## School of Computer Science

## M. Sc. (Information Technology) Programme

**Subject Code & Name:** MITME205-2 – Essentials of Data Science using Python

**Teaching And Evaluation Scheme:**

| Teaching Scheme | | | | Examination Scheme | | | |
|---|---|---|---|---|---|---|---|
| Credits | | | | Component Weightage | | | |
| | | | | CCE | | SEE | |
| L | T | P | Total | TH | PWE | TH | PWE |
| 2 | - | 2 | 4 | 25% | 25% | 25% | 25% |

| | |
|---|---|
| **Programme Name** | M. Sc. (Information Technology) |
| **Semester** | 2 |
| **Course Code** | MITME105-2 |
| **Course Title** | Essentials of Data Science using Python |
| **Course Content Type (Th. / Pr. )** | Th. + Pr. |
| **Course Credit** | 2+2 |
| **Sessions + Lab Per Week** | 2+4 |
| **Total Teaching/ Lab Hours** | 30+60 |
| *** 2 Lab = 1 session** | |

| **Learning Objectives** |
|---|
| Upon successful completion of this course, students will be able to:<br>• Develop advanced programming skills in Python with emphasis on data handling and analysis.<br>• Provide in-depth understanding of Pandas DataFrame and Series operations for structured data manipulation.<br>• Enable students to perform data acquisition, cleaning, transformation, and integration from multiple sources (CSV, JSON, Excel, text). |

- Introduce data preprocessing techniques, including handling missing values and normalization methods.
- Equip students with skills to perform statistical analysis and aggregation operations using Pandas.
- Develop competency in data visualization using Matplotlib for effective interpretation and presentation.
- Encourage problem-solving and analytical thinking through real-world data-driven applications.
- Promote writing efficient, modular, and maintainable Python code for data-centric tasks.

**Learning Outcomes**

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

- Apply advanced Python concepts to manipulate and analyse structured datasets using Pandas DataFrames and Series.
- Perform data input/output operations by reading and writing data in CSV, JSON, Excel, and text formats.
- Analyse datasets using data processing techniques, including sorting, filtering, grouping, and aggregation functions.
- Implement data cleaning and preprocessing techniques, including handling missing values, duplicates, and normalization methods.
- Interpret data using statistical measures and transformation techniques for meaningful insights.
- Design and generate data visualizations using Matplotlib to effectively communicate analytical results.

**CLO – PLO Mapping Matrix**

| CLO/ PLO | PLO1 | PLO2 | PLO3 | PLO4 | PLO5 | PLO6 | PLO7 | PLO8 | PLO9 |
|---|---|---|---|---|---|---|---|---|---|
| **CLO1** | 3 | 3 | 2 | 3 | 1 | – | – | 1 | 3 |
| **CLO2** | 2 | 2 | 3 | 3 | 1 | – | 1 | 1 | 3 |
| **CLO3** | 3 | 3 | 2 | 3 | 2 | – | 1 | 1 | 3 |
| **CLO4** | 2 | 3 | 3 | 3 | 2 | 1 | 1 | 1 | 3 |
| **CLO5** | 2 | 3 | 2 | 2 | 3 | – | 1 | 2 | 3 |
| **CLO6** | 2 | 2 | 3 | 3 | 2 | – | 3 | 2 | 3 |

**Detailed Contents**

| Unit | Topic/Sub Topic | Teaching Hours |
|---|---|---|
| I<br>Working with DataFrame ans Series using Pandas | Pandas Dataframe:<br>• DataFrame Creation<br>• Understanding Dataframe Index and Access<br>• Indexing and Selecting Data with Pandas and Slicing Operations | 15 |

