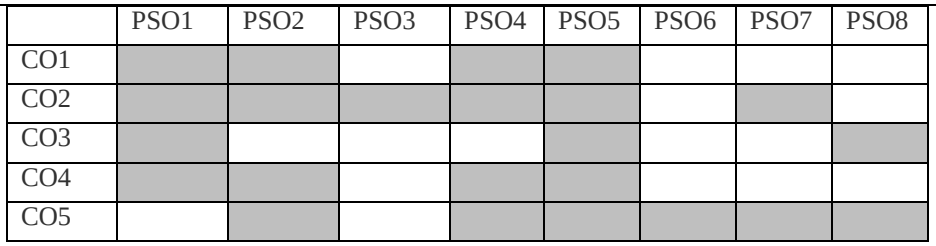
Course Code: 105-04
Course Title: Basics of Data Science

Course Code	105-04
Course Title	Basics of Data Science
Credits	4
Course Category	Major Course
Level of Course	200-299 (Intermediate Level)
Teaching per Week	Theory/Lectures: 2 Hours/Week and Lab./Practical Hours: 4 Hours/Week
Minimum weeks per Semester	15 (Including class work, examination, preparation etc.)
Review / Revision	-
Implementation Year:	A.Y. 2024-2025
Purpose of Course	This course is designed to introduce the fundamental concepts of Data Science to beginners. It covers essential methodologies and tools needed to understand, analyse, and interpret data effectively. The course provides hands-on experience with real-world data, preparing students for further studies and careers in Data Science.
Course Objective	To learn and obtain the skills related to i) Understand the basics of Data Science and its applications. ii) Gain proficiency in statistical analysis and handling of various types of data. iii) Learn to use python for Data Science tasks. iv) Understand and apply machine learning algorithms. v) Develop skills in data visualization and interpretation.
Pre-requisite	-
Course Outcomes	**CO1:** Identify and apply the right Concepts of Data Science and tools to solve problems. **CO2:** Manipulate and analyze data effectively using python. **CO3:** Implement basic machine learning algorithms to solve real- world problems. **CO4:** Create meaningful data visualizations to interpret and present data insights. **CO5:** Critically evaluate data analytics techniques in the context of various business and research scenarios.
Mapping between Course Outcomes(CO) with Program Specific Outcomes(PSO)	
Course Content	**Unit 1: Concepts of Data Science** 1.1 Introduction to Data Science 1.2 Definition of Data Science 1.3 Importance and Scope of Data Science 1.4 Data Analytics life cycle 1.5 Understanding data: 1.5.1 Types of Data: Numeric, Categorical, Graphical, High Dimensional Data **Unit-2 : Classification of digital Data:** 2.1 Structured, Semi-Structured and Un Structured 2.2 Sources of Data: 2.2.1 Time Series

	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6	PSO7	PSO8
CO1								
CO2								
CO3								
CO4								
CO5								

	2.2.2 Transactional Data 2.2.3 Biological Data 2.2.4 Spatial Data 2.2.5 Social Network Data 2.2.6 Example Applications. **UNIT-3: Concepts of Database:** 3.1 Database characteristics: 3.1.1 Data Independence (Logical and Physical) 3.1.2 Components of Database (User, Application , DBMS, Database) 3.1.3 Database Architecture (1-tier, 2-tier, 3-tier) 3.1.3.1 Comparison, advantages and disadvantages. 3.2 Database Models (Hierarchical, Network, E/R, Relational) 3.2.1 E/R model : Entity, Relationship, Attribute 3.2.2 E/R Diagram : One to one, one to many , many to one, many to many 3.2.3 Strong entity, weak entity 3.2.4 key attribute, derived attribute, Multi-valued attribute 3.3 Types of keys : 3.3.1 Super key, candidate key, Primary key, Composite key, Foreign key, Unique key. **UNIT-4: Normalization and Introduction of SQL:** (Max.Weightage: 20%) 4.1 Normalization (Insertion, Updating, Deletion anomalies) 4.2 Normalization Rules: 4.2.1 Concepts of Dependency, Transitive Dependency 4.2.2 Armstrong Axioms 4.2.3 1st Normal Form, 2nd Normal Form, 3rd Normal Form, B.C.N.F. **Unit-5: Structured Query Language:** (Max.Weightage: 25%) 5.1 Concepts of Structure Query Language (SQL) 5.1.1 SQL datatypes : int, float, double, char, varchar, number, varchar2, Text, date 5.2 DDL Statements : 5.2.1 Create , Drop, Truncate, Rename, Alter 5.3 DML and DQL(Data Query Language) Statements : 5.3.1 Insert, Update, Delete 5.3.2 select 5.4 Working with Tables. 5.4.1 Create,Drop,Alter tables 5.4.2 Insert, Update, Delete, Select queries on table 5.5 Constraints (Table level and Attribute Level) 5.5.1 NOT NULL, CHECK, DEFAULT 5.5.2 UNIQUE, Primary Key, Foreign Key 5.5.3 On Delete Cascade
Reference Books	1. "Data Science from Scratch" by Joel Grus - A clear introduction to the field, using Python. 2. "Python Data Science Handbook" by Jake VanderPlas - Comprehensive guide to the Python ecosystem for data analysis and visualization. 3. "Practical Statistics for Data Scientists" by Peter Bruce and Andrew Bruce - Covers essential statistical methods for data analysis in Data Science. 4. "Machine Learning Yearning" by Andrew Ng - Insights on how to structure Machine Learning projects. 5. "Storytelling with Data" by Cole Nussbaumer Knaflic - Teaches the fundamentals of data visualization and how to communicate effectively with data.

	6. "Data Science from Scratch" by Joel Grus - A clear introduction to the field, using Python. 7. "Python Data Science Handbook" by Jake VanderPlas - Comprehensive guide to the Python ecosystem for data analysis and visualization. 8. "Practical Statistics for Data Scientists" by Peter Bruce and Andrew Bruce - Covers essential statistical methods for data analysis in Data Science. 9. "Machine Learning Yearning" by Andrew Ng - Insights on how to structure Machine Learning projects. 10. "Storytelling with Data" by Cole Nussbaumer Knaflic - Teaches the fundamentals of data visualization and how to communicate effectively with data.
Teaching Methodology	Class Work, Discussion, Lab work, Self-Study, Seminars and/or Assignments
Evaluation Method	50% Internal assessment. 50% External assessment.