Course Code: 205-02
Course Title: Data Retrieval and Analysis (D.R.A.)

Course Code	205-01
Course Title	Data Retrieval and Analysis (D.R.A.)
Credits	4
Course Category	Major Course
Level of Course	200-299 (Intermediate Level)
Teaching per Week	4 Hours
Minimum weeks per Semester	15 (Including class work, examination, preparation etc.)
Review / Revision	-
Implementation Year:	A.Y. 2024-2025
Purpose of Course	This syllabus covers essential SQL commands and techniques for statistical analysis of data, integrating Python for data manipulation and interaction with SQL databases. It also covers executing SQL commands using Python, converting CSV into SQL tables and SQL tables into CSV using python libraries.
Course Objective	1. To Learn fundamental SQL commands such as SELECT, INSERT, UPDATE, DELETE, and JOIN. Understand database creation, modification, and management techniques. 2. To Perform Statistical Analysis with SQL. Utilize SQL aggregate functions for statistical analysis. Execute complex queries to summarize and report data. 3. To Integrate Python with SQL Databases. Use Python libraries like `sqlite3`, `pandas`, and `SQLAlchemy` for database interaction. Execute SQL commands within Python scripts for seamless data manipulation. 4. Convert CSV files to SQL tables and vice versa using Python. 5. Develop Practical knowledge about Data Analysis: Apply SQL and Python skills to real-world data scenarios through hands-on practical experience. Design and implement efficient databases and perform insightful data analysis.
Pre-requisite	Basic knowledge of data, python, sql and csv.
Course Outcomes	**CO1:** Demonstrate proficiency in using SQL to create, modify, and query databases, including performing complex joins and aggregations for data analysis. **CO2:** Apply Python programming skills to manipulate data, including reading and writing data from various file formats like CSV and xls. **CO3:** Utilize Python libraries such as pandas and NumPy for advanced data manipulation and transformation tasks, ensuring data is in the required format for analysis. **CO4:** Integrate Python with SQL databases using libraries like SQLAlchemy and sqlite3, executing SQL commands within Python scripts for seamless data handling. **CO5:** Develop and implement complete data analysis processes using Python, automating the workflow from data extraction to transformation and loading into databases or files.

Mapping between Course Outcomes(CO) with Program Specific Outcomes(PSO)

	PSO 1	PSO2	PSO 3	PSO 4	PSO 5	PSO 6	PSO 7	PSO 8
CO1	■	■	■	■	■			
CO2	■	■		■	■			
CO3	■	■		■	■	■		
CO4	■	■		■	■	■		
CO5		■	■	■	■		■	
CO6		■	■		■		■	■

Course Content	**Unit 1: Introduction to SQL and Data Storage** 1.1 Basic SQL Commands (Create, Read, Update, Delete) 1.2 Data Storage and Management 1.2.1 Creating and Dropping Tables 1.2.2 Understanding Data Types 1.2.3 ALTER , MODIFY, DROP 1.2.4 ROWID, ROWNUM 1.3 Concepts of NULL, NOT NULL 1.4 Aggregate Functions (COUNT, SUM, AVG, MAX, MIN) 1.5 Statistical Functions 1.5.1 STDDEV (Standard Deviation), VARIANCE, MEDIAN 1.5.2 PERCENTILE_CONT (Continuous Percentile) 1.5.3 PERCENTILE_DISC (Discrete Percentile) 1.5.4 CORR (Correlation) **Unit 2: Data Retrieval and Basic Analysis** 2.1 Exporting and Importing Data using Python (Pandas, Numpy Libraries) 2.1.1 Converting SQL Tables to CSV 2.1.2 Importing CSV Files into SQL 2.2 Using SQL with Python 2.2.1 Setting up SQL Database Connections in Python 2.2.2 Executing SQL Commands using Python (SQLite, MySQL, PostgreSQL) 2.2.3 Retrieving and Manipulating Data with Pandas and SQLAlchemy **UNIT 3: Working with multiple tables:** 3.1 Filtering Data : (WHERE Clause, IN, BETWEEN, LIKE) 3.2 Joining Tables 3.2.1 INNER JOIN 3.2.2 LEFT JOIN , RIGHT JOIN, FULL JOIN, SELF JOIN 3.3 Sorting Data : 3.3.1 ORDER BY Clause 3.3.2 Sorting by Multiple Columns 3.4 Grouping Data 3.4.1 GROUP BY Clause 3.4.2 HAVING Clause 3.5 SET operations (UNION, INTERSECT, EXCEPT) **Unit 4: Advanced Data Retrieval and Manipulation :** 4.1 Subqueries 4.1.1 Nested SELECT Statements 4.1.2 Correlated Subqueries 4.2 Common Table Expressions (CTEs) 4.2.1 Writing Recursive and Non-Recursive CTEs 4.2.2 Simplifying Complex Queries with CTEs 4.2.3 Cumulative Sums **Unit 5: Performance Optimization and Practical Applications:** 5.1 Query Optimization 5.1.1 Strategies for Optimizing SQL Queries 5.1.2 Identifying Performance Bottlenecks 5.2 Using EXPLAIN or EXPLAIN PLAN 5.2.1 Analyzing Query Execution Plans 5.2.2 Improving Query Performance Based on Execution Plans 5.3 Indexing 5.3.1 Creating and Managing Indexes 5.3.2 Impact of Indexes on Query Performance 5.4 Advanced Data Manipulation - INSERT INTO ... SELECT - UPDATE ... FROM, - DELETE with Conditions

Reference Books	1. Learning SQL by Alan Beaulieu, O'Reilly Media, ISBN: 978-0596520830 2. Python for Data Analysis by Wes McKinney, O'Reilly Media, ISBN: 978-1491957660 3. SQL for Data Scientists: A Beginner's Guide for Building Datasets for Analysis by Renee M. P. Teate, Wiley, ISBN: 978-1119669363 4. Python Data Science Handbook by Jake VanderPlas, O'Reilly Media, ISBN: 978-1491912058 5. SQL in 10 Minutes, Sams Teach Yourself by Ben Forta, Sams Publishing, ISBN: 978-0672336072 6. SQL, PL/SQL: The Programming Language of Oracle by Ivan Bayross, BPB Publications, ISBN: 978-8183331168 7. Python Programming: A Modern Approach by Vamsi Kurama, Pearson India, ISBN: 978-9332585342 8. Data Science and Analytics (with Python, R and SPSS Programming) by V.K. Jain, Khanna Publishing, ISBN: 978-9382609800 9. Mastering Python for Data Science by Samir Madhavan, Packt Publishing (India), ISBN: 978-1783553358 10. Database Management Systems by Raghu Ramakrishnan and Gehrke, McGraw Hill Education (India), ISBN: 978-0071231510
Teaching Methodology	Class Work, Discussion, Lab work, Self-Study, Seminars and/or Assignments
Evaluation Method	50% Internal assessment. 50% External assessment.