

**Common Pool of Generic Electives (GE) Courses
Offered by Department of Computer Sciences
Category-IV**

GENERIC ELECTIVES (GE-2a): Data Analysis and Visualization

Credit distribution, Eligibility and Pre-requisites of the Course

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre- requisite of the course
		Lecture	Tutorial	Practical/ Practice		
GE2a Data Analysis and Visualization using Python	4	3	0	1	Class XII pass with Mathematics	knowledge of Python

Learning Objectives

This course is designed to introduce the students to real-world data analysis problems, their analysis and interpretation of results in the field of exploratory data science using Python.

Learning outcomes

On successful completion of the course, students will be able to:

- Apply descriptive statistics to obtain a deterministic view of data
- Apply basic and advanced level statistical function on data
- Perform data handling using Numpy arrays
- Do data cleaning and transformation before extracting useful information
- Visualize data for ease of understanding the revealed information

SYLLABUS OF GE-2a

UNIT – I & II (09 Hours)

Introduction to basic statistics and analysis: Fundamentals of Data Analysis, Statistical foundations for Data Analysis, Types of data, Descriptive Statistics, Python Libraries: NumPy, Pandas, Matplotlib

Array manipulation using NumPy: NumPy array: Creating NumPy arrays, various data types of NumPy arrays

UNIT – I & II (09 Hours)

Introduction to basic statistics and analysis: contd..

Correlation and covariance, Linear Regression, Statistical Hypothesis Generation and Testing

Unit 2 Array manipulation using Numpy: contd..

Indexing and slicing, swapping axes, transposing arrays, data processing using Numpy arrays

UNIT – III (15 Hours)

Data Manipulation using Pandas: Data Structures in Pandas: Series, Data Frame, Index objects, loading data into Panda's data frame, Working with Data Frames: Arithmetics, Statistics, Binning, Indexing, Reindexing, Filtering, Handling missing data, Hierarchical indexing, Data wrangling: Data cleaning, transforming, merging and reshaping

UNIT – IV (12 Hours)

Plotting and Visualization: Using Matplotlib to plot data: figures, subplots, markings, color and line styles, labels and legends, Plotting functions in Pandas: Lines, bar, Scatter plots, histograms, stacked bars, Heatmap

Practical component (if any) – 30 Hours

Use data set of your choice from Open Data Portal ([https:// data.gov.in/](https://data.gov.in/), UCI repository) or load from scikit, seaborn library for the following exercises to practice the concepts learnt.

1. Load a Pandas data frame with a selected dataset. Identify and count the missing values in a data frame. Clean the data after removing noise as follows
 - a. Drop duplicate rows.
 - b. Detect the outliers and remove the rows having outliers
 - c. Identify the most correlated positively correlated attributes and negatively correlated attributes
2. Import iris data using sklearn library or (Download IRIS data from: <https://archive.ics.uci.edu/ml/datasets/iris> or import it from sklearn.datasets)
 - a. Compute mean, mode, median, standard deviation, confidence interval and standard error for each feature
 - b. Compute correlation coefficients between each pair of features and plot heatmap
 - c. Find covariance between length of sepal and petal
 - d. Build contingency table for class feature
3. Load Titanic data from sklearn library , plot the following with proper legend and axis labels:
 - a. Plot bar chart to show the frequency of survivors and non-survivors for male and female passengers separately
 - b. Draw a scatter plot for any two selected features
 - c. Compare density distribution for features age and passenger fare

- d. Use a pair plot to show pairwise bivariate distribution
4. Using Titanic dataset, do the following
 - a. Find total number of passengers with age less than 30
 - b. Find total fare paid by passengers of first class
 - c. Compare number of survivors of each passenger class

Project students are encouraged to work on a good dataset in consultation with their faculty and apply the concepts learned in the course.

Essential/recommended readings

1. McKinney W. *Python for Data Analysis: Data Wrangling with Pandas, NumPy and IPython*. 2nd edition, O'Reilly Media, 2018.
2. Molin S. *Hands-On Data Analysis with Pandas*, Packt Publishing, 2019.
3. Gupta S.C., Kapoor V.K., *Fundamentals of Mathematical Statistics*, Sultan Chand & Sons, 2020.

