

B.A. Prog. Sem II

A2: Data Interpretation and Visualization using Python (Admission 2022 onwards)

	TOPICS/UNITS	Chapter	Ref
Week 1 to 2	Unit 1 Introduction to basic statistics and analysis: Fundamentals of Data Analysis, Statistical foundations for Data Analysis, Types of data, Descriptive Statistics, Correlation and covariance, Linear Regression, Statistical Hypothesis Generation and Testing Python Libraries: NumPy, Pandas, Matplotlib	Ch1: pg 11-22, pg 29-31 pg 33-34, pg 37-p38	[2]
		Ch 1: 1.3 (pg 4-6)	[1]
Week 3 to 5	Unit 2 Array manipulation using Numpy: NumPy array: Creating NumPy arrays, various data types of NumPy arrays, Indexing and slicing, swapping axes, transposing arrays, data processing using Numpy arrays	Ch4:4.1, Usage of rand(), nrand() and randint() functions of NumPy	[1]
Week 6 to 9	Unit 3 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	Ch 5: 5.1, 5.2(upto pg 149), 5.3 Ch 6: 6.1 (pg 169-172, 175) Ch 7: 7.1, 7.2 (upto pg 206) Ch 8: 8.1 (pg 221-223), 8.2 (pg 227-231), 8.3 (pg 243-245)	[1]
Week 10 to 13	Unit 4 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	Ch 9: 9.1 (pg 253-264, 267), 9.2 (pg 269-282)	[1]
		Ch 5 : pg 281-282	[2]
Week 14 to 15	Data Aggregation and Group operations: Group by Mechanics, Data aggregation, General split-apply combine, Pivot tables and cross tabulation	Ch 10: 10.1(upto pg 293), 10.2, 10.3 (upto pg 303), 10.4	[1]

References:

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.

Suggested Practical List Data Interpretation and Visualization using Python

Note:

- Any platform for Python can be used for lab exercises.
 - Use a 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 dataframe with a selected dataset. Identify and count the missing values in a dataframe. 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)
 - i. Compute mean, mode, median, standard deviation, confidence interval and standard error for each feature
 - ii. Compute correlation coefficients between each pair of features and plot heatmap
 - iii. Find covariance between length of sepal and petal
 - iv. 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
 5. Download any dataset and do the following:
 - a. Count number of categorical and numeric features
 - b. Remove one correlated attribute (if any)
 - c. Display five-number summary of each attribute and show it visually