

M.Sc.(DS).I/11.23.001 Reg.No.

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

B

**M.SC. COMPUTER SCIENCE WITH SPECIALIZATION IN DATA SCIENCE
FIRST SEMESTER EXAMINATION NOVEMBER 2023**

22-359-0101 STATISTICAL FOUNDATIONS FOR DATA SCIENCE

(Regular)

Time: 3 Hours

Maximum Marks :50

Each question carries 10 Marks

No		QUESTIONS	MARKS	CO	BL	PI																																												
1	a.	<p>Explain the term Univariate Analysis?</p> <p>Given a dataset, perform the following univariate analysis on the variable <i>Household Size</i>:</p> <table border="1" style="margin: 10px auto; border-collapse: collapse; text-align: center;"> <thead> <tr> <th>Household ID</th> <th>Household Size</th> <th>Annual Income</th> <th>Number of Pets</th> </tr> </thead> <tbody> <tr><td>1</td><td>2</td><td>\$37,000</td><td>0</td></tr> <tr><td>2</td><td>4</td><td>\$49,000</td><td>0</td></tr> <tr><td>3</td><td>4</td><td>\$58,000</td><td>1</td></tr> <tr><td>4</td><td>1</td><td>\$68,000</td><td>3</td></tr> <tr><td>5</td><td>3</td><td>\$61,000</td><td>2</td></tr> <tr><td>6</td><td>5</td><td>\$64,000</td><td>2</td></tr> <tr><td>7</td><td>6</td><td>\$79,000</td><td>1</td></tr> <tr><td>8</td><td>4</td><td>\$89,000</td><td>1</td></tr> <tr><td>9</td><td>7</td><td>\$104,000</td><td>1</td></tr> <tr><td>10</td><td>2</td><td>\$95,000</td><td>0</td></tr> </tbody> </table>	Household ID	Household Size	Annual Income	Number of Pets	1	2	\$37,000	0	2	4	\$49,000	0	3	4	\$58,000	1	4	1	\$68,000	3	5	3	\$61,000	2	6	5	\$64,000	2	7	6	\$79,000	1	8	4	\$89,000	1	9	7	\$104,000	1	10	2	\$95,000	0	10	CO1	L3	2.8.4
Household ID	Household Size	Annual Income	Number of Pets																																															
1	2	\$37,000	0																																															
2	4	\$49,000	0																																															
3	4	\$58,000	1																																															
4	1	\$68,000	3																																															
5	3	\$61,000	2																																															
6	5	\$64,000	2																																															
7	6	\$79,000	1																																															
8	4	\$89,000	1																																															
9	7	\$104,000	1																																															
10	2	\$95,000	0																																															
	b.	Describe the variable <i>Household Size</i> using summary statistics.																																																
	c.	Create a frequency distribution table for the variable <i>Household Size</i> .																																																
	d.	Visualize the distribution of values for the variable <i>Household Size</i> using <ol style="list-style-type: none"> i. Box plot ii. Histogram 																																																
OR																																																		
2	a.	Define Covariance and identify the interpretation of covariance values.	4	CO1	L2	2.8.4																																												
	b.	The table below shows the Sample data of two new company stocks between 2015 and 2020. Find Sample Covariance of the below data and identify the relation between stock values of the company's X and Y.	6	CO1	L3	2.8.4																																												

Year	Company X Stock Value (\$)	Company Y Stock Value (\$)
2015	1245	100
2016	1415	123
2017	1312	129
2018	1427	143
2019	1510	150
2020	1590	197

3	a.	The average hourly wage of a sample of 150 workers in a plant 'A' was Rs. 2.56 with a standard deviation of Rs. 1.08. The average wage of a sample of 200 workers in plant 'B' was Rs. 2.87 with a standard deviation of Rs. 1.28. Can an applicant safely assume that the hourly wages paid by plant 'B' are higher than those paid by plant 'A' with 1% level of significance?	6	CO2	L3	1.7.1
	b.	Suppose the null hypothesis, H_0 , is: The victim of an automobile accident is alive when he arrives at the emergency room of a hospital. (i) State Type I and Type II errors for the above hypothesis. (ii) Which type of error has the greater consequence, Type I or Type II? Explain why?	4	CO2	L3	1.7.1

