Question Paper

Exam Date & Time: 18-Jan-2024 (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 - S2]

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) Define a python function to check whether a number (user-defined) is palindrome or no palindrome. The function raises an exception if the argument is not a number. Call the function by call by user-specified value and print the result, and accept the exception in case the user inputs non-integer data. (5)

C) Write python code to create the following DataFrame. Insert a new column 'Sum' and assign row-wise sum to 'Sum'. (2.5)

	one	two
a	1.0	4.0
b	2.0	3.0
c	3.0	2.0
d	4.0	1.0

2) A) Write a program to read text file data and create a dictionary of all keywords in the text file. The program should count how many times each word is repeated inside the text file and then find the keyword with a highest repeated number. The program should display both the keywords dictionary and the most repeated word. (5)

B) Write python code : (5)
i) to find all words with even length in a given string.
ii) to replace all occurrences of space, comma, or dot with a colon.
iii) to find all words having substring "the" in a string.
iv) to separate and print the numbers of a given string.
v) to reverse a string.

3) A) Consider Cricket data set and answer the following: [assume Country_Name column which contains cricketers' country name] (6)
a) Convert data table into data frame and display the description. Display the length of the dataset. Display first 10 records of the dataset. Display last five records of the data set.
b) i) Display the name of the player/players who played minimum match, minimum innings, scored minimum runs(non zero).
ii) Display total number of players and their average performance, country-wise.
c) Display the longest span player's details. Display the number of players from this country .
d) Display the details of the players who are currently active.
e) Display the details of the players having 'M'/ 'm' as one of the character in their name.

c) Display the details of the players having 'm' or 'M' as one of the characters in their name.

	Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100	50	0	4s	6s
0	SR Tendulkar (INDIA)	1989-2013	664	782	74	34357	248	49	50817	67.6	100	164	34	4076	264
1	KC Sangakkara (Asia/ICC/SL)	2000-2015	594	666	67	28016	319	47	42086	66.6	63	153	28	3015	159
2	RT Ponting (AUS/ICC)	1995-2012	560	668	70	27483	257	46	40130	68.5	71	146	39	2781	246
3	DPMD Jayawardene (Asia/SL	1997-2015	652	725	62	25957	374	39	40100	64.7	54	136	47	2679	170
4	JH Kallis (Afr/ICC/SA)	1995-2014	519	617	97	25534	224	49	45346	56.3	62	149	33	2455	254
5	R Dravid (Asia/ICC/INDIA)	1996-2012	509	605	72	24208	270	45	46564	52	48	146	21	2604	66
6	V Kohli (INDIA)	2008-2022	473	527	77	24130	254	54	30483	79.2	71	125	33	2400	258
7	BC Lara (ICC/WI)	1990-2007	430	521	38	22358	400	46	32839	68.1	53	111	33	2601	221
8	ST Jayasuriya (Asia/SL)	1989-2011	586	651	35	21032	340	34	25910	81.2	42	103	53	2486	352
9	S Chanderpaul (WI)	1994-2015	454	553	94	20988	203	46	40150	52.3	41	125	21	2041	126
10	Inzamam-ul-Haq (Asia/ICC/P	1991-2007	499	551	76	20580	329	43	32172	64	35	129	35	2076	193
11	AB de Villiers (Afr/SA)	2004-2018	420	484	68	20014	278	48	26787	74.7	47	109	20	2004	328
12	CH Gayle (ICC/WI)	1999-2021	483	551	35	19593	333	38	25370	77.2	42	105	44	2332	553
13	HM Amla (SA/World)	2004-2019	349	437	36	18672	311	47	28718	65	55	88	19	2138	93

B) Describe 'NAN' in python. What are the different ways to clean data? Support your answer with example codes. (4)

4) What are the ways to encode categorical features? Discuss their pros & cons. (2.5)

A)

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

C) Write short note on interaction features. (2.5)

5) What is KNN machine learning algorithm? Define each step of KNN. What are the advantages and disadvantages of KNN? (4)

A)

B) Derive the linear regression model using the following data: (6)

STUDY HOURS (X)	TOPICS SOLVED (Y)
1	1.5
1.2	2
1.5	3
2	1.8
2.3	2.7
2.5	4.7
2.7	7.1
3	10
3.1	6
3.2	5

3.6 8.9

Use derived model to predict topic covered in 8 hours.

-----End-----