

Department of Computer Science Engineering – (Data Science)

COURSE MODULE OF THE SUBJECT TAUGHT FOR THE SESSION 2026-27 (ODD SEM)

Course Syllabi with CO's

Faculty Name: Mr. Chandan A			Academic Year: 2026 - 2027				
Department: Computer Science & Engineering- Data Science							
Course Code	Course Title	Core / Elective	Prerequisite	Contact Hours			Total Hrs/ Sessions
				L	T	P	
1BCSL306	EXPLORATORY DATA ANALYSIS LAB		AEC	0	0	2	24
Topics Covered as per Syllabus							
List of Experiments							
1. Load a real-world dataset such as Student Performance, Employee Information, or Sales Data using Pandas in Jupyter Notebook. Explore the dataset by displaying the first few records, identifying the number of rows and columns, examining data types, and generating a summary of the dataset to understand its structure and contents. a. Expected Input: CSV dataset containing records and attributes. b. Expected Output: Display of dataset records, dimensions, column names, data types, and summary information. c. Learning Outcome: Ability to load and explore datasets using Pandas and understand the structure of real-world data. 2.Extract numerical attributes from a dataset and perform basic statistical analysis using NumPy and Pandas. Compute measures such as mean, median, mode, standard deviation, minimum, and maximum values, and interpret the results to understand the distribution and characteristics of the data. a. Expected Input: Dataset containing numerical attributes. b. Expected Output: Statistical summary showing central tendency and dispersion measures. c. Learning Outcome: Ability to perform basic statistical analysis and interpret numerical data characteristics.							

Department of Computer Science Engineering – (Data Science)

3. Analyze the quality of a dataset by identifying missing values, counting null entries in each attribute, and calculating the percentage of missing data. Generate a simple data quality report highlighting attributes that require cleaning before further analysis.

- a. Expected Input: Dataset containing missing or incomplete records.
- b. Expected Output: Missing value count, percentage of missing data, and data quality report.
- c. Learning Outcome: Ability to assess data quality and identify issues affecting analysis.

4. Perform data cleaning by handling missing values using suitable techniques such as replacement with mean, median, or mode values and removal of incomplete records. Compare the dataset before and after cleaning and analyze the impact of the cleaning process.

- a. Expected Input: Dataset containing missing values.
- b. Expected Output: Cleaned dataset and comparison report before and after preprocessing.
- c. Learning Outcome: Ability to apply data cleaning techniques and prepare datasets for analysis.

5. Detect and remove duplicate records from a dataset using Pandas functions. Compare the number of records before and after duplicate removal and discuss the importance of maintaining data consistency in analytical applications.

- a. Expected Input: Dataset containing duplicate records.
- b. Expected Output: Duplicate record count and cleaned dataset without duplicate entries.
- c. Learning Outcome: Ability to identify and eliminate duplicate records to improve data quality.

6. Perform data filtering, sorting, and selection operations on a dataset to answer business-oriented questions such as identifying high-performing students, top-selling products, or employees with the highest salaries. Present the extracted information in a meaningful format.

- a. Expected Input: Dataset with categorical and numerical attributes.
- b. Expected Output: Filtered and sorted data satisfying specified conditions.
- c. Learning Outcome: Ability to retrieve and analyze relevant information from large datasets using query operations.

Department of Computer Science Engineering – (Data Science)

7. Apply grouping and aggregation techniques using Pandas to generate summaries from a dataset. Compute category-wise counts, averages, totals, and other statistical measures to derive meaningful insights from grouped data.

- a. Expected Input: Dataset containing categorical and numerical attributes.
- b. Expected Output: Group-wise summary tables and aggregated results.
- c. Learning Outcome: Ability to summarize and analyze data using grouping and aggregation techniques.

8. Conduct exploratory statistical analysis on a dataset by generating descriptive statistics and identifying patterns, variations, and unusual observations. Interpret the results to understand the overall behavior of the dataset and support data-driven decision-making.

- a. Expected Input: Real-world dataset.
- b. Expected Output: Statistical report highlighting key observations and trends.
- c. Learning Outcome: Ability to perform exploratory statistical analysis and derive meaningful insights.

9. Create bar charts and pie charts using Matplotlib or Seaborn to visualize categorical data from a real world dataset. Analyze the visualizations and draw conclusions regarding trends, distributions, and comparisons among different categories.

- a. Expected Input: Dataset containing categorical attributes.
- b. Expected Output: Bar chart and pie chart visualizations.
- c. Learning Outcome: Ability to visualize categorical data and communicate insights effectively.

10. Develop line plots and scatter plots to visualize trends and relationships among variables in a dataset. Analyze the generated plots to identify growth patterns, variations, correlations, and potential business insights.

- a. Expected Input: Dataset containing numerical attributes.
- b. Expected Output: Line plots and scatter plots showing trends and relationships.
- c. Learning Outcome: Ability to analyze trends and relationships using graphical visualization techniques.