OR

4	a.	A random sample of 20 graduate college students and 18 undergraduate college students indicated these results concerning the amount of time spent in volunteer service per week. <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th></th> <th>Graduate</th> <th>Undergraduate</th> </tr> </thead> <tbody> <tr> <td>Sample Mean</td> <td>3.8 hrs/week</td> <td>2.5 hrs/week</td> </tr> <tr> <td>Sample Variance</td> <td>3.5 hrs/week</td> <td>2.2 hrs/week</td> </tr> <tr> <td>Sample Size</td> <td>20</td> <td>18</td> </tr> </tbody> </table> At $\alpha = 0.01$ level of significance, is there sufficient evidence to conclude that graduate students have a higher standard deviation of the number of volunteer hours per week compared to undergraduate students? Assume that number of volunteer hours per week is normally distributed.		Graduate	Undergraduate	Sample Mean	3.8 hrs/week	2.5 hrs/week	Sample Variance	3.5 hrs/week	2.2 hrs/week	Sample Size	20	18	6	CO2	L3	1.7.1
		Graduate	Undergraduate															
Sample Mean	3.8 hrs/week	2.5 hrs/week																
Sample Variance	3.5 hrs/week	2.2 hrs/week																
Sample Size	20	18																
b.	How do you reduce the risk of making (i) Type I Error (ii) Type II Error	4	CO2	L2	1.7.1													

5	a.	If X is a random variable following Binomial distribution with mean 2.4 and variance 1.44. Find $P(X \geq 5)$?	3	CO3	L3	1.7.1
	b.	If X is a random variable which follows Poisson distribution. Given that $P(X = 2) = P(X = 3)$. Find $P(X = 4)$?	3	CO3	L3	1.7.1
	c.	If the probability that an individual suffers a bad reaction from the injection of a serum is 1%, find the probability that out of 800 individuals: (i) No individual will suffer a bad reaction. (ii) Exactly two individuals will suffer a bad reaction. (iii) More than 2 individuals will suffer a bad reaction	4	CO3	L3	1.7.1

OR

6	a.	The number of messages sent per hour over a computer network has the following probability distribution: <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td>x</td> <td>10</td> <td>11</td> <td>12</td> <td>13</td> <td>14</td> <td>15</td> </tr> <tr> <td>P(X=x)</td> <td>0.08</td> <td>0.15</td> <td>0.30</td> <td>0.20</td> <td>0.20</td> <td>0.07</td> </tr> </table> Determine the mean and standard deviation of the number of messages sent per hour.	x	10	11	12	13	14	15	P(X=x)	0.08	0.15	0.30	0.20	0.20	0.07	5	CO3	L3	1.7.1
	x	10	11	12	13	14	15													
P(X=x)	0.08	0.15	0.30	0.20	0.20	0.07														
b.	Six dice are thrown 729 times. How many times do you expect at least three dice to show 5 or 6?	5	CO3	L3	1.7.1															

7	a.	Explain Simple Linear Regression.	2	CO4	L2	4.6.2											
	b.	Fit a Least Squares Regression line for the five-point data set. <table style="margin-left: auto; margin-right: auto;"> <tr> <td>x</td> <td>2</td> <td>2</td> <td>6</td> <td>8</td> <td>10</td> </tr> <tr> <td>y</td> <td>0</td> <td>1</td> <td>2</td> <td>3</td> <td>3</td> </tr> </table> Verify that it fits the data better by estimating the Sum of Squared Errors (SSE).	x	2	2	6	8	10	y	0	1	2	3	3	8	CO4	L3
x	2	2	6	8	10												
y	0	1	2	3	3												

OR

8		The age x and resting heart rate y were measured for ten men, with the results shown in the table: <table style="margin-left: auto; margin-right: auto;"> <tr> <td>x</td> <td>20</td> <td>23</td> <td>30</td> <td>37</td> <td>35</td> <td>45</td> <td>51</td> <td>55</td> <td>60</td> <td>63</td> </tr> <tr> <td>y</td> <td>72</td> <td>71</td> <td>73</td> <td>74</td> <td>74</td> <td>73</td> <td>72</td> <td>79</td> <td>75</td> <td>77</td> </tr> </table>	x	20	23	30	37	35	45	51	55	60	63	y	72	71	73	74	74	73	72	79	75	77	10	CO4	L3	4.6.2
	x	20	23	30	37	35	45	51	55	60	63																	
y	72	71	73	74	74	73	72	79	75	77																		
a.	Compute the linear correlation coefficient for these sample data.																											
b.	Interpret its meaning in the context of the problem.																											

