

MID-TERM EXAMINATION
(Course Name :CSE-AI/AI&ML) (Semester : II)
(May, 2023) OFF LINE mode

Subject Code: BAI-104	Subject: Introduction To Data Science																																																				
Time :1 ½ Hours	Maximum Marks : 30																																																				
Note: Q. 1 is compulsory.																																																					
Q1	(2.5*4)																																																				
(a) "Database contains data in its raw format". Is this statement Valid? Justify.																																																					
(b) Differentiate between Diversity and Inclusion. Further mention how are they interlinked.																																																					
(c) Write code using NumPy for following problem. You work as a meteorologist. Your job is to maintain hourly rainfall data in millimeters in the array. rain_mm_by_hour. On a particularly rainy day, you get the results stored in the np array: (i) At the end of the day, you are asked to report the number of hours with heavy rainfall (i.e., rainfall greater than 4 mm). (ii) You want to know how much it rained during the even-numbered hours.																																																					
(d) Write a python code to construct the following dataframe and then sort the names in alphabetical order.																																																					
<table style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th></th> <th>RollNo</th> <th>Name</th> <th>Eco</th> <th>Maths</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>1</td> <td>Arnab</td> <td>18</td> <td>57</td> </tr> <tr> <td>1</td> <td>2</td> <td>Kritika</td> <td>23</td> <td>45</td> </tr> <tr> <td>2</td> <td>3</td> <td>Divyam</td> <td>51</td> <td>37</td> </tr> <tr> <td>3</td> <td>4</td> <td>Vivaan</td> <td>40</td> <td>60</td> </tr> <tr> <td>4</td> <td>5</td> <td>Aaroosh</td> <td>18</td> <td>27</td> </tr> </tbody> </table>			RollNo	Name	Eco	Maths	0	1	Arnab	18	57	1	2	Kritika	23	45	2	3	Divyam	51	37	3	4	Vivaan	40	60	4	5	Aaroosh	18	27																						
	RollNo	Name	Eco	Maths																																																	
0	1	Arnab	18	57																																																	
1	2	Kritika	23	45																																																	
2	3	Divyam	51	37																																																	
3	4	Vivaan	40	60																																																	
4	5	Aaroosh	18	27																																																	
Q2	(5,5)																																																				
(Attempt any Two Parts) UNIT-1																																																					
(a) What are the different phases of a data science process? Explain																																																					
(b) Explain the different C's that are required to be followed while handling data.																																																					
(C) (i) "In big data, veracity corresponds to the accuracy of the considered Data". Is this statement true? Justify. (ii) Explain the different V's of Big data.																																																					
Q3	(5,5)																																																				
(Attempt any Two Parts) UNIT-2																																																					
(a) What are the various methods to replace values inside an array? Give an example for each.																																																					
(b) Write a python program to construct the following dataframe :																																																					
<table style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th></th> <th>Player</th> <th>Year</th> <th>Points</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>Superman</td> <td>2000</td> <td>23</td> </tr> <tr> <td>1</td> <td>Batman</td> <td>2000</td> <td>43</td> </tr> <tr> <td>2</td> <td>Thanos</td> <td>2000</td> <td>45</td> </tr> <tr> <td>3</td> <td>Batman</td> <td>2001</td> <td>65</td> </tr> <tr> <td>4</td> <td>Thanos</td> <td>2001</td> <td>76</td> </tr> <tr> <td>5</td> <td>Superman</td> <td>2002</td> <td>34</td> </tr> <tr> <td>6</td> <td>Batman</td> <td>2002</td> <td>23</td> </tr> <tr> <td>7</td> <td>Thanos</td> <td>2002</td> <td>78</td> </tr> <tr> <td>8</td> <td>Black Widow</td> <td>2003</td> <td>89</td> </tr> <tr> <td>9</td> <td>Batman</td> <td>2004</td> <td>76</td> </tr> <tr> <td>10</td> <td>Thanos</td> <td>2004</td> <td>92</td> </tr> <tr> <td>11</td> <td>Superman</td> <td>2005</td> <td>87</td> </tr> </tbody> </table>			Player	Year	Points	0	Superman	2000	23	1	Batman	2000	43	2	Thanos	2000	45	3	Batman	2001	65	4	Thanos	2001	76	5	Superman	2002	34	6	Batman	2002	23	7	Thanos	2002	78	8	Black Widow	2003	89	9	Batman	2004	76	10	Thanos	2004	92	11	Superman	2005	87
	Player	Year	Points																																																		
0	Superman	2000	23																																																		
1	Batman	2000	43																																																		
2	Thanos	2000	45																																																		
3	Batman	2001	65																																																		
4	Thanos	2001	76																																																		
5	Superman	2002	34																																																		
6	Batman	2002	23																																																		
7	Thanos	2002	78																																																		
8	Black Widow	2003	89																																																		
9	Batman	2004	76																																																		
10	Thanos	2004	92																																																		
11	Superman	2005	87																																																		
Perform the following functions:																																																					
(i) Segregate the elements and group them player-wise.																																																					
(ii) Write a function to print the players with points above 50.																																																					
(iii) Write a function to print last 4 elements of the dataframe.																																																					

(C) Write code using NumPy

Using the stock data set, analyze the data to programmatically print answers to the following business questions.

First few observations of Amazon stock dataset in csv format are:

Date	Close/Last	Volume	Open	High	Low
12/31/2021	3334.34	2391519	3379.12	3387	3331.17
12/30/2021	3372.89	1879155	3394	3417.76	3370.48
12/29/2021	3384.02	1787702	3416.8	3424.24	3372.0101

- (i). What was the lowest and highest opening price for the AMZN stock?
- (ii). How many days saw a high trading price that exceeded 3,500?
- (iii). Did the AMZN stock ever have an opening price less than 2,500? Your answer must include the phrase "did have" or "did not have."
- (iv). What were the top 3 opening prices for the AMZN stock?