

CBCS SCHEME

USN

--	--	--	--	--	--	--	--	--	--

BAI515E

Fifth Semester B.E./B.Tech. Degree Examination, June/July 2026 Exploratory Data Analysis

Time: 3 hrs.

Max. Marks: 100

*Note: 1. Answer any FIVE full questions, choosing ONE full question from each module.
2. M : Marks , L: Bloom's level , C: Course outcomes.*



Module – 1				M	L	C
Q.1	a.	Interpret the purpose of the np.sort () function and np.argsort () function in NumPy with an example program.	8	L3	CO1	
	b.	Explain how array indexing works in multidimensional arrays. Discuss array slicing and its use in NumPy. Demonstrate a program to extract subarrays using slicing.	8	L3	CO1	
	c.	What are the advantages of using NumPy arrays over python lists?	4	L2	CO1	
OR						
Q.2	a.	Explain how to create an array from a python list using NumPy. What is the use of the dtype attribute in NumPy arrays? Interpret an example program.	8	L3	CO1	
	b.	Define array concatenation in NumPy. How is it different from splitting arrays? Write a python program that demonstrates array reshaping and splitting operations using NumPy.	8	L3	CO1	
	c.	Describe the difference between reshaping and slicing of NumPy arrays.	4	L2	CO1	
Module – 2						
Q.3	a.	What is the purpose of the pivot-table () function and how does it differ from groupby ()? How can you create multi level pivot tables in Pandas and why are they useful for data aggregation?	7	L2	CO2	
	b.	With an programming example, explain the following isnull (), notnull (), dropna (), fillna () method in Pandas.	8	L2	CO2	
	c.	Execute how the aggfunc parameter in the pivot-table () function works and how it allows for custom aggregation.	5	L3	CO2	
OR						
Q.4	a.	Describe series and Dataframe in Pandas. List and explain three different methods to create a DataFrame in Pandas.	7	L2	CO2	

	b.	What is hierarchical indexing, and what are its advantages? Interpret the methods of creating a MultiIndex. Using <code>pd.multiindex.from_tuples ()</code> and <code>pd.MultiIndex.from_Product ()</code> .	8	L3	CO2
	c.	Explain the function of the <code>dropna ()</code> method and its parameters for handling missing data.	5	L2	CO2
Module – 3					
Q.5	a.	Demonstrate the following with programming example, <code>Strlen ()</code> , <code>str.split ()</code> , <code>str.lower ()</code>	6	L3	CO3
	b.	Discuss the purpose of the <code>eval</code> and <code>query</code> methods in Pandas. How do they improve performance in comparison to traditional operation?	7	L2	CO3
	c.	Explain the <code>rolling ()</code> method in Pandas for calculating rolling statistics. How can it be combined with aggregation functions?	7	L2	CO3
OR					
Q.6	a.	Demonstrate three Pandas str methods and mirror python's string methods and briefly explain their functions.	6	L3	CO3
	b.	What types of operations are supported by the <code>eval</code> function in Pandas? Provide examples of arithmetic, comparison and bitwise operations.	7	L2	CO3
	c.	How does the <code>query</code> method, simplify filtering rows in a DataFrame? Provide examples using column conditions and local variables.	7	L3	CO3
Module – 4					
Q.7	a.	Draw the Scatter. Plot to illustrate the relationship between number of friends and the number of minutes spend on every day. Friends = [70, 65, 72, 63, 71, 64, 60, 64, 67] Minutes = [175, 170, 205, 120, 220, 130, 105, 145, 190] Labels = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i']	6	L3	CO4
	b.	Implement the process of creating a basic line plot in Matplotlib using the object-oriented interface. How does the <code>plt.subplots ()</code> function help in creating multiple plots?	8	L2	CO4
	c.	Describe the differences between the MATLAB-style and object-oriented interfaces in Matplotlib. Demonstrate examples of both approaches.	6	L3	CO4

OR

Q.8	a.	Interpret the differences between plt.scatter () and plt.plot () for creating scatter plots. When should you use plt.scatter ()?	8	L3	CO4
	b.	Explain in detail how seaborn plots useful for statistical Data Exploration.	6	L3	CO4
	c.	How can you control axis limits and labels in Matplotlib? Explain the differences between plt.xlim () and plt.ylim () with examples.	6	L2	CO4

Module – 5

Q.9	a.	How does the Iris dataset demonstrate data representation in scikit-Learn? What are the features matrix and target array?	8	L3	CO5
	b.	Explain clustering and dimensionality reduction as examples of unsupervised learning.	7	L2	CO5
	c.	Operate the purpose of GridSearch CV in Scikit-Learn. How is it used to optimize model hyperparameters?	5	L3	CO5

OR

Q.10	a.	What are classification and regression in supervised learning? How do they differ explain with examples.	8	L2	CO5
	b.	What are the steps involved in the Scikit-Learn. Estimator API? Use and describe each step.	7	L3	CO5
	c.	What is Machine Learning? Differentiate machine learning from traditional programming.	5	L3	CO5

* * * * *