Suggestive readings

- (i) Chen D. Y, *Pandas for Everyone: Python Data Analysis*, Pearson, 2018.
- (ii) Miller J.D. *Statistics for Data Science*, Packt Publishing, 2017.

GENERIC ELECTIVES (GE-2b): Data Analysis and Visualization using Spreadsheet

Credit distribution, Eligibility and Pre-requisites of the Course

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course
		Lecture	Tutorial	Practical/ Practice		
GE2b Data Analysis and Visualization using Spreadsheet	4	3	0	1	Class XII pass with Mathematics	Nil

Learning Objectives

This course is designed to inculcate statistical thinking about data to the students who have studied Mathematics up to Class 10th ONLY. It gently introduces the students to basic statistics, and builds gradually to cover advanced functionalities for data analysis in spreadsheets. The objective is to enhance the knowledge of statistics and enable students to make sense of data by analyzing and visualizing it using spreadsheets, interpreting the results and gaining insights.

Learning outcomes

On successful completion of the course, students will be able to:

- Analyze and visualize data using spreadsheets
- Apply basic and advanced level statistical functions in spreadsheets
- Gain practical, hands-on experience of data analysis using spreadsheet

SYLLABUS OF GE-2b

UNIT – I (09 Hours)

Introduction to Basic Statistics

Fundamentals of Data Analysis, Statistical foundations for Data Analysis, Types of data, Descriptive Statistics, Correlation and covariance, Linear Regression.

UNIT – II (15 Hours)

Data Handling

Spreadsheet concepts, managing worksheets, formatting cells, entering data, Handling operators in formula, Cell referencing and naming of cells and cell ranges, Sorting, Multilayer sorting, Data validation, Find and Replace, Paste special, Filter and advanced filter, Formatting as table, Pivot tables, Formulae vs functions, Cell formulae vs Array formulae.

Mathematical functions, Statistical functions, Logical functions, Date and Time functions, Lookup and reference: Hlookup, and Vlookup, Index and Match functions, Text functions.

What-if-analysis: Goal-seek, Data tables, Scenario manager.

UNIT – III (12 Hours)

Data Analysis

Explore a data model: its content, and its structure, using the Power Pivot add-in. Learning DAX formula language. Create calculated fields and calculated measure for each cell, filter context for calculation, and explore several advanced DAX functions.

Cube formulas to retrieve data from data model.

UNIT – IV (09 Hours)

Data Visualization

Different types of charts including Pivot charts: Column, Line, Pie, Bar, Scatter charts. Fine tuning of charts: Chart Elements, Chart Styles, Chart Filters, Box Plot.

Practical component (if any) – 30 Hours

1. In a meeting of a marketing department of an organization it has been decided that price of selling an item is fixed at Rs. 40. It was resolved to increase the selling of more items and getting the profit of Rs. 50000/-. Use Goal Seek to find out how many items you will have to sell to meet your profit figure.
2. Create worksheet related to crop production of various crops in Indian states in last five years (wheat, rice, pulses, soya-bean, and cane-sugar etc).
 - i) Make a bar chart
 - ii) Make a pie chart
 - iii) Make a box plot
3. Study and perform the various DAX functions to analyse the data.
4. Create workbook related to sales of Business Company having various products in last four quarters for 10 sales persons.
 - i) Make a line graph to show the growth/decline in the sales
 - ii) Show the graph of each sales person sales
 - iii) Find the two sales persons done in last 2 quarters
 - iv) Find the sales persons consistent in last four quarters
 - v) Find the most popular product of the company and the current popular product of company.
5. Create a Pivot-table showing the Customer Names who placed orders with GSS during 2019-2022. For each customer, also show the total number of orders, Total Sales, and Total Profit. Add a Slicer or a Filter that can be used to show the information specifically for each Customer Segment. Use information from the Pivot-table to answer the following questions (Hint: Filter and sort the data in the Pivot-table to locate the answer):
 - i) Which small business customer had the highest sales?
 - ii) Which corporate customer placed the greatest number of orders in 2019-2022? How many orders were placed by the corporate customer?
 - iii) Which consumer customer was the most profitable one?
 - iv) What is the sales figure of the least profitable home office customer?
6. Consider the following worksheet: (enter 5 records)

FULL NAME	GRADE 1/2/3	BASIC SALARY	HRA	PF	GROSS	NET	VA	VA>HRA

HRA is calculated as follows:

Grade	HRA (% of basic)
1	40%
2	35%
3	30%

PF is 8% for all grades

VA is 15000, 10000, 7000 for Grades 1, 2 and 3.

