Question Paper

Exam Date & Time: 04-Feb-2023 (09:30 AM - 12:30 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

INTERNATIONAL CENTRE FOR APPLIED SCIENCES
END SEMESTER THEORY EXAMINATION - DECEMBER 2022
III SEMESTER B.Sc (Applied Sciences) in Engg.
INTRODUCTION TO DATA ANALYTICS WITH PYTHON [ICS 235]

Marks: 50 **Duration: 180 mins.**

Answer all the questions.

Missing data, if any, may be suitably assumed

1) Describe any five basic features of python. [Elaborated] (2.5)

A)

B) Open a user-specified file in read mode and display the content of the file and the set of 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 smallest and largest element in a user-defined array. Demonstrate the utility of the 'None' data type in this code. (2.5)

2) What is a Data Frame? How to create a Data Frame using a List, Dictionary, and Tuple? [support your answer with example codes] (5)

A)

B) Write python code : (2.5)
i) to find all words starting with 'a' or 'e' in a given string.
ii) to replace all occurrences of space, comma, or dot with a colon.
iii) to find all five characters long word in a string.
iv) to separate and print the numbers of a given string.
v) to find all words which are at least 4 characters long in a 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 'FINISH.' (2.5)

3) Create a data frame called df from the following tabular data dictionary that has these index labels: ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j'] (5)

A)

	Animal	Age	Priority	Visits
a	cat	2.5	yes	1
b	cat	3.0	yes	3
c	snake	0.5	no	2
d	dog	NaN	yes	3
e	dog	5.0	no	2

f	cat	2.0	no	3
g	snake	4.5	no	1
h	cat	NaN	yes	1
i	dog	7.0	no	2
j	dog	3.0	no	1

a. Display a summary of the data frame's basic information.
b. Return the first three rows of the data frame df.
c. Select just the animal and age columns from the data frame df
d. Count the visit priority per animal
e. Find the mean of the animals' ages.
f. Display a summary of the below data set. [find the data set from the given dataset and then summarize it. Hint: Calculate statistics on age column.]

	count	mean	std	min	25%	50%	75%	max
Animal								
cat	3.0	2.5	0.500000	2.0	2.25	2.5	2.75	3.0
dog	3.0	5.0	2.000000	3.0	4.00	5.0	6.00	7.0
snake	2.0	2.5	2.828427	0.5	1.50	2.5	3.50	4.5

B) Consider below data set and perform the following: (5)

Date	Location	Min Temp	Max Temp	Rainfall	Evaporation	Sunshine	Wind Gust Dir	WindGustSpeed	Wind Dir9am	Wind Dir3pm	Wind Speed 9am	Wind Speed 3pm	Humidity 9am	Humidity 3pm	Pressure 9am	Pressure 3pm	Cloud 9am	Cloud 3pm	Temp 9am	Temp 3pm	Rain Today
01-12-2008	Albury	13.4	22.9	0.6	NA	NA	W	44	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	44	NNW	WSW	4	22	44	25	1010.6	1007.8	NA	NA	17.2	24.3	No
03-12-2008	Albury	12.9	25.7	0	NA	NA	WSW	46	W	WSW	19	26	38	30	1007.6	1008.7	NA	2	21	23.2	No
04-12-2008	Albury	9.2	28	0	NA	NA	NE	24	SE	E	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	ENE	NW	7	20	82	33	1010.8	1006	7	8	17.8	29.7	No
06-12-2008	Albury	14.6	29.7	0.2	NA	NA	WNW	56	W	W	19	24	55	23	1009.2	1005.4	NA	NA	20.6	28.9	No
07-12-2008	Albury	14.3	25	0	NA	NA	W	50	SW	W	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	SSE	W	6	17	48	19	1013.4	1010.1	NA	NA	16.3	25.5	No
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	S	SSE	15	11	58	27	1007	1005.7	NA	NA	20.1	28.2	Yes
11-12-2008	Albury	13.4	30.4	0	NA	NA	N	30	SSE	ESE	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	ENE	15	13	89	91	1010.5	1004.2	8	8	15.9	17	Yes
13-12-2008	Albury	15.9	18.6	15.6	NA	NA	W	61	NNW	NNW	28	28	76	93	994.3	993	8	8	17.4	15.8	Yes
14-12-2008	Albury	12.6	21	3.6	NA	NA	SW	44	W	SSW	24	20	65	43	1001.2	1001.8	NA	7	15.8	19.8	Yes
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
16-12-2008	Albury	9.8	27.7	NA	NA	NA	WNW	50	NA	WNW	NA	22	50	28	1013.4	1010.3	0	NA	17.3	26.2	NA
17-12-2008	Albury	14.1	20.9	0	NA	NA	ENE	22	SSW	E	11	9	69	82	1012.2	1010.4	8	1	17.2	18.1	No
18-12-2008	Albury	13.5	22.9	16.8	NA	NA	W	63	N	WNW	6	20	80	65	1005.8	1002.2	8	1	18	21.5	Yes

1. Create data frame and display the information of the data set.
2. Treat NA values by mean value of the column values.
3. Display average min and average max temperature city-wise.
4. Plot bar graph to showcase min, max temperature and rainfall . (x-axis: time, y-axis: country, Title: 'Min-Max Temperature')
5. Display the details where wind direction does not get change at different observation time.

4) Why Feature Engineering is important in model building. Explain in detail any two techniques used (5)

for Feature Engineering.

A)

B) What in the One-Hot method to encode categorical features? [explain with an example] (2.5)

C) Write short note on Correlation. (2.5)

5) What is KNN machine learning algorithm? Define each step of KNN. What are the advantages and disadvantages of KNN? (5)

A)

B) We have data from the questionnaires survey (to ask people opinion) and objective testing with two attributes (acid durability and strength) to classify whether a special paper tissue is good or not. Here is four training samples (5)

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 he classification of this new tissue Bad or Good. [Consider K=3]

-----End-----