Part – A (8 × 4 = 32 Marks)

Note: Answer any eight questions

1. Write simple program using pickle module.
   Write any six methods of os.path module.
   How do you read a JSON file line by line in Python?
   Explain basic HTML tags with its attributes.
5. Write any four method of regular expression with syntax.
6. List any five meta characters in regular expression with example.
7. How to create table using MYSQL explain with syntax and example?
8. What are the advantages of Numpy?
9. Write about any four array creation functions in Numpy.
10. Write a program in python to create a series of first five even numbers.
11. How to create series from dictionaries?
12. Define DataFrame in Panda.

Part – B (4 × 12 = 48 Marks)

Note: Answer all the questions

13. (a) Define data analysis sequence. What are the different types of data in data science?

    OR

    (b) Explain how to create CSV file. Explain how to extract data from a CSV file in python using program?

    (a) Explain text processing with normalization.

    OR

    (b) Explain different regex functions with suitable program.

15. (a) Explain how to read data from database table using fetchone( ) and fetchall( ) function using example program.

    OR

    (b) Explain following operations with example.

    (i) reshape

    (ii) Concatenation

    (iii) Transpose

    (iv) Any three arithmetic operations

16. (a) Explain how to add a column into a Panda DataFrame using program?

    OR

    (b) Explain various functions of Pandas with example.

OSMANIA UNIVERSITY

FACULTY OF SCIENCE

B.Sc. (CBCS) III-Semester Examination

December-2023/January-2024

Subject: Data Science

Paper-III: Data Engineering with Python

Time: 3 Hours   Max. Marks: 80

Part – A (8 × 4 = 32 Marks)

Note: Answer any eight questions.

1. How do you read a JSON file line by line in Python?
2. Discuss how to perform read and write operations on binary file.
3. What is the difference object's writerow( ) and writerows( ) functions in CSV.
4. Discuss about various HTML tags with its attributes.
5. Explain stemming technique in NLP.
6. Write a short note on glob module.
   Explain how to create database table in MySQL with example.
   Discuss about MongoDB.
9. Discuss about applications of Numpy.
10. Define Pandas and its applications.
11. Explain how to create series from scalar values.
12. Define matplotlib in python.

Part – B (4 × 12 = 48 Marks)

Note: Answer all the questions.

13. (a) What are the steps in a data acquisition pipeline? Explain with neat diagram.

    OR

    (b) Explain JSON package. Discuss how JSON package works in Python and write short note on serializing & deserializing methods in JSON.

    (a) Explain Tokenization Lemmatization techniques in NLP with simple programs.

    OR

    (b) Define regular expression and discuss about special sequence characters with proper example.

15. (a) How to create database & table using MySQL. Explain how to perform insert, update and delete elements in the database with syntax and example.

    OR

    (b) Explain following with example.

    (i) Reshape (ii) Concatenation (iii) Transpose (iv) Any three arithmetic operations

16. (a) Define data frame in Pandas. Discuss about various attributes of data frames with example.

    OR

    (b) Define data visualization. Explain any different photos used in matplotlib with simple example.