



SRI VENKATESWARA UNIVERSITY

**4-Year UG Honours in B.Sc. (Artificial Intelligence) as Major in consonance
with Curriculum framework w.e.f. AY 2026-27**

COURSE STRUCTURE (for Semester III)

Year	Semester	Course	Title of the Course	No. of Hrs /Week	No. of Credits
II	III	5	Data Base Management System	3	3
			Data Base Management System Lab	2	1
		6	Statistical Foundation of AI	3	3
			Statistical Foundation of AI Lab	2	1
		7	Exploratory Data Analysis & Data Visualization	3	3
			Exploratory Data Analysis & Data Visualization Lab	2	1



SRI VENKATESWARA UNIVERSITY

II Year Syllabus for 4-Year UG Honours in B.Sc. (Artificial Intelligence) as Major in consonance with the Curriculum framework w.e.f. AY 2026-27

II Year III Semester

Sl. No.	Category	Course No.	Course Name	No. of Hours		Total No. of Hours	No. of Credits		Total No. of Credits
				Theory	Practical		Theory	Practical	
1	Major-Core	5	Data Base Management System	3	2	5	3	1	4
2	Major-Core	6	Statistical Foundation of AI	3	2	5	3	1	4
3	Major-Core	7	Exploratory Data Analysis & Data	3	2	5	3	1	4
4	Minor	1		3	2	5	3	1	4
5	AECC	1	English	4	0	4	3	0	3
6	AECC	1	MIL (Telugu/Hindi/Sanskrit)	4	0	4	3	0	3
7	MDC	2	-	2	0	2	2	0	2
8	SEC	3	SEC: Design Thinking/Problem solving/ others	2	0	2	2	0	2
				24	8	32	22	4	26

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Artificial Intelligence Honours
II year III Semester
COURSE 7: EXPLORATORY DATA ANALYSIS & DATA VISUALIZATION

Theory

Credits: 3

3 hrs/week

Course Objectives

1. To introduce the concepts, importance, and workflow of exploratory data analysis.
2. To impart practical skills in data pre-processing, cleaning, and manipulation using Python.
3. To equip students with techniques in univariate, bivariate, and multivariate data analysis.
4. To develop proficiency in data visualization and interpretation using standard Python libraries.
5. To enable students to build reports and dashboards to communicate data insights effectively.

Course Outcomes

After completing this course, students will be able to:

1. Explain the importance and process of EDA in the data science pipeline.
2. Perform robust data pre-processing including missing value handling, outlier detection, transformation, and encoding.
3. Manipulate, filter, and reshape datasets using Python libraries like Pandas and NumPy.
4. Conduct data analysis and visualize results using Matplotlib, Seaborn, and Plotly.
5. Design interactive visualizations and dashboards to communicate findings in real-world scenarios.

Syllabus:

Unit 1: Introduction to Exploratory Data Analysis (EDA)

Significance and objectives of EDA, Types and scales of data (nominal, ordinal, interval, ratio), Concepts of variability and central tendency, Role of EDA in the data science pipeline, Introduction to data quality and data issues, Differences between EDA, classical, and Bayesian analysis, Basic terminology and concepts

Unit 2: Data Pre-processing and Cleaning

Overview of data pre-processing, Handling missing types of missingness (MCAR, MAR, MNAR), Techniques for missing data imputation (mean, median, mode, forward/backward fill), Detection and treatment of duplicates, Outlier detection methods (Z-score, IQR method), Data transformation: normalization and standardization, Encoding categorical variables: label encoding, one-hot encoding, Data integration and reduction techniques, Importance of pre-processing for machine learning models

Unit 3: Data Manipulation using Python with Pandas and NumPy

NumPy Array Creation: Use `np.array()`, `np.zeros()`, `np.ones()` to create arrays of various dimensions. **Array Manipulation:** Reshape, transpose, flatten arrays for flexible data formats. Element-wise Arithmetic Operations, Statistical Functions on NumPy Arrays **Indexing & Slicing:** Access array elements via slicing, masking, and fancy indexing for selective data analysis.

Pandas DataFrames: Series and Data frames, Create from dictionaries, Series, CSV, Excel, and JSON files using Dataframe or reader functions

Data Frame Manipulation: Indexing & Selection - Label-based and Position-based indexing; conditional filtering with boolean masks, Aggregation & Grouping- Group data by columns with `.groupby()`, summarize using `.sum()`, `.mean()`, `.count()`. Data Reshaping & Pivoting and stack/unstack hierarchical index data.

Merging & Joining Data Frames

Datetime & Categorical Data: Handle and transform temporal data with `pd.to_datetime()`, categorical conversions with `.astype('category')`.

Unit 4: Basic Data Analysis and Visualization

What is visualization, what is its importance? Types: histograms, bar charts, Boxplots and violin plots for data distribution, Scatter plots and correlation analysis (Pearson, Spearman), Cross-tabulation and contingency tables, Analysing relationships: correlation vs causation, Visualizing categorical vs numerical data

Unit 5: Advanced Visualization and Reporting

Principles of effective graphical representation, Introduction to visualization libraries:

Matplotlib, Seaborn, Plotly, Advanced plots: heatmaps, pair plots, joint plots, Interactive visualizations and dashboards, Geographical data visualization basics, Storytelling with using visualization for communication, Introduction to reporting tools and best practices

Activities

Unit 1: Introduction to EDA

Activity: Analyse the Titanic dataset to explore data types, central tendency, and variability.

