

COURSE TITLE	EXPLORATORY DATA ANALYTICS AND VISUALIZATION
COURSE CODE	05MD0301
COURSE CREDITS	5

Course Outcomes: After completion of this course, student will be able to:

- 1 Understanding of Exploratory data analysis & data frame.
- 2 Handling of data & data cleaning before analysis.
- 3 Able to understand data grouping and aggregation.
- 4 Visualization of various graphs to get insight of the data.
- 5 Implementation of applicability of data analysis in real life.

Pre-requisite of course: Knowledge of Programming Languages, statistical concepts.

Teaching and Examination Scheme

Theory Hours	Tutorial Hours	Practical Hours	ESE	IA	CSE	Viva	Term Work
3	0	4	50	30	20	25	25

Contents : Unit	Topics	Contact Hours
1	Exploratory data analysis: Introduction of EDA, Types of EDA: Univariate, bi-variate and multi variate analysis, Creating Data Frames from a Dict of Series or Dicts, Creating Data Frames from a Dict of Ndarrays/Lists, Creating Data Frames from a structured or Record Array, Creating Data Frames from a List of Dict, Creating Data Frames from a Dict of Tuples, Selecting- Adding-and Deleting Data Frame Columns, Reading CSV file with pandas, Reading CSV file with pandas	15
2	Data Gathering and Cleaning: Cleaning Data – Checking of Missing value, Handling the Missing values, Reading and cleaning CSV Data, Merging, integrating Data	10
3	Data Analysis Statistical Analysis-1, Statistical Analysis-2, Statistical Analysis-3, Data Grouping, Iterating through Groups, Aggregation, , Transformations, Filteration	10
4	Data Visualization DirectPlotting, LinePlot, BarPlot, PieChart, Histogram, ScatterPlot, Matplotlib Plot, Line plot, Bar chart, Histogram Plot, Scatter Pie Chart	10

Contents : Unit	Topics	Contact Hours
5	Case study case study-1, case study-2	5
Total Hours		50

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
1	<p>Practical</p> <p>1. Write a Pandas program to create a dataframe from a dictionary and display it. 2. Write a Pandas program to create and display a DataFrame from a specified dictionary data which has the index labels. 3. Write a Pandas program to display a summary of the basic information about a specified DataFrame and its data. 4. Write a Pandas program to get the first 3 rows of a given DataFrame. 5. Write a Pandas program to get list from DataFrame column headers. 6. Write a Pandas program to write a DataFrame to CSV file using tab separator. 7. Write a program to create dataframe using dictionary of lists. 8. Write a program to create dataframe by passing lists variables to dictionary. 9. Write a program to create pandas Dataframe from the dictionary of dictionaries. 10. Write a program to creating a Pandas dataframe using list of tuples., Data Cleaning programs : 1. Write a program to clean the empty cells. 2. Write a program to replace the empty cells. 3. Write a program to convert into correct format. 4. Write a program to remove the rows. 5. Write a program to fix the wrong data using pandas. 6. Write a program that replaces the wrong data. 7. Write a program that discovers and removes the duplicate data. 8. Write a Pandas program to join the two given dataframes along rows and merge with another dataframe along the common column id.9. Write a program that reads data from csv file and cleans it. 10. Write a Pandas program to join the two given dataframes along rows and merge with another dataframe along the common column id., Write a Python script to read the data in an Excel file named movies.xlsx and save this data in a dataframe called mov. Perform the following steps: 1. Read the contents of the second sheet that is named 2000s in the Excel file (movies.xlsx)and store this content in a data frame called Second_sheet. 2. Write the code needed to show the first seven rows from the data frame Second_sheet using an appropriate method. 3. Write the code needed to show the last five rows using an appropriate method. 4. Use a suitable command to show only one column that is named Budget. 5. Use a suitable command to show the total rows in the first sheet that is called 2000s, 1. Create a data frame called df from the following tabular data dictionary that has these index labels:['a','b','c','d','e','f','g', 'h','i', 'j']. 2. Display a summary of the dataframe's basic information. 3. Return the first three rows of the dataframe df. 4. Select just the animal and age columns from the dataframe df. 5. Count the visit priority per animal. 6. Find the mean of the animals' ages. 7. Display a summary of the data set. 8. Return the first three rows of the</p>	30

<p>dataframe df. 9. Extract first and last column in one table. 10. Observe output of ndim(), shape(), Create 500 random temperature readings for six cities over a season and then plot the generated data using Matplotlib. 2. Load the well-known Iris data set, which lists measurements of petals and sepals of three iris species. Then plot the correlations between each pair using the .pairplot() method. 3. Load the well-known Tips data set, which shows the number of tips received by restaurant staff based on various indicator data; then plot the percentage of tips per bill according to staff gender. 4. Load the well-known Tips data set, which shows the number of tips received by restaurant staff based on various indicator data; then implement the factor plots to visualize the total bill per day according to staff gender. 5. 6. Reimplement the above exercise using the Seaborn joint plot distributions. 7. Python program of Barplot with all parameters of a sample data. 8. Python program of Pie-chart with all parameters of a sample data. 9. Python program of Histogram with all parameters of a sample data. 10. Python program of Line Plot with all parameters of a sample data. 11. Python program of Scatter plot with all parameters of a sample data.</p>	
Total Hours	30

Textbook :

- 1 Data Analysis and visualization using Python, Dr. Ossama Embarak, Apress Berkeley, CA, 2018

References:

- 1 Introduction to Computation and Programming Using Python, Introduction to Computation and Programming Using Python, John V Guttag., Prentice Hall of India, 2ndE
- 2 Core Python Programming, Core Python Programming, R Nageswara Rao, Dreamtech Press, 2ndE
- 3 Core Python Applications Programming, Core Python Applications Programming, Wesley J Chun, Pearson, 3rdE

Suggested Theory Distribution:

The suggested theory distribution as per Bloom's taxonomy is as follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process

Distribution of Theory for course delivery and evaluation					
Remember / Knowledge	Understand	Apply	Analyze	Evaluate	Higher order Thinking
10.00	10.00	30.00	20.00	20.00	10.00

Instructional Method:

- 1 Demo
- 2 Presentation
- 3 Videos

Supplementary Resources:

- 1 "Python Programming", http://en.wikibooks.org/wiki/Python_Programming
- 2 "The Python Tutorial", <http://docs.python.org/release/3.0.1/tutorial/>