Introduction to Data Science Using Python (CSE 3054)

MAJOR ASSIGNMENT-1

1 Objective

The objective of the assignment is to provide programming practice regarding reading datasets (using mathematical and data visualization tools) in Python.

2 Data and Problems

Consider the following list of students. student is a list of 20 dictionaries, and each dictionary contains the student's id and marks in $Math$, $Computer\ Science$, and $Science$.

student=[{"$student\_id$" : 1, "$Math$" : 50, "$Computer\ Science$" : 60, "$Science$" : 73},
{"$student\_id$" : 2, "$Math$" : 40, "$Computer\ Science$" : 50, "$Science$" : 55},
{"$student\_id$" : 3, "$Math$" : 90, "$Computer\ Science$" : 70, "$Science$" : 95},
{"$student\_id$" : 4, "$Math$" : 80, "$Computer\ Science$" : 62, "$Science$" : 72},
{"$student\_id$" : 5, "$Math$" : 80, "$Computer\ Science$" : 90, "$Science$" : 45},
{"$student\_id$" : 6, "$Math$" : 84, "$