9	<p>Consider the two-dimensional patterns:</p> <table border="1" data-bbox="359 190 1045 291"> <tr> <td>x</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>6</td> <td>7</td> </tr> <tr> <td>y</td> <td>1</td> <td>5</td> <td>3</td> <td>6</td> <td>7</td> <td>8</td> </tr> </table> <p>Compute the principal component using PCA Algorithm.</p>	x	2	3	4	5	6	7	y	1	5	3	6	7	8	10	CO5	L3	4.6.2				
x	2	3	4	5	6	7																	
y	1	5	3	6	7	8																	
OR																							
10	<p>Assume, we have the below dataset which has 3 features.</p> <table border="1" data-bbox="363 548 790 801"> <tr> <th>1</th> <th>2</th> <th>3</th> </tr> <tr> <td>1</td> <td>2</td> <td>3</td> </tr> <tr> <td>5</td> <td>5</td> <td>6</td> </tr> <tr> <td>1</td> <td>4</td> <td>2</td> </tr> <tr> <td>5</td> <td>3</td> <td>2</td> </tr> <tr> <td>8</td> <td>1</td> <td>2</td> </tr> </table> <p>Apply the following steps of Principal Component Analysis (PCA):</p> <ol style="list-style-type: none"> Standardize the Dataset by applying the standardization formula: $Z = \frac{x-\mu}{\sigma}$ Explain the significance of Standardization in PCA. Calculate the Covariance matrix for the above dataset. What do the entries of the calculated Covariance matrix tell us about the correlations between the variables? 	1	2	3	1	2	3	5	5	6	1	4	2	5	3	2	8	1	2	10	CO5	L3	4.6.2
1	2	3																					
1	2	3																					
5	5	6																					
1	4	2																					
5	3	2																					
8	1	2																					

Reference: Z Table

Critical Values (z_{α})	CRITICAL VALUES (z_{α}) OF Z		
	Level of significance (α)		
	1%	5%	10%
Two-tailed test	$ Z_{\alpha} = 2.58$	$ Z_{\alpha} = 1.96$	$ Z_{\alpha} = 1.645$
Right-tailed test	$Z_{\alpha} = 2.33$	$Z_{\alpha} = 1.645$	$Z_{\alpha} = 1.28$
Left-tailed test	$Z_{\alpha} = -2.33$	$Z_{\alpha} = -1.645$	$Z_{\alpha} = -1.28$

M.Sc.(DS).I/11.23.005 Reg.No.

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

C

**M.SC. COMPUTER SCIENCE WITH SPECIALIZATION IN DATA SCIENCE
FIRST SEMESTER EXAMINATION NOVEMBER 2023**

22-359-0105- MATHEMATICS FOR MACHINE LEARNING

(Regular)

Time: 3 Hours

Maximum Marks :50

Each question carries 10 Marks

No		QUESTIONS	MARKS	CO	BL	PI
1	a.	Check whether the given system of linear equations is consistent. If yes find the solution. $3x + 4y + 4z = 2$ $2x + 6y + z = 7$ $4x + 10y - z = 1$	10	CO1	L3	1.2.1
OR						
2	a.	Solve the following system of equations using Jacobi Method $3x_1 - x_2 + x_3 = 1$ $3x_1 + 3x_2 + 8x_3 = 4$ $3x_1 + 6x_2 + 2x_3 = 0$	10	CO1	L3	1.2.1
3	a.	Find the angles of the triangle ABC if the sides consist of the following vectors: Vector AB [3, 2, 6], Vector BC [2, 6, -3] and Vector AC [5, 8, 3].	5	CO2	L3	1.2.1
	b.	Determine if the following set of vectors is linearly independent: [2, 2, 0] [1, -1, 1] [4, 2, -2].	5			
OR						
4	a.	Check whether the vectors (1,3,2, -2), (4,1, -1,3), (1,1,2,0), (0,0,0,1) form a basis of the vector space R^4 .	7	CO2	L3	1.2.1
	b.	Find vector norms of the vector $X^T = [3 -3 0 1 -4]$.	3			

5	<p>a. Consider three vectors representing the preferences of three users for movies in a movie recommendation system. Each vector represents the ratings given by the users to a set of movies. Suppose we have the following three user preference vectors, A's preferences: [5, 4, 2, 0, 3] B's preferences: [4, 5, 3, 0, 2], and C's preferences: [2,0,5,4,7]. These vectors represent the ratings (on a scale of 1 to 5) for five different movies. Compute the cosine similarity between A, B and C based on their preferences.</p>	10	CO3	L3	1.2
OR					
6	<p>a. Find the derivative of the following:</p> $f(x)=(x+1)e^x$ $f(x)=(x^2+3)(2x^3+5x)$ $f(x)=3x-x^2y^2+2x^3y$ <p>b. Find all first order and second order partial derivatives of the following:</p> $z=xy \cos(xy)$ $z=xy+x^2+2y$ <p>c. Find slope of the surface $Z=xe^{-x}+5y$ in y- direction at (4,5).</p>	3	CO3	L3	1.2.1
		4			
		3			
7	<p>Find the eigenvalues and corresponding eigen vectors of the following matrix</p> $A = \begin{bmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{bmatrix}$	10	CO4	L2	1.2.1
OR					
8	<p>a. Determine the rank of the matrix.</p> $\begin{bmatrix} 4 & 3 & 4 & 5 \\ 2 & 4 & 5 & 6 \\ 4 & 5 & 6 & 7 \\ 9 & 10 & 11 & 12 \end{bmatrix}$	5	CO4	L3	1.2.1

