# **Question Paper**

Exam Date & Time: 06-Dec-2023 (09:30 AM - 12:30 PM)



#### MANIPAL ACADEMY OF HIGHER EDUCATION

### INTERNATIONAL CENTRE FOR APPLIED SCIENCES END SEMESTER THEORY EXAMINATION - NOVEMBER/DECEMBER 2023 III SEMESTER B.Sc. (APPLIED SCIENCES) IN ENGG.

INTRODUCTION TO DATA ANALYTICS WITH PYTHON [ICS 235]

Marks: 50

## Answer all the questions.

#### Missing data, if any, may be suitably assumed

1)		Why do modern businesses need data visualization? [any 5 well-explained points]	(2.5)	
	A)			
	B)	Open a user-specified file in read mode and display the content of the file and the set of even length words used in the file with their frequency of occurrence; if the user inputs a mode other than read, then raise an exception "INVALID MODE.	(5)	
	C)	Write a python code to find the second largest element in a user-defined array. Demonstrate the utility of the 'None' data type in this code.	(2.5)	
2)		Write python code to plot pie chart and bar graph. [ Include libraries ]	(2.5)	
	A)			
	B)	Write Python program: i) to replace whitespaces with an underscore and vice versa . ii) to extract year, month and date from an url.	(5)	
		iv) to replace a specific word in the given string by user defined word. v) to separate and print the numbers of a given string.		
	C)	Create a subdirectory of your name in the working directory. Open an existing file to append the user-specified content till the user enters 'END.'	(2.5)	
3)		Consider Weather data set and answer the following:		(6)
	A)	<ul> <li>a) i) Create data frame and display the information of the data set.</li> <li>ii) Treat NA values by '0'.</li> <li>iii) Treat NA values by mean value of the column values.</li> <li>b) Display average min and average max temperature city-wise.</li> <li>c) Plot bar graph to showcase min, max temperature and rainfall . (x-axis: time, y-axis: country, Title</li> </ul>	:	
		'Min-Max Temperature') d) i) Which day highest air pressure has been observed and which place/places?		

- ii) Display total number of records year-wise.
- e) How many records have 'NA' entry?

		Min	Max				Wind Gust	Windo	e Dir	ind 9a	Wind Dir3p	Wind Speed	Wind Speed	Humidi	ity	Humidity	Pressure	Pressure	Cloud	d Cloud	Temp	Temp	Rain Tod
Date	Location	Temp	Temp	Rainfall	Evaporation	Sunshine	Dir	d	m		m	9am	3pm	9am		3pm	9am	3pm	9am	3pm	9am	3pm	ay
01-12-2008	Albury	13.4	22.9	0.6	NA	NA	W	4	4 W		WNW	20	24		71	22	1007.7	1007.1		8 NA	16.9	21.8	No
02-12-2008	Albury	7.4	25.1	0	NA	NA	WNW	1 4	4 NN	w	WSW	4	22		44	25	1010.6	1007.8	NA	NA	17.2	24.3	No

Duration: 180 mins.

03-12-2008 Albury	12.9	25.7	0 NA	NA	WSW	46	s w	WSW	1	19	26	38	30	1007.6	1008.7 NA		2	21	23.2 No	3 1
04-12-2008 Albury	9.2	28	0 NA	NA	NE	24	SE	E	1	11	9	45	16	1017.6	1012.8 NA	NA		18.1	26.5 No	
05-12-2008 Albury	17.5	32.3	1 NA	NA	w	41	L ENE	NW		7	20	82	33	1010.8	1006	7	8	17.8	29.7 No	1
05-12-2008 Albury	14.6	29.7	0.2 NA	NA	WNW	56	5 W	w	1	19	24	55	23	1009.2	1005.4 NA	NA		20.6	28.9 No	1
07-12-2008 Albury	14.3	25	0 NA	NA	w	50	SW 0	w	2	20	24	49	19	1009.6	1008.2	1 NA		18.1	24.6 No	
08-12-2008 Albury	7.7	26.7	0 NA	NA	w	35	S SSE	w		6	17	48	19	1013.4	1010.1 NA	NA		16.3	25.5 No	1
09-12-2008 Albury	9.7	31.9	0 NA	NA	NNW	80	SE	NW		7	28	42	9	1008.9	1003.6 NA	NA		18.3	30.2 No	· ·
10-12-2008 Albury	13.1	30.1	1.4 NA	NA	W	28	8 5	SSE	1	15	11	58	27	1007	1005.7 NA	NA		20.1	28.2 Ye	s I
11-12-2008 Albury	13.4	30.4	0 NA	NA	N	30	SSE	ESE	1	17	6	48	22	1011.8	1008.7 NA	NA		20.4	28.8 No	
12-12-2008 Albury	15.9	21.7	2.2 NA	NA	NNE	31	NE NE	ENE	1	15	13	89	91	1010.5	1004.2	8	8	15.9	17 Ye	s '
13-12-2008 Albury	15.9	18.6	15.6 NA	NA	w	61	NNW	NNW	2	28	28	76	93	994.3	993	8	8	17.4	15.8 Ye	5 1
14-12-2008 Albury	12.6	21	3.6 NA	NA	SW	44	w	SSW	2	24	20	65	43	1001.2	1001.8 NA		7	15.8	19.8 Ye	s I
15-12-2008 Albury	8.4	24.6	0 NA	NA	NA	NA	s	WNW		4	30	57	32	1009.7	1008.7 NA	NA		15.9	23.5 No	2 1
16-12-2008 Albury	9.8	27.7 NA	A NA	NA	WNW	/ 50	NA C	WNW	NA		22	50	28	1013.4	1010.3	0 NA		17.3	26.2 NA	4 1
17-12-2008 Albury	14.1	20.9	0 NA	NA	ENE	22	2 SSW	E	1	11	9	69	82	1012.2	1010.4	8	1	17.2	18.1 No	
18.12.2008 Albury	12 5	77 9	16 8 NA	NΔ	w	63	N	WNW		6	20	80	65	1005.8	1002.2	8	1	18	71 5 Ve	e '
What is	a Da	ataFra	ame ? (	Compa	re th	e dif	fere	nt w	ays	s to	o ad	d new	colu	imns i	n an ex	iting	)			(4)
DataFra	ame.	[ Sup	oport yc	our ans	wer \	with	exa	mple	e co	bde	es].									
Define f	eatu	re sc	aling/no	ormaliza	ation	. Exp	olair	n coi	mm	nor	n sca	aling o	pera	tions.						(4)

4)

5)

.

B)

A)
 B) Elaborate: "Correlation is not causal "
 C) Define feature selection. Explain different classes of feature engineering techniques.
 (4) Define each step in the K-Means Clustering algorithm. What are the disadvantages of the algorithm?

A) B)

We have data from the questionnaires survey (to ask people opinion) and objective testing with <sup>(6)</sup> two attributes (acid durability and strength) to classify whether a special paper tissue is good or not. Here is four training samples

X1 = Acid Durability (seconds)	X2 = Strength(kg/square meter)	Y = Classification
7	7	Bad
7	4	Bad
3	4	Good
1	4	Good

Now the factory produces a new paper tissue that pass laboratory test with X1 = 3 and X2 = 7.

Without another expensive survey, predict the classification Bad/Good of this new tissue using KNN algorithm. [Consider K=3]

-----End-----