Roll No: | | | | | | | | | | | |

PRESIDENCY UNIVERSITY

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

Presidency University Act, 2013 of the Karnataka Act No. 41 of 2013 | Established under Section 2(f) of UGC Act, 1956
Approved by AICTE, New Delhi | Approved By BCI
Bengaluru

Mid Term Examinations - October 2026

Date: 05/10/2026 **Time:** 11:45 AM - 01:15 PM

Course Code: IST3410 **Course Name:** Data Handling and Visualization Techniques

Semester: Seventh Semester **Max. Marks:** 50 **Weightage:** 50%

CO - Levels	CO1	CO2
Marks	46	44

PART A: Answer Following Questions. 10 M

Q.N.	Questions	M	CO	BT
1	Data acquisition is the process of collecting raw data from various sources for further processing and analysis. (1M) Two methods: (i) Web scraping – extracting data from websites/HTML pages, (ii) APIs – programmatically fetching data from a service/server (e.g., REST APIs). (1M)	2	CO1	BT1
2	(i) NumPy – used for numerical/array computations. (1M) (ii) Pandas – used for handling tabular/structured data (DataFrames), cleaning and transformation. (1M)	2	CO1	BT1
3	Structured data is organized in a fixed, tabular format (rows/columns) with a defined schema, e.g., data in a relational database/CSV file. (1M) Unstructured data has no predefined format/schema, e.g., text documents, images, social media posts. (1M)	2	CO1	BT2
4	Hierarchical indexing (MultiIndex) allows multiple (two or more) index levels on an axis of a Pandas object, enabling representation of higher-dimensional data in a lower-dimensional (2D) structure such as a DataFrame. (2M)	2	CO2	BT1
5	Data profiling is the process of examining a dataset to understand its structure, content, quality, and statistics (e.g., missing values, data types, distributions). (1M) Tool: Pandas Profiling (or DataPrep). (1M)	2	CO2	BT1

PART B : Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
6	- Data Acquisition: collecting raw data from sources such as web scraping, APIs, files (2M) - Data Cleaning: handling missing values, removing duplicates/inconsistencies, AI-assisted anomaly detection (2M) - Data Transformation: converting/normalizing data, string operations, feature preparation (2M) - Data Storage: storing processed data in appropriate formats/databases for later use (2M) - Neat labeled diagram showing the pipeline flow: Acquisition → Cleaning → Transformation → Storage (2M)	10	CO1	BT2
7	Technique 1: Deletion (row/column removal) of missing data – explanation with when to use (2.5M) - Technique 2: Imputation (mean/median/mode or model-based) – explanation with example (2.5M) - AI-assisted anomaly detection: uses ML models to automatically flag outliers/inconsistent records (2.5M) - How AI-based imputation strategy selection improves accuracy/efficiency over manual methods (2.5M)	10	CO1	BT2

PART C: Answer Any 1 Following Questions. 10 M

ANSWER KEY

Q.N.	Questions	M	CO	BT
8	- Web scraping definition + example (e.g., using BeautifulSoup to extract table data from a webpage) (2.5M) - REST API definition + example (e.g., using the requests library to call a weather API and get a JSON response) (2.5M) - General steps: identify source → send request/parse HTML → extract required fields → clean & store data (3M) - Correct example/illustration of the acquired data (2M)	10	CO1	BT3
9	String operations: strip, replace, split, regex-based cleaning of text fields (2.5M) - Data preparation steps: type conversion, handling missing/duplicate values (2.5M) - Worked example on a sample raw dataset (e.g., cleaning inconsistent name/date fields) (3M) - Correct final prepared output illustrated (2M)	10	CO1	BT3

PART D: Answer Any 1 Following Questions. 10 M				
Q.N.	**Questions**	**M**	**CO**	**BT**
10	Q: Explain hierarchical indexing and data merging in Pandas with suitable examples. Answer / Scheme: - Hierarchical indexing (MultiIndex) concept explained with example (3M) - Data merging concept (merge/join operations – inner, outer, left, right) explained (3M) - Example of merging two DataFrames with output illustration (4M)	10	CO2	BT2
11	- Pivoting: converting long to wide format – example (3M) - Stacking/unstacking: converting between hierarchical row/column index – example (3M) - Illustration with sample dataset/output for at least one technique (4M)	10	CO2	BT2

PART E: Answer Any 1 Following Questions. 10 M				
Q.N.	**Questions**	**M**	**CO**	**BT**
12	- Memory optimization techniques: chunking, appropriate dtypes, downcasting (3M) - Role of Polars: columnar, multi-threaded processing for large data (2.5M) - Role of DuckDB: out-of-core SQL querying on large datasets without loading fully into memory (2.5M) - Applied example/workflow showing how these tools process a large dataset (2M)	10	CO2	BT3
13	- Case study description (e.g., movie/product recommendation) (2M) - Data wrangling steps applied: merging, reshaping, cleaning the case-study data (3M) - Predictive analysis / recommender approach explained (e.g., basics of collaborative filtering) (3M) - Expected outcome/insight illustrated (2M)	10	CO2	BT3

ANSWER KEY