#  Cover Page

23CD22SC - EXPLORATORY DATA ANALYSIS WITH PYTHON

(Skill Enhancement Course)

(CSE(DS))

|  |  |  |  |
| --- | --- | --- | --- |
| Course Category: | Skill Enhancement Course | Credits: | 2 |
| Course Type: | Practical | Lecture-Tutorial-Practical: | 1-0-2 |
| Prerequisite: | Proficiency in **Python** Familiarity with **libraries for data manipulation** such as **pandas** and **NumPy**. Basic knowledge of **descriptive statistics** and **inferential statistics** | Sessional Evaluation:Univ. Exam Evaluation:Total Marks: | 3070100 |
| Objectives: | **Upon successful completion of the course, the students will be able to:** |
| * This course introduces the fundamentals of Exploratory Data Analysis
* It covers essential exploratory techniques for understanding multivariate data by summarizing it through statistical methods and graphical methods.
 |

|  |  |
| --- | --- |
| Course Outcomes | **Upon successful completion of the course, the students will be able to:** |
| CO1 | Understand the fundamentals of exploratory data analysis. (L2) |
| CO2 | Implement the data visualization using Matplotlib. (L5) |
| CO3 | Perform univariate data exploration and analysis |
| CO4 | Apply bivariate data exploration and analysis. (L3) |
| CO5 | Use Data exploration and visualization techniques for multivariate and time series data (L3) |
| Course Content | UNIT-IExploratory 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. Download Dataset from Kaggle using the following link :<https://www.kaggle.com/datasets/sukhmanibedi/cars4u>
	2. Install python libraries required for Exploratory Data Analysis (numpy, pandas,matplotlib,seaborn)
1. Perform Numpy Array basic operations and Explore Numpy Built-in functions.
2. Loading Dataset into pandas dataframe

4. Selecting rows and columns in the dataframeUNIT-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
2. Line Chart b) Bar Chart c) Scatter Plots d)Bubble Plot
3. Generate Scatter Plot using seaborn library for iris dataset
4. Apply following visualization Techniques for a sample dataseta) Area Plot b) Stacked Plot c) Pie chart d) Table Char
5. Generate the following charts for a dataset.

a) Polar Chart b)Histogram c)Lollipop chart1. Case Study: Perform Exploratory Data Analysis with Personal Email Data

UNIT-III**Data Transformation:** Merging database-style dataframes, Concatenating along with anaxis, Merging on index, Reshaping and pivoting, Transformation techniques, Handlingmissing 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
	* 1. Merging Dataframes
		2. Reshaping with Hierarchical Indexing
		3. Data Deduplication
		4. Replacing Values
2. Apply different Missing Data handling techniques

 a) NaN values in mathematical Operations1. Filling in missing data
2. Forward and Backward filling of missing values
3. Filling with index values
4. Interpolation of missing values
5. Apply different data transformation techniques
	* 1. Renaming axis indexes
		2. Discretization and Binning
		3. Permutation and Random Sampling
		4. Dummy variables

UNIT-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
	* 1. Uniform Distribution
		2. Normal Distribution
		3. Gamma Distribution
		4. Exponential Distribution
		5. Poisson Distribution
		6. Binomial Distribution
2. Perform Data Cleaning on a sample dataset.
3. Compute measure of Central Tendency on a sample dataset
	* 1. Mean b)Median c)Mode
4. Explore Measures of Dispersion on a sample dataset
	* 1. 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 Plots1. Perform the following analysis on automobile dataset.
	* 1. Bivariate analysis b)Multivariate analysis
2. Perform Time Series Analysis on Open Power systems dataset

UNIT-V**Model Development and Evaluation:** Unified machine learning workflow, Data pre processing, 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 statsmodels 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
 |
| Text Books &ReferencesBooks | **TEXTBOOKS:**1. Suresh Kumar Mukhiya, Usman Ahmed, Hands-On Exploratory Data Analysis with Python, Packt Publishing, 2020.

**REFERENCE BOOKS:**1. Ronald K. Pearson, Exploratory Data Analysis Using R, CRC Press, 2020
2. Radhika Datar, Harish Garg, Hands-On Exploratory Data Analysis with R: Become an expert in exploratory data analysis using R packages, Ist Edition, Packt Publishing, 2019
 |
| E-Resources | 1. <https://github.com/PacktPublishing/Hands-on-Exploratory-Data-Analysis-withPython>
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-PythonCookbook>
 |