	b	Check whether the given matrix A is an orthogonal matrix. If yes find the inverse. $A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix}$	5													
9	a.	Apply Principal Component analysis to do the dimensionality reduction on following data. <table border="1" style="margin-left: 40px;"> <tr> <td>X</td> <td>4</td> <td>8</td> <td>13</td> <td>7</td> </tr> <tr> <td>Y</td> <td>11</td> <td>4</td> <td>5</td> <td>14</td> </tr> </table>	X	4	8	13	7	Y	11	4	5	14	10	CO5	L3	1.2.1
X	4	8	13	7												
Y	11	4	5	14												
OR																
10	a.	Explain the steps in density estimation with gaussian mixture model.	7	CO5	L3	1.2.1										
	b.	Find directional derivative of $x^3y - yz^3 + z$ at $(2, -3, 0)$ in the direction of $3i + 4j + 3k$	3													

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

**M.SC. COMPUTER SCIENCE WITH SPECIALIZATION IN DATA SCIENCE
FIRST SEMESTER EXAMINATION NOVEMBER 2023**

22-359-0103 DATA STRUCTURES AND ALGORITHMS

(Regular)

Time: 3 Hours

Maximum Marks :50

Each question carries 10 Marks

No	QUESTIONS	MARKS	CO	BL	PI
1	Write an algorithm to perform insert and delete operations on a min-heap implemented on arrays. Also, analyze the time complexity of the algorithm.	10	CO1	L2	1.7.1
OR					
2	Write an algorithm to find the k-th smallest element of an array of numbers. Analyze the time complexity of your algorithm. Also, discuss the notions of best-case and worst-case in the context of the problem.	10	CO1	L2	1.7.1
3	<p>a. Write an algorithm to perform enqueue operation on a queue implemented using linked lists.</p> <p>b. Do the following operations on the given binary search tree .</p> <div style="text-align: center;"> <pre> graph TD 15((15)) --- 6((6)) 15 --- 18((18)) 6 --- 3((3)) 6 --- 7((7)) 3 --- 2((2)) 3 --- 4((4)) 7 --- 13((13)) 13 --- 9((9)) 18 --- 17((17)) 18 --- 20((20)) </pre> </div> <ul style="list-style-type: none"> ● Left rotation with respect to 7 and 13. ● Right rotation with respect to 6 and 15. 	5 5	CO2 CO2	L2 L2	1.7.1 1.7.1

OR

4	<p>Perform delete() operation on the Fibonacci heap shown below. Show the heap after every intermediate transformation.</p> <div style="text-align: center; margin: 10px 0;"> </div>	10	CO2	L3	1.7.1
---	--	----	-----	----	-------

5	<p>a.</p> <table border="1" style="margin: 0 auto; text-align: center;"> <tr> <td>-3</td><td>0</td><td>0</td><td>-1</td><td>-1</td><td>6</td><td>-2</td><td>-1</td><td>-1</td><td>-1</td> </tr> <tr> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td> </tr> </table> <p>Do the following unions on the disjoint set shown above using union by size.</p> <ul style="list-style-type: none"> ● Union (3,0) ● Union (4,6) ● Union (8,9) <p>Show the disjoint set after each of the operations.</p> <p>b. Perform Find (7) & Find (5) with path compression in the following disjoint set.</p> <table border="1" style="margin: 0 auto; text-align: center;"> <tr> <td>-1</td><td>0</td><td>1</td><td>-1</td><td>-1</td><td>6</td><td>-1</td><td>2</td><td>-1</td><td>-1</td> </tr> <tr> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td> </tr> </table> <p>Also, write the algorithm for performing find operations with path compression.</p>	-3	0	0	-1	-1	6	-2	-1	-1	-1	0	1	2	3	4	5	6	7	8	9	-1	0	1	-1	-1	6	-1	2	-1	-1	0	1	2	3	4	5	6	7	8	9	5	CO3	L3	1.7.1
-3	0	0	-1	-1	6	-2	-1	-1	-1																																				
0	1	2	3	4	5	6	7	8	9																																				
-1	0	1	-1	-1	6	-1	2	-1	-1																																				
0	1	2	3	4	5	6	7	8	9																																				
	OR																																												
6	<table border="1" style="margin: 0 auto; text-align: center;"> <tr> <td>-1</td><td>0</td><td>1</td><td>-1</td><td>-1</td><td>6</td><td>-1</td><td>2</td><td>-1</td><td>-1</td> </tr> <tr> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td> </tr> </table> <p>Implement the above disjoint sets using linked lists. Write the pseudocode for performing Union and Find operations on the linked list implementation of disjoint sets.</p>	-1	0	1	-1	-1	6	-1	2	-1	-1	0	1	2	3	4	5	6	7	8	9	10	CO3	L3	1.7.1																				
-1	0	1	-1	-1	6	-1	2	-1	-1																																				
0	1	2	3	4	5	6	7	8	9																																				