| | | |
|---|---|---|
| | • Dataframe filtering , Merging, Joining and Concatenating<br>• Sorting and Searching in DataFrame<br>Pandas Series<br>• Series creation (From array and Sequences)<br>• Accessing elements and Binary Operations on Series<br>• Pandas Series Index() Methods | |
| **II**<br>Data Reading and Writing Operations | Data Reading and Writing Operations:<br>• Read/Writing CSV Files<br>• Import CSV files to Dataframe and Export DataFrame to a CSV file<br>• Read JSON File and Parsing JSON Dataset<br>• Exporting Pandas DataFrame to JSON File<br>• Working with Excel Files in Pandas<br>• Read Text Files with Pandas<br>• Text File to CSV using Pandas | 15 |
| **III**<br>Data Processing , Normalization and Analysis | Data Processing with Pandas.<br>• Loading Data in Pandas DataFrame<br>• Printing rows and column name of the Data<br>• Summarizing , Sorting and Merging of Data Frame<br>• Descriptive Statistical Measures of a DataFrame<br>• Missing Data Handing<br>Data Normalization in Pandas<br>• Steps for Data Normalization in Pandas<br>• Normalization Techniques in Pandas<br>• Normalization Techniques<br>  ○ Maximum Absolute Scaling<br>  ○ The min-max feature scaling<br>  ○ The z-score method<br>Data Analysis using Pandas<br>• dataframe.groupby() Method<br>• Apply Transformation Methods<br>• Using Filtration Methods<br>• Aggregation Functions(sum,min,max, mean,size,describe,first, last, count, std, var, sem)<br>• Grouping , Aggregating and Join Operations | 15 |
| **IV**<br>Data Cleaning and Visualization | Data Cleaning<br>• Handling Missing and Duplicate Data<br>• Handling various datatypes in Pandas<br>• Drop Empty Columns in Pandas<br>• String handling in Pandas using string methods<br>• Detect Mixed Data Types and Fix it<br>Data visualization with Matplotlib<br>• Creating basic plots (line, bar, scatter)<br>• Customizing plots (titles, labels, legends, colors) | 15 |

| | | |
|---|---|---|
| | • Saving plots to files<br>Data Wrangling<br>Introduction to Machine Learning<br>Mini Project | |

**Reference Books**


6. Thareja, R. (2023). *Python Programming* (4th ed.). Oxford University Press. ISBN: 978-9367257654
7. Rao, R. N. (2022). *Core Python Programming* (3rd ed.). Dreamtech Press. ISBN: 978-9390457151
8. Srivastava, D. (2023). *Ultimate Python Programming*. BPB Publications. ISBN: 978-9355516558
9. Dhamdhere, V. S., et al. (2024). *Python for Data Analysis*. RK Publication.
10. Khanna Publishing House. (2025). *Data Analysis & Visualization Using Python*. Khanna Publishing House. ISBN: 978-9355387967
11. McKinney, W. (2026). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Jupyter* (3rd ed.). O'Reilly Media. ISBN: 978-1098104030
12. Nelli, F. (2023). *Python Data Analytics: Data Analysis and Science Using Pandas, NumPy and Matplotlib* (3rd ed.). Apress. ISBN: 978-1484295311
13. VanderPlas, J. (2024). *Python Data Science Handbook* (2nd ed.). O'Reilly Media. ISBN: 978-1098121228
14. Harrison, M. (2023). *Effective Pandas: Patterns for Data Manipulation*. Treading on Python Press. ISBN: 978-1951412463
15. Bruce, P., Bruce, A., & Gedeck, P. (2020). *Practical Statistics for Data Scientists* (2nd ed.). O'Reilly Media. ISBN: 978-1492072942


**Web References:**

1. https://docs.python.org/3/
2. https://pandas.pydata.org/docs/
3. https://numpy.org/doc/
4. https://matplotlib.org/stable/users/index.html
5. https://scikit-learn.org/stable/user_guide.html
6. https://www.kaggle.com/learn
7. https://jupyter.org/documentation
8. https://realpython.com
9. https://www.w3schools.com/python/pandas/
10. https://www.geeksforgeeks.org/data-science-with-python-tutorial/

**L::** Lecture, **T::** Tutorial, **P::** Practical

**CCE::** Continuous & Comprehensive Evaluation

(CCE theory includes Mid Semester Examination, Assignment, MCQ Quizzes, Seminar, Reflective Notes, Class Participation, Case Analysis and Presentation, Slip Tests (Announced/ Surpirsed), Attendance etc. or any combination of these)

**PWE:** Practical Work Examination

(PWE includes Laboratory Practical Work, Project Work, Viva simulation exercise work etc.)

