Unique Paper Code: 42343307
Name of the Course: B Sc Programme / B Sc Mathematical Science SEC
Name of the Paper: Data Analysis using Python Programming
Semester: III
Year of Admission: 2019
Duration: 1 Hour Maximum Marks: 25

Instructions for Candidates
Attempt any two questions. All questions carry equal marks.

Q1. Consider the salary in INR lakhs of 10 employees of a company:

```
Salary = [1.50, 0.80, 1.45, 7.0, 7.70, 0.75, 1.10, 1.50, 1.10, 1.30 ]
```

- Import appropriate Python libraries and give Numpy commands to display the mean and median of the given data.
- Mention a Numpy command to display the salaries less than the median salary.
- What is a box plot? Give a command to visually represent the above data as a boxplot.
- Write command(s) to display the interquartile range (IQR) of the given data.
- What are outliers and how are they determined in the data? Mention any two means to handle the outliers.

Q2. Consider the following data for weather for 15 instances where the different environment conditions help decide whether the weather is suitable to play.

S.no.	Outlook	Temperature	Humidity	Wind	Play
1	Rainy	NULL	NULL	NULL	No
2	Rainy	Hot	90	Yes	No
3	Overcast	Hot	86	No	Yes
4	Sunny	Mild	96	No	Yes
5	Sunny	Cool	80	No	Yes
6	Sunny	Cool	70	Yes	No
7	Overcast	Cool	65	Yes	Yes
8	Rainy	Mild	95	No	No
9	Rainy	Cool	NULL	No	Yes
10	Sunny	Mild	80	No	Yes
11	Rainy	Mild	70	Yes	Yes
12	Overcast	Mild	90	Yes	Yes
13	Overcast	Hot	75	No	Yes
14	Sunny	Mild	91	Yes	No
15	Rainy	NULL	70	Yes	Yes

- What is the role of pre-processing techniques in data analysis?
- Assume that the data is loaded in a Pandas dataframe `df`. Write a command to drop the data instance 1.
- For data instance 9, it is intended to fill in the missing value by a reasonable value. Give the necessary command(s) to replace the attribute having NULL value accordingly. Give reason for your choice.
- For data instance 15, it is intended to fill in the missing value by a reasonable value. Give the necessary command(s) to replace the attribute having NULL value accordingly. Give reason for your choice.
- Give a Pandas command to convert the categorical attribute, `Play` into dummy variables.
- From the dataframe `df`, drop the column `Play` and include the dummy variables as new columns.

Q3. Consider the given data with two numerical attributes, *`Hours_studied`* and `Marks_obtained` loaded in a dataframe `df`.

Hours_studied	3	4.5	9	10	1.5	2	8	1	11	6
Marks_obtained	45	50	90	95	33	20	89	0	80	72

- Give a Pandas command for computing the covariance among *`Hours_studied`* and *`Marks_obtained`*.
- What is meant by correlation between a pair of numerical attributes?
- Give a Pandas command to compute the correlation coefficient between `Hours_studied` and `Marks_obtained`.
- What is the possible range of values obtained in correlation and how is it interpreted?

Consider the following categorical data

Gender	F	M	M	M	M	M	F	F	F	F	F	F	M
Grades	B	C	C	D	B	C	B	A	A	A	B	B	D

- Describe a technique for correlating categorical data.
- Give a code segment/command to illustrate the correlation between the Gender and the Grades obtained. Show the output table.