KUMARAGURU
college of technology
character is life

Register Number:……..…

M.TECH DEGREE EXAMINATIONS: DEC 2022

(Regulation 2018)

First Semester

DATA SCIENCE

P18ITI1204: Data Science and Analytics with Python

COURSE OUTCOMES

CO1: Explain the roles and stages of data science projects

CO2: Describe the data structures provided by numpy library for arrays and vectorized computation

CO3: Explain data structures provided by pandas library for data analysis

CO4: Perform data wrangling, cleaning and transformation using python

CO5: Use matplot lib for plotting and visualizing the datasets

CO6: Demonstrate data aggregation and time series analysis using python programming Language

Time: Three Hours **Maximum Marks: 100**

Answer all the Questions:-

PART A (10 x 1 = 10 Marks)

1. Find the correct sequence of data science process: CO1 [K_1]

   i. Data Acquisition ii. Business Understanding

   iii. Modeling iv. Deployment

   a) ii, i, iii, iv b) i, ii, iii, iv

   c) iii, i, ii, iv d) ii, iv, iii, i

2. What will be the minimum number of arguments require to pass in pandas series? CO3 [K_1]

   a) 1 b) 2

   c) 3 d) 4

3. Name a Python built-in string method which uses string as delimiter for concatenating a sequence of other strings. CO4 [K_1]

   a) concat b) index

   c) join d) merge

4. Matching type item with multiple choice code CO3 [K_2]

List I	List II
A. Panda's data structure	i. holding the axis label and meta

	data
B. panda's index	ii. Series
C. Panda's reindex method	iii. multiple index levels on an axis
D. Panda's Hierarchical indexing	iv. backfill

	A	B	C	D
a)	i	ii	iii	iv
b)	ii	i	iv	iii
c)	iv	iii	ii	i
d)	i	iv	ii	iii

5. A ---------- is a data summarization tool frequently found in spreadsheet programs and other data analysis software. CO4 [K_L]

a) Sort table b) Index table

c) Data table d) pivot table

6. In Pandas, ---------- connects rows in DataFrames based on one or more keys CO3 [K_L]

a) pandas.merge b) pandas.sort

c) pandas.align d) pandas.key

7. Assertion (A): Resampling refers to the process of converting a time series from one frequency to another. CO5 [K_L]

Reason (R): All resampling falls into either upsampling or downsampling.

a) Both A and R are Individually true and R is the correct explanation of A b) Both A and R are Individually true but R is not the correct explanation of A

c) A is true but R is false d) A is false but R is true

8. Timestamps can be converted to periods using the ---------- method CO6 [K_L]

a) change_period b) to_period

c) convert_period d) timestamp_period

9. In Pandas, a Data frame object is ---------- mutable. CO4 [K_L]

a) value b) size

c) frame d) count

10. Which of the following are not array creation functions in NumPy? CO2 [K_1]

i. array ii. asarray iii. aqueue iv. acreate

a) i and ii b) i and iv

c) iii and iv d) ii and iv

PART B (10 x 2 = 20 Marks)

11.	List the significant advantages of data science.	CO1	[K_1]
12.	Identify the data types supported in NumPy.	CO2	[K_2]
13.	Tell the uses of the following NumPy random function: shuffle, permutation	CO3	[K_1]
14.	Infer the need of 'Series' data structure in Pandas.	CO3	[K_2]
15.	How do you read excel file in Pandas?	CO4	[K_1]
16.	What are the fundamental operations for rearranging tabular data in Pandas?	CO4	[K_1]
17.	Show the use of plot function in matplotlib.	CO5	[K_2]
18.	List some Series.plot function method arguments in Pandas.	CO5	[K_1]
19.	What is the use of apply function in Group-wise operations?	CO6	[K_1]
20.	What are the types used for datetime module?	CO6	[K_1]

PART C (6 x 5 = 30 Marks)

21.	Outline the stages of data science project life cycle.	CO1	[K_2]
22.	Explain the various array set operations in NumPy.	CO2	[K_2]
23.	List the commonly used linear algebra functions in NumPy.	CO3	[K_1]
24.	Explain the use of HDF5 in data file format.	CO4	[K_2]
25.	Assess the various plotting functions in Pandas.	CO5	[K_3]
26.	Summarize the various optimized Groupby methods with their uses.	CO6	[K_2]

Answer any FOUR Questions

PART D (4 x 10 = 40 Marks)

27.	Compile the various data processing tasks performed in NumPy.	CO1	[K_3]
28.	Analyze how to handle missing data in Pandas.	CO2	[K_3]
29.	Discuss the data transformation operations with examples.	CO3	[K_3]
30.	Discuss the Python Visualization Tool Ecosystem in detail.	CO4	[K_3]
31.	Illustrate how to handle time zone in Python.	CO5	[K_3]

****