Department of Computer Science Engineering – (Data Science)

11. Perform correlation analysis on a dataset by computing the correlation matrix and visualizing it using a heat map. Identify strongly and weakly related attributes and discuss the significance of correlation analysis in exploratory data analysis and machine learning applications.
 - a. Expected Input: Dataset with multiple numerical attributes.
 - b. Expected Output: Correlation matrix and heat map visualization.
 - c. Draft Syllabus Learning Outcome: Ability to analyze relationships among variables and identify influential attributes.

12. Carry out a comprehensive Exploratory Data Analysis (EDA) micro-project on a real-world dataset from domains such as education, healthcare, finance, retail, sports, or social media. The analysis shall include dataset understanding, data quality assessment, data cleaning, statistical analysis, visualization, correlation analysis, and presentation of key findings and recommendations through a Jupyter Notebook report.
 - a. Expected Input: Real-world dataset selected by the student.
 - b. Expected Output: Complete EDA report including visualizations, statistical summaries, observations, and recommendations.
 - c. Learning Outcome: Ability to independently perform end-to-end exploratory data analysis on real-world datasets and communicate actionable insights.

List of Textbook and URLs

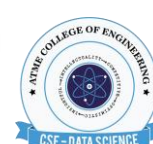
Text books:

1. Jake VanderPlas, Python Data Science Handbook: Essential Tools for Working with Data, 2nd Edition, O'Reilly Media, 2022.
2. Elbert L. Heaton, Data Analytics with Applications in Business, LibreTexts Open Educational Resource, 2025.

Reference books / Manuals:

1. Wayne L. Winston, Microsoft Excel 2019 Data Analysis and Business Modeling, 6th Edition, Microsoft Press, 2019.
2. Wes McKinney, Python for Data Analysis: Data Wrangling with Pandas, NumPy and Jupyter, 3rd Edition, O'Reilly Media, 2022.
3. Documentation of NumPy, Pandas, Matplotlib and Seaborn Libraries available at:
<https://numpy.org> <https://pandas.pydata.org> <https://matplotlib.org> <https://seaborn.pydata.org>

Course Outcomes	At the end of the course, the student will be able to: CO1- Compare real-world datasets having different dataset structures and characteristics using Jupyter Notebook, NumPy, and Pandas. CO2- Apply descriptive statistical techniques to summarize, analyze, and interpret data for exploratory analysis. CO3- Make use of appropriate preprocessing techniques to handle data quality issues such as missing values, duplicate records, and inconsistencies. CO4 - Perform data filtering, sorting, grouping, and aggregation operations to extract meaningful information from datasets. CO5 - Interpret visualisations using Matplotlib and Seaborn to identify trends, patterns, and relationships in data.
------------------------	--



Department of Computer Science Engineering – (Data Science)

Assessment Detail: The assessment for each course is equally divided between Continuous Internal Evaluation (CIE) and the Semester End Examination (SEE), with each component carrying 50% weightage (i.e., 50 marks each).

The CIE marks awarded shall be based on the continuous evaluation of the laboratory report using a defined set of rubrics. Each experiment report can be evaluated for 30 marks. Total marks for the all report shall be scaled down to 30 marks. The laboratory test (duration 03 hours) at the end of the last week of the semester /after completion of all the experiments (whichever is early) shall be conducted for 50 marks and Marks scored shall be scaled down to 20 marks. Final average of marks for report and test shall be out of 50 marks

Department of Computer Science Engineering – (Data Science)

The Correlation of Course Outcomes (CO's) and Program Outcomes (PO's)

Course Code:	1BAIL307A		TITLE: Exploratory Data Analysis Lab					Faculty Name:	Mr. Chandan A		
List of Course Outcomes	Program Outcomes										
	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11
CO-1	2	2	-	-	2	-	-	-	-	-	-
CO-2	2	2	1	3	3	-	-	-	-	-	-
CO-3	-	-	1	-	1	-	-	-	-	-	1
CO-4	2	-	1	3	3	-	-	-	-	-	-
CO-5	-	-	-	2	2	-	-	-	-	-	-
CO-6	-	-	1	-	1	-	-	-	1	-	1

Note: 3 = Strong Contribution, 2 = Average Contribution, 1 = Weak Contribution, - = No Contribution

The Correlation of Course Outcomes (CO's) and Program Specific Outcomes (PSO's)

Course Code	1BAIL307A	Title: EDA Lab	Faculty name: Mr. Chandan A
List of Course Outcomes	Program Specific Outcomes		
	PSO1	PSO2	PSO3
CO1	2	-	-
CO2	2	-	-
CO3	2	-	-
CO4	2	-	-
CO5	2	-	-
CO6	2	-	-

Note: 3 = Strong Contribution, 2 = Average Contribution, 1 = Weak Contribution, - = No Contribution