**SEE::** Semester End Evaluation

**Sample Practical Exercises:**

**Unit-I**

1. Create a DataFrame using a Python dictionary.
2. Create a DataFrame using a list of lists.
3. Create a DataFrame using a NumPy array.
4. Create a DataFrame from a list of dictionaries.
5. Display DataFrame dimensions using shape.
6. Display DataFrame column names.
7. Display DataFrame index values.
8. Use head() and tail() methods.
9. Display DataFrame information using info().
10. Access a single column and multiple columns from a DataFrame.
11. Access rows using loc[] and iloc[].
12. Select a range of rows using slicing , specific rows and columns.
13. Retrieve scalar values using at[] and iat[].
14. Change index values of a DataFrame.
15. Reset DataFrame index.
16. Filter records based on a single condition and multiple conditions.
17. Filter records using logical AND and OR.
18. Filter records using isin() between().
19. Find records containing specific strings.
20. Exclude records matching a condition.
21. Apply conditional filtering on numerical data.
22. Sort DataFrame by one column and by multiple columns.
23. Sort in ascending order or descending order.
24. Search records using conditional expressions.
25. Find maximum and minimum values.
26. Locate duplicate values and identify unique values in a column.
27. Count occurrences of unique values.
28. Search for missing values in a DataFrame.

**Unit-II**

1. Read a CSV file into DataFrame.
2. Display first 10 records from CSV.
3. Write DataFrame to CSV file.
4. Export selected columns to CSV.
5. Import CSV with custom delimiter.
6. Read large CSV file in chunks.
7. Append records to existing CSV.
8. Read CSV without header.
9. Assign custom headers while importing CSV.

10. Export filtered data to CSV.
11. Read JSON file.
12. Convert JSON to DataFrame.
13. Parse nested JSON structure.
14. Export DataFrame to JSON.
15. Convert CSV data into JSON format.
16. Convert JSON into CSV.
17. Read JSON from URL source.
18. Handle malformed JSON data.
19. Extract specific fields from JSON.
20. Read Excel workbook.
21. Read specific worksheet and multiple worksheets.
22. Export DataFrame to Excel.
23. Write multiple DataFrames into multiple sheets.
24. Append new sheet to existing workbook.
25. Read Excel using custom column selection.
26. Export filtered data to Excel.
27. Create summary report in Excel.
28. Convert Excel to CSV.
29. Read text file using Pandas.
30. Convert text file into DataFrame.
31. Clean text data and export.
32. Convert text file to CSV.
33. Count words from text file.
34. Extract lines containing specific keywords.
35. Merge multiple text files.
36. Analyze text file statistics.
37. Save processed text data.
38. Mini project on multi-format file conversion.

**Unit-III**

1. Load dataset into DataFrame.
2. Display rows and columns.
3. Generate dataset summary.
4. Display statistical information.
5. Sort dataset by multiple attributes.
6. Merge two DataFrames.
7. Concatenate multiple DataFrames.
8. Rename columns.
9. Change data types of columns.
10. Drop unnecessary columns.
11. Detect missing values and count missing values.
12. Replace missing values with zero, mean or median.
13. Forward fill missing values and Backward fill missing values.
14. Remove rows and columns with missing values.
15. Apply Max Absolute Scaling and Apply Min-Max Scaling.
16. Apply Z-score normalization.

17. Normalize selected columns and complete dataset.
18. Visualize normalized data.
19. Perform grouping using groupby().
20. Calculate sum and mean using aggregation.
21. Calculate maximum and minimum values.
22. Generate descriptive statistics.
23. Perform multi-level grouping.
24. Join grouped datasets.

**Unit-IV:**

1. Detect and Remove duplicate records.
2. Identify mixed data types.
3. Convert mixed data types into standard format.
4. Drop empty columns.
5. Replace null values.
6. Detect outliers using statistical methods.
7. Clean inconsistent text data.
8. Remove unwanted spaces.
9. Standardize date formats.
10. Convert categorical values into standard format.
11. Create data cleaning pipeline.
12. Generate data quality report.
13. Convert text to uppercase or lowercase.
14. Replace substrings.
15. Split strings into columns.
16. Extract specific patterns using regex.
17. Count occurrences of words.
18. Remove special characters.
19. Trim whitespace.
20. Concatenate string columns.
21. Create Line Chart, Multiple Line Chart, Bar Chart, Horizontal Bar Chart, Scatter Plot, Pie Chart and Histogram.
22. Add Titles and Labels, Legends, customized colours and Styles.
23. Save chart as PNG or as PDF.
24. Compare multiple charts.