7	<p>Five students applied to four different internship positions within a company. The preferences given by the students are as follows. The students are ordered based on their rank in the qualifying test. Do bipartite matching using Ford-Fulkerson algorithm and find the maximum matching.</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Student</th> <th>Positions</th> </tr> </thead> <tbody> <tr> <td>S1</td> <td>P1, P3</td> </tr> <tr> <td>S4</td> <td>P1</td> </tr> <tr> <td>S3</td> <td>P4</td> </tr> <tr> <td>S2</td> <td>P2, P3</td> </tr> <tr> <td>S5</td> <td>P3, P4</td> </tr> </tbody> </table>	Student	Positions	S1	P1, P3	S4	P1	S3	P4	S2	P2, P3	S5	P3, P4	10	CO4	L3	1.7.1
Student	Positions																
S1	P1, P3																
S4	P1																
S3	P4																
S2	P2, P3																
S5	P3, P4																

OR

8	<p>Compute the convex hull of the following set of points using Graham's scan algorithm.</p> <p>(1,1), (9,2), (10,3), (7,8), (6,7), (4,11), (1,9), (11,7), (3,8) & (2,8).</p>	10	CO4	L3	1.7.1
---	---	----	-----	----	-------

9	<p>Encode the character string CABDIG using Huffman codes. The character frequency table is as follows. How much space could be saved here in comparison with the fixed length encoding of the string?</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Character</th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> <th>F</th> <th>G</th> <th>H</th> <th>I</th> </tr> </thead> <tbody> <tr> <td>Frequency (%)</td> <td>19</td> <td>3</td> <td>7</td> <td>4</td> <td>21</td> <td>15</td> <td>6</td> <td>20</td> <td>5</td> </tr> </tbody> </table>	Character	A	B	C	D	E	F	G	H	I	Frequency (%)	19	3	7	4	21	15	6	20	5	10	CO5	L3	1.7.1
Character	A	B	C	D	E	F	G	H	I																
Frequency (%)	19	3	7	4	21	15	6	20	5																

OR

10	a.	<p>A multi-processing environment with three processors is given a set of jobs whose running times are as shown in the table below. Do a greedy scheduling of the jobs. Is your schedule optimal? Justify your answer.</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Job</th> <th>J1</th> <th>J2</th> <th>J3</th> <th>J4</th> <th>J5</th> <th>J6</th> <th>J7</th> <th>J8</th> <th>J9</th> </tr> </thead> <tbody> <tr> <td>Time</td> <td>3</td> <td>5</td> <td>7</td> <td>9</td> <td>11</td> <td>14</td> <td>15</td> <td>18</td> <td>20</td> </tr> </tbody> </table>	Job	J1	J2	J3	J4	J5	J6	J7	J8	J9	Time	3	5	7	9	11	14	15	18	20	6	CO5	L3	1.7.1
	Job	J1	J2	J3	J4	J5	J6	J7	J8	J9																
Time	3	5	7	9	11	14	15	18	20																	
b.	<p>Do bin packing with items having sizes 0.5, 0.6, 0.7, 0.1, 0.4, 0.2, 0.7, 0.5, 0.1, 0.8, 0.4, 0.9, 0.3 & 0.1 using First-fit (Online) and Best-fit (Online). Assume a bin size of 1.5.</p>	4	CO5	L3	1.7.1																					

M.Sc.(DS).I/11.23.004 Reg.No.

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

B

**M.SC. COMPUTER SCIENCE WITH SPECIALIZATION IN DATA SCIENCE
FIRST SEMESTER EXAMINATION NOVEMBER 2023**

22-359-0104 PYTHON FOR DATA ANALYTICS

(Regular)