Outcome: Develop a foundational understanding of data structures and EDA significance.

Evaluation Method: Lab notebook submission and oral quiz.

Unit 2: Data Pre-processing and Cleaning

Activity: Clean a health survey dataset: identify/impute missing values, standardize columns, encode categories, remove duplicates/outliers.

Outcome: Demonstrate proficiency in data cleaning and pre-processing techniques in Python.

Evaluation Method: Script submission and report on pre-processing steps with before/after summary.

Unit 3: Data Manipulation using Python

Activity: Manipulate a retail transaction dataset by indexing, filtering, grouping, merging, and handling date/time columns.

Outcome: Apply multiple Pandas/NumPy techniques for dataset transformation.

Evaluation Method: Graded Jupyter notebook, peer code review.

Unit 4: Basic Data Analysis and Visualization

Activity: Explore demographic data with descriptive statistics, create histograms, boxplots, scatter plots, and cross-tabulations.

Outcome: Visualize and interpret univariate and bivariate data insights.

Evaluation Method: Visualization notebook and critical analysis write-up.

Unit 5: Advanced Visualization and Reporting

Activity:

- Build a dashboard on sales or air quality data using Matplotlib/Seaborn/Plotly.
- Incorporate interactive and geospatial visualizations.
- Develop professional dashboards and communicate insights through visual storytelling.

Evaluation Method: Final project dashboard evaluation and presentation.

Preferred Textbooks

1. Suresh Kumar Mukhiya, Usman Ahmed, “Hands-On Exploratory Data Analysis with Python”, Packt Publishing, 2020.
2. Jake Vander Plas, “Python Data Science Handbook: Essential Tools for Working with Data”, O’Reilly, 2017.
3. Catherine Marsh, Jane Elliott, “Exploring Data: An Introduction to Data Analysis for Social Scientists”, Wiley, 2008.

References

1. Eric Pimpler, “Data Visualization and Exploration with R”, GeoSpatial Training Service, 2017.
2. Claus O. Wilke, “Fundamentals of Data Visualization”, O’Reilly, 2019.
3. Matthew O. Ward, Georges Grinstein, Daniel Keim, “Interactive Data Visualization: Foundations, Techniques, and Applications”, CRC Press, 2015.

SEMESTER-III

COURSE 7: EXPLORATORY DATA ANALYSIS & DATA VISUALIZATION

Practical

Credits: 1

2 hrs/week

Note: This lab has to be executed using interactive computing environments like Jupyter Notebook or Google Colab.

1. Importing Data and Basic Exploration
 - Load datasets from CSV/Excel/JSON
 - Use `.head()`, `.info()`, `.describe()`, `.shape` to examine structure and summary statistics
2. Examining Data Types and Missing Values
 - Identify data types for each column
 - Detect and quantify missing values with `.isnull()` and `.sum()`.
3. Handling Missing Data Techniques
 - Impute missing values using mean, median, mode, forward, and backward fill
 - Drop rows/columns with missing data and compare results
4. Dealing with Duplicates and Outliers
 - Find and remove duplicate entries
 - Detect outliers using Z-score or IQR methods
 - Treat or remove outliers and compare distributions
5. Data Transformation and Scaling
 - Normalize, standardize, log-transform columns
 - Encode categorical variables (Label Encoding, One-Hot Encoding)
6. Indexing, Filtering, and Slicing DataFrames
 - Select specific rows/columns using index and Boolean filters
 - Conditional selections and advanced slicing
7. Aggregating, Grouping, and Pivoting Data
 - Use `.groupby()` for aggregate statistics
 - Create pivot tables for categorical summaries
8. Merging and Concatenating Multiple Datasets
 - Join, merge, concatenate DataFrames
 - Handle merging keys and missing arguments
9. Univariate Visualization Techniques
 - Create histograms, boxplots, violin plots for distributions
 - Visualize frequency and central tendency
10. Bivariate and Multivariate Visualization
 - Display scatter plots, correlation matrices, pair plots
 - Visualize relationships and patterns
11. Time Series Data Exploration
 - Parse and index datetime columns
 - Visualize trends and seasonality
12. Interactive Visualizations and Dashboards
 - Build simple interactive graphs with Plotly
 - Design a dashboard showing multiple aspects of a dataset

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Artificial Intelligence Honours
II year III Semester
COURSE 7: EXPLORATORY DATA ANALYSIS & DATA VISUALIZATION

MODEL QUESTION PAPER

Time: 3 Hours

Max. Marks: 75

PART-A

Answer any **FIVE** of the following. Each Question Carries 3 marks.

(5X3=15)

1. Short answer question from unit-I
2. Short answer question from unit-I
3. Short answer question from unit-II
4. Short answer question from unit-II
5. Short answer question from unit-III
6. Short answer question from unit-III
7. Short answer question from unit-IV
8. Short answer question from unit-IV
9. Short answer question from unit-V
10. Short answer question from unit-V

PART-B

Answer **FIVE** of the following. Each Question Carries 12 marks.

(5X12=60)

11. a) Essay question from Unit-I
OR
b) Essay question from Unit-I
12. a) Essay question from Unit-II
OR
b) Essay question from Unit-II
13. a) Essay question from Unit-III
OR
b) Essay question from Unit-III
14. a) Essay question from Unit-IV
OR
b) Essay question from Unit-IV
15. a) Essay question from Unit-V
OR
b) Essay question from Unit-V