Gross=Basic + HRA+VA

Net=Gross - PF

- i) Find max, min and average salary of employees in respective Grade.
 - ii) Count no. of people where VA>HRA
 - iii) Find out most frequently occurring grade.
 - iv) Extract records where employee name starts with "A" has HRA>10000
 - v) Print Grade wise report of all employees with subtotals of net salary and also grand totals.
 - vi) Use subtotal command.
 - vii) Extract records where Grade is 1 or 2 and salary is between 10000 and 20000 both inclusive.
7. Create workbook related to sales of Business Company having various product in last ten quarters for 20 sales persons. Perform the following on workbook:
- i) Create and modify a Pivot-table
 - ii) Apply Pivot-table styles and formatting
 - iii) Filter a Pivot-table
 - iv) Insert a slicer to filter a Pivot-table
 - v) Create a Pivot Chart
8. Create a PivotTable showing Total Sales breakdown by Region, Product Category, and Product Sub-Category. Use information from the PivotTable to answer the following questions:
- i) What was the Total Sales figure included in this data set?
 - ii) Which Product Category had the highest sales?

- iii) Which Region had the lowest sales?
iv) What was the Total Sales of Appliances in Delhi?

9. You are required to prepare a payroll statement in the given format making maximum use of cell referencing facility:

Code	Name	Category	Is HRA to be Paid	Basic	DP	DA	HRA	TA	CCA	Gross
1			Y							
2			N							
	Total									

Required:

- Basic salary (Allow any Basic salary in the range of Rs.10000-35000)
- DP is 50% of Basic Salary.
- DA (as a Percentage of Basic + DP) is more than 35000 then 40% of basic else 30% of basic.
- HRA is to be paid @ 40% of (Basic plus DP) to those whom HRA payable is yes.
- TA is to be paid @ Rs. 800 PM if Basic Salary is Less than Rs.12000, otherwise the TA is Rs. 1000 PM)
- CCA is to be paid @ Rs. 300 PM if Basic Salary is less than Rs.12000/- otherwise the CCA is Rs. 500 PM)
- Gross salary is the sum of Salary and all other allowances
- Deduction: a) GPF 10% of (Basic +DP) subject to a minimum of Rs.2000/- b) IT 10% of Gross Salary
- Net salary is Gross salary minus total deductions.

10. Consider the following worksheet for APS 1st year students:

S.No.	Name	Physics	Chem	Bio	Maths	CS	Total	%	Grade

1									
2									
3									
4									
5									

The value of Grade is calculated as follows:

If % ≥ 90	Grade A
If % ≥ 80 & < 90	Grade B
If % ≥ 70 & < 80	Grade C
If % ≥ 60 & < 70	Grade D

Otherwise, students will be declared fail.

- i) Calculate Grade using if function
- ii) Sort the data according to total marks
- iii) Apply filter to display the marks of the students having more than 65% marks.
- iv) Enter the S.No. of a student and find out the Grade of the student using VLOOKUP.
- v) Extract all records where name
 - a) Begins with "A"
 - b) Contains "A"
 - c) Ends with "A"

Essential/recommended readings

1. Gupta, S.P., *Elementary Statistical Methods*, Sultan Chand and Sons, New Delhi, 2017.

2. Goldmeier, J., *Advanced Excel Essentials*, Apress, 2014.
3. Slager, D., *Essential Excel 2016: A Step-by-Step Guide*, Apress, 2016.
4. Valerie M. Sue and Matthew T. Griffin, *Data Visualization and Presentation with Microsoft Office*, SAGE, 2016.
5. Schmuller, J., *Statistical Analysis with Excel for Dummies*, 4th edition., Wiley India Pvt Ltd., 2020.

Note: Examination scheme and mode shall be as prescribed by the Examination Branch, University of Delhi, from time to time.