Time: 3 Hours

Maximum Marks :50

Each question carries 10 Marks

Qn No		Questions	Marks	CO	BL	PI
1	a	Differentiate between qualitative and quantitative data.	4	CO1	L2	1.7.1
	b	Explain the data analysis process in detail.	6	CO1	L2	1.7.1
OR						
2	a	Name a few libraries in python used for Data Analysis and Scientific Computations. Write short note on each.	4	CO1	L2	1.7.1
	b	Differentiate between categorical and numerical data	6	CO1	L2	1.7.1
3	a	Predict the output for the given code <pre>A = np.arange(10, 19).reshape((3, 3)) print(A) print(np.sum(A)) print(A*A) print(A[0,:])</pre>	4	CO4	L3	1.7.1
	b	With suitable example explain read and write, data on Microsoft Excel files in python.	6	CO4	L3	1.7.1
OR						
4	a	Consider the given 1-D array: [1,2,3,4,5,6,7,8,9,10,11,12,13,14,15] Write a program to i. Create the above array using <i>arange ()</i> . ii. Negate all the elements between 5 and 10, in place. iii. Half the elements which are even. iv. Double the elements which are odd.	5	CO4	L3	1.7.1
	b	Consider the given DataFrame Salary.	5	CO4	L3	1.7.1

Emp_ID	Name	Dept	Salary	Status
100	Kabir	IT	34000	Regular
110	Rishav	Finance	28500	Regular
120	Seema	IT	13500	Contract
130	David	IT	41000	Regular
140	Ruchi	HRD	17000	Contract

Write python code

- i. To create DataFrame Salary
- ii. To display the 4th record.
- iii. To delete the column status.
- iv. To delete the first and last record

5	a	<table border="1"> <thead> <tr> <th>Emp_Id</th> <th>F_Name</th> <th>S_Name</th> <th>Year_Hiring</th> </tr> </thead> <tbody> <tr> <td>1001</td> <td>Peter</td> <td>Vargas</td> <td>2001</td> </tr> <tr> <td>1002</td> <td>John</td> <td>Smith</td> <td>2001</td> </tr> <tr> <td>1003</td> <td>Karen</td> <td>Partners</td> <td>2001</td> </tr> <tr> <td>1004</td> <td>John</td> <td>Alberto</td> <td>2001</td> </tr> <tr> <td>1005</td> <td>Randall</td> <td>Matos</td> <td>2001</td> </tr> </tbody> </table> <p>Table:1</p> <table border="1"> <thead> <tr> <th>Emp_Id</th> <th>F_Name</th> <th>S_Name</th> <th>Year_Hiring</th> </tr> </thead> <tbody> <tr> <td>2001</td> <td>Max</td> <td>James</td> <td>2002</td> </tr> <tr> <td>2002</td> <td>Adam</td> <td>Fripp</td> <td>2002</td> </tr> <tr> <td>2003</td> <td>Steven</td> <td>Markle</td> <td>2002</td> </tr> <tr> <td>2004</td> <td>James</td> <td>Marlow</td> <td>2002</td> </tr> <tr> <td>2005</td> <td>Jason</td> <td>Mallin</td> <td>2002</td> </tr> </tbody> </table> <p>Table:2</p> <p>Write the above given data into two separate sheets of the excel file (employee.xlsx)</p>	Emp_Id	F_Name	S_Name	Year_Hiring	1001	Peter	Vargas	2001	1002	John	Smith	2001	1003	Karen	Partners	2001	1004	John	Alberto	2001	1005	Randall	Matos	2001	Emp_Id	F_Name	S_Name	Year_Hiring	2001	Max	James	2002	2002	Adam	Fripp	2002	2003	Steven	Markle	2002	2004	James	Marlow	2002	2005	Jason	Mallin	2002	5	CO4		1.7.1
	Emp_Id	F_Name	S_Name	Year_Hiring																																																		
1001	Peter	Vargas	2001																																																			
1002	John	Smith	2001																																																			
1003	Karen	Partners	2001																																																			
1004	John	Alberto	2001																																																			
1005	Randall	Matos	2001																																																			
Emp_Id	F_Name	S_Name	Year_Hiring																																																			
2001	Max	James	2002																																																			
2002	Adam	Fripp	2002																																																			
2003	Steven	Markle	2002																																																			
2004	James	Marlow	2002																																																			
2005	Jason	Mallin	2002																																																			
	b	How can you reverse a NumPy array using slicing method? Illustrate with example.	5	CO4		1.7.1																																																

OR

6	a	What happens when we use the <i>arrays_split()</i> method for splitting the NumPy array? Illustrate with example.	4	CO4	L2	1.7.1																																		
	b	<p>Consider the given dataframe labeled IPL.</p> <table border="1"> <thead> <tr> <th>Player</th> <th>Team</th> <th>Category</th> <th>BidPrice</th> <th>Runs</th> </tr> </thead> <tbody> <tr> <td>Hardik Pandya</td> <td>Mumbai Indians</td> <td>Batsman</td> <td>13</td> <td>1000</td> </tr> <tr> <td>KL Rahul</td> <td>Kings Eleven</td> <td>Batsman</td> <td>12</td> <td>2400</td> </tr> <tr> <td>Andre Russel</td> <td>Kolkata Knight riders</td> <td>Batsman</td> <td>7</td> <td>900</td> </tr> <tr> <td>Jasprit Bumrah</td> <td>Mumbai Indians</td> <td>Bowler</td> <td>10</td> <td>200</td> </tr> <tr> <td>Virat Kohli</td> <td>RCB</td> <td>Batsman</td> <td>17</td> <td>3600</td> </tr> <tr> <td>Rohit Sharma</td> <td>Mumbai Indians</td> <td>Batsman</td> <td>15</td> <td>3700</td> </tr> </tbody> </table> <p>Write python program to</p> <ol style="list-style-type: none"> i. Create the dataframe IPL. ii. To print total players per team. iii. To find player who had highest BidPrice from each team. 	Player	Team	Category	BidPrice	Runs	Hardik Pandya	Mumbai Indians	Batsman	13	1000	KL Rahul	Kings Eleven	Batsman	12	2400	Andre Russel	Kolkata Knight riders	Batsman	7	900	Jasprit Bumrah	Mumbai Indians	Bowler	10	200	Virat Kohli	RCB	Batsman	17	3600	Rohit Sharma	Mumbai Indians	Batsman	15	3700	6	CO4	L3
Player	Team	Category	BidPrice	Runs																																				
Hardik Pandya	Mumbai Indians	Batsman	13	1000																																				
KL Rahul	Kings Eleven	Batsman	12	2400																																				
Andre Russel	Kolkata Knight riders	Batsman	7	900																																				
Jasprit Bumrah	Mumbai Indians	Bowler	10	200																																				
Virat Kohli	RCB	Batsman	17	3600																																				
Rohit Sharma	Mumbai Indians	Batsman	15	3700																																				

7	a	Explain hist () function with example.	4	CO3	L3	1.7.1																																																																																																																				
	b	Use the following company_sales_data.csv file <table border="1"> <thead> <tr> <th>month_number</th> <th>facecream</th> <th>facewash</th> <th>toothpaste</th> <th>bathingsoap</th> <th>shampoo</th> <th>moisturizer</th> <th>total_units</th> <th>total_profit</th> </tr> </thead> <tbody> <tr><td>1</td><td>2500</td><td>1500</td><td>5200</td><td>9200</td><td>1200</td><td>1500</td><td>21100</td><td>211000</td></tr> <tr><td>2</td><td>2630</td><td>1200</td><td>5100</td><td>6100</td><td>2100</td><td>1200</td><td>18330</td><td>183300</td></tr> <tr><td>3</td><td>2140</td><td>1340</td><td>4550</td><td>9550</td><td>3550</td><td>1340</td><td>22470</td><td>224700</td></tr> <tr><td>4</td><td>3400</td><td>1130</td><td>5870</td><td>8870</td><td>1870</td><td>1130</td><td>22270</td><td>222700</td></tr> <tr><td>5</td><td>3600</td><td>1740</td><td>4560</td><td>7760</td><td>1560</td><td>1740</td><td>20960</td><td>209600</td></tr> <tr><td>6</td><td>2760</td><td>1555</td><td>4890</td><td>7490</td><td>1890</td><td>1555</td><td>20140</td><td>201400</td></tr> <tr><td>7</td><td>2980</td><td>1120</td><td>4780</td><td>8980</td><td>1780</td><td>1120</td><td>29550</td><td>295500</td></tr> <tr><td>8</td><td>3700</td><td>1400</td><td>5860</td><td>9960</td><td>2860</td><td>1400</td><td>36140</td><td>361400</td></tr> <tr><td>9</td><td>3540</td><td>1780</td><td>6100</td><td>8100</td><td>2100</td><td>1780</td><td>23400</td><td>234000</td></tr> <tr><td>10</td><td>1990</td><td>1890</td><td>8300</td><td>10300</td><td>2300</td><td>1890</td><td>26670</td><td>266700</td></tr> <tr><td>11</td><td>2340</td><td>2100</td><td>7300</td><td>13300</td><td>2400</td><td>2100</td><td>41280</td><td>412800</td></tr> <tr><td>12</td><td>2900</td><td>1760</td><td>7400</td><td>14400</td><td>1800</td><td>1760</td><td>30020</td><td>300200</td></tr> </tbody> </table> i. Write python code to read total profit of all months and show it using a line plot. Generated line plot must include the following properties: — a. X label name = Month Number b. Y label name = Total profit ii. Calculate total sale data for each product and show it using a Pie chart.	month_number	facecream	facewash	toothpaste	bathingsoap	shampoo	moisturizer	total_units	total_profit	1	2500	1500	5200	9200	1200	1500	21100	211000	2	2630	1200	5100	6100	2100	1200	18330	183300	3	2140	1340	4550	9550	3550	1340	22470	224700	4	3400	1130	5870	8870	1870	1130	22270	222700	5	3600	1740	4560	7760	1560	1740	20960	209600	6	2760	1555	4890	7490	1890	1555	20140	201400	7	2980	1120	4780	8980	1780	1120	29550	295500	8	3700	1400	5860	9960	2860	1400	36140	361400	9	3540	1780	6100	8100	2100	1780	23400	234000	10	1990	1890	8300	10300	2300	1890	26670	266700	11	2340	2100	7300	13300	2400	2100	41280	412800	12	2900	1760	7400	14400	1800	1760	30020	300200	6	CO3	L3
month_number	facecream	facewash	toothpaste	bathingsoap	shampoo	moisturizer	total_units	total_profit																																																																																																																		
1	2500	1500	5200	9200	1200	1500	21100	211000																																																																																																																		
2	2630	1200	5100	6100	2100	1200	18330	183300																																																																																																																		
3	2140	1340	4550	9550	3550	1340	22470	224700																																																																																																																		
4	3400	1130	5870	8870	1870	1130	22270	222700																																																																																																																		
5	3600	1740	4560	7760	1560	1740	20960	209600																																																																																																																		
6	2760	1555	4890	7490	1890	1555	20140	201400																																																																																																																		
7	2980	1120	4780	8980	1780	1120	29550	295500																																																																																																																		
8	3700	1400	5860	9960	2860	1400	36140	361400																																																																																																																		
9	3540	1780	6100	8100	2100	1780	23400	234000																																																																																																																		
10	1990	1890	8300	10300	2300	1890	26670	266700																																																																																																																		
11	2340	2100	7300	13300	2400	2100	41280	412800																																																																																																																		
12	2900	1760	7400	14400	1800	1760	30020	300200																																																																																																																		

