulative distribution functions, quantiles and random numbers, Graphics in R - histograms, empirical cumulative distribution, **QQ-plots, box plots, bar plots**, dot charts and pie charts.

UNIT-IV

Introduction to data analytics with Python: Importance of Python Programming, features of Python, Brief background to python, introduction to Jupyter and numpy, pandas, visualization. Significance of data analytics of Python programming.

REFERENCE BOOKS:

1. Joe Zhu ,Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets and DEA Excel Solver, Springer US
2. Peter Weverka, Office 2019 A L L - I N - O N E for dummies
3. Robert Gentleman Kurt Hornik Giovanni Parmigiani, Applied Spatial Data Analysis with R, Springer
4. Paul Cornell, Beginning Excel what-if data analysis tools : getting started with Goal Seek, data tables, scenarios and Solver , Apress
5. Eric Mayor, Learning Predictive Analytics with R: Get to grips with key data visualization and predictive analytic skills using R, Packt Publishing
6. Gerhard Svolba, Data Preparation for Analytics Using SAS, SAS Press
7. Anil Maheshwari, Data analytics , McGraw-Hill Education 2017
8. Bharti Motwani, Data Analytics using Python, Wiley, 2020
9. John Paul Mueller, PYTHON FOR DATA SCIENCE FOR DUMMIES, Wiley 2015

