PVP 23

PRASAD V. POTLURI SIDDHARTHA INSTITUTE OF TECHNOLOGY

(Autonomous)

Kanuru, Vijayawada-520007

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (AI&ML)

III B Tech – I Semester

Exploratory Data Analysis with Python

Course Code	**23AM4501D**	**Year**	III	**Semester**	I
Course Category	PEC	**Branch**	CSE (AI&ML)	**Course Type**	Theory
Credits	3	**L-T-P**	3-0-0	**Prerequisites**	Python Programming
Continuous Internal Evaluation	30	**Semester End Evaluation**	70	**Total Marks**	100

Course Outcomes		
Upon Successful completion of course, the student will be able to		
CO1	Describe EDA concepts, data types, and Python-based tools to understand data-driven analysis	**L2**
CO2	Apply data visualization and transformation techniques to analyze structured datasets using Python tools like Matplotlib, Seaborn, and Pandas	**L3**
CO3	Use datasets to perform statistical analysis, handle missing data, and explore patterns including time series	**L3**
CO4	Analyze and deploy predictive models using evaluation metrics to solve real-world data problems	**L4**

Contribution of course outcomes towards achievement of program outcomes & Strength of correlations (3: Substantial,2: Moderate,1: Slight)

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**P010**	**PO11**	**PSO1**	**PSO2**
CO1	**2**												
CO2	**3**												
CO3	**3**												
CO4		**3**									**2**		

PRASAD V. POTLURI SIDDHARTHA INSTITUTE OF TECHNOLOGY

(Autonomous)
Kanuru, Vijayawada-520007

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (AI&ML)

III B Tech – I Semester

Syllabus		
Unit No	**Contents**	**Mapped CO**
I	**Exploratory Data Analysis Fundamentals**: Understanding data science, the significance of EDA, steps in EDA, making sense of data, Numerical data, Categorical data, Measurement scales, Comparing EDA with classical and Bayesian analysis, Software tools available for EDA, getting started with EDA. **Sample Experiments:** 1. a) Download Dataset from Kaggle using the following link: https://www.kaggle.com/datasets/sukhmanibedi/cars4u b) Install python libraries required for Exploratory Data Analysis (NumPy, pandas, matplotlib, seaborn) 2. Perform NumPy Array basic operations and Explore NumPy Built-in functions. 3. Loading Dataset into pandas dataframe 4. Selecting rows and columns in the dataframe.	CO1
II	Visual Aids for EDA: Technical requirements, Line chart, Bar charts, Scatter plot using seaborn, Polar chart, Histogram, Choosing the best chart **Case Study:** EDA with Personal Email, Technical requirements, Loading the dataset, Data transformation, Data cleansing, Applying descriptive statistics, Data refactoring, Data analysis. **Sample Experiments:** 1. Apply different visualization techniques using sample dataset a. Line Chart b. Bar Chart c. Scatter Plots d. Bubble Plot 2. Generate Scatter Plot using seaborn library for iris dataset 3. Apply following visualization Techniques for a sample dataset a. Area Plot b. Stacked Plot c. Pie chart d. Table Chart 4. Generate the following charts for a dataset. a. Polar Chart b. Histogram c. Lollipop chart 5. Case Study: Perform Exploratory Data Analysis with Personal Email Data	CO1, CO2, CO4
III	Data Transformation: Merging database-style data frames, Concatenating along with an axis, merging on index, Reshaping and pivoting, Transformation techniques, handling missing data, Mathematical operations with NaN, Filling missing values, Discretization and binning, Outlier detection and filtering, Permutation and random sampling, Benefits of data transformation, Challenges. **Sample Experiments:** 1. Perform the following operations a) Merging Dataframes b) Reshaping with Hierarchical Indexing c) Data Deduplication d) Replacing Values 2. Apply different Missing Data handling techniques a) NaN values in mathematical Operations b) Filling in missing data c) Forward and Backward filling of missing values d) Filling with index values e) Interpolation of missing values 3. Apply different data transformation techniques a) Renaming axis indexes b) Discretization and Binning c) Permutation and Random Sampling d) Dummy variables	CO1, CO2, CO4

PRASAD V. POTLURI SIDDHARTHA INSTITUTE OF TECHNOLOGY

(Autonomous)
Kanuru, Vijayawada-520007

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (AI&ML)

III B Tech – I Semester

IV	**Descriptive Statistics:** Distribution function, Measures of central tendency, Measures of dispersion, Types of kurtosis, Calculating percentiles, Quartiles, Grouping Datasets, Correlation, Understanding univariate, bivariate, multivariate analysis, Time Series Analysis **Sample Experiments:** 1. Study the following Distribution Techniques on a sample data a) Uniform Distribution b) Normal Distribution c) Gamma Distribution d) Exponential Distribution e) Poisson Distribution f) Binomial Distribution 2. Perform Data Cleaning on a sample dataset. 3. Compute measure of Central Tendency on a sample dataset a) Mean b) Median c) Mode 4. Explore Measures of Dispersion on a sample dataset a) Variance b) Standard Deviation c) Skewness d) Kurtosis 5. a) Calculating percentiles on sample dataset b) Calculate Inter Quartile Range(IQR) and Visualize using Box Plots 6. Perform the following analysis on automobile dataset. a) Bivariate analysis b) Multivariate analysis 7. Perform Time Series Analysis on Open Power systems dataset	CO1, CO3, CO4
V	**Model Development and Evaluation**: Unified machine learning workflow, Data preprocessing, Data preparation, Training sets and corpus creation, Model creation and training, Model evaluation, Best model selection and evaluation, Model deployment Case Study: EDA on Wine Quality Data Analysis **Sample Experiments:** 1. Perform hypothesis testing using stats models library a) Z-Test b)T-Test 2. Develop model and Perform Model Evaluation using different metrics such as prediction score, R2 Score, MAE Score, MSE Score. 3. Case Study: Perform Exploratory Data Analysis with Wine Quality Dataset	CO1, CO3, CO4

Learning Resources
Text Books
1. Hands-On Exploratory Data Analysis with Python, Suresh Kumar Mukhiya, Usman Ahmed, 1st Edition, 2020, Packt Publishing.
References
1. Exploratory Data Analysis Using R, Ronald K. Pearson, 1st Edition, 2020, CRC Press. 2. Hands-On Exploratory Data Analysis with R: Become an Expert in Exploratory Data Analysis Using R Packages, Radhika Datar, Harish Garg, 1st Edition, 2019, Packt Publishing.
E-Recourses and other Digital Material

PVP 23

PRASAD V. POTLURI SIDDHARTHA INSTITUTE OF TECHNOLOGY

(Autonomous)
Kanuru, Vijayawada-520007

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (AI&ML)

III B Tech – I Semester

1. https://github.com/PacktPublishing/Hands-on-Exploratory-Data-Analysis-with-Python
2. https://www.analyticsvidhya.com/blog/2022/07/step-by-step-exploratory-dataanalysis-eda-using-python/#h-conclusion
3. https://github.com/PacktPublishing/Exploratory-Data-Analysis-with-Python-Cookbook