OR

8	a	What is the purpose of the plt.subplots() function in Matplotlib? Explain with example.	3	CO3	L2	1.7.1											
	b	Draw a pie chart to represent the following data: <table border="1"> <thead> <tr> <th>Classes</th> <th>Percentage of number of students</th> </tr> </thead> <tbody> <tr><td>Pre-primary</td><td>25</td></tr> <tr><td>I to IV</td><td>20</td></tr> <tr><td>V to VI</td><td>15</td></tr> <tr><td>VII to VIII</td><td>30</td></tr> <tr><td>IX to X</td><td>10</td></tr> </tbody> </table> Generated pie chart must include the following properties: – i. The starting angle is 90 ii. Explode the two highest percentage	Classes	Percentage of number of students	Pre-primary	25	I to IV	20	V to VI	15	VII to VIII	30	IX to X	10	7	CO3	L3
Classes	Percentage of number of students																
Pre-primary	25																
I to IV	20																
V to VI	15																
VII to VIII	30																
IX to X	10																

9	a	Differentiate between supervised and unsupervised learning? List the algorithms in each category.	4	CO2, CO5	L2	1.7.1
	b	Predict the salary based on year of experience, using linear regression model in supervised learning.	6	CO2, CO5	L3	

		<table border="1"> <thead> <tr> <th>YearsExperience</th> <th>Salary</th> </tr> </thead> <tbody> <tr><td>1.1</td><td>39,343</td></tr> <tr><td>1.3</td><td>46,205</td></tr> <tr><td>1.5</td><td>37,731</td></tr> <tr><td>2.0</td><td>43,525</td></tr> <tr><td>2.2</td><td>39,891</td></tr> <tr><td>2.9</td><td>56,642</td></tr> <tr><td>3.0</td><td>60,150</td></tr> <tr><td>3.2</td><td>54,445</td></tr> <tr><td>3.2</td><td>64,445</td></tr> <tr><td>3.7</td><td>57,189</td></tr> </tbody> </table>	YearsExperience	Salary	1.1	39,343	1.3	46,205	1.5	37,731	2.0	43,525	2.2	39,891	2.9	56,642	3.0	60,150	3.2	54,445	3.2	64,445	3.7	57,189				
YearsExperience	Salary																											
1.1	39,343																											
1.3	46,205																											
1.5	37,731																											
2.0	43,525																											
2.2	39,891																											
2.9	56,642																											
3.0	60,150																											
3.2	54,445																											
3.2	64,445																											
3.7	57,189																											

OR

10	a	Differentiate between classification and regression.	4	CO2, CO5	L2	1.7.1
	b	Classify the Iris dataset with the help of scikit library using KNN algorithm.	6	CO2, CO5	L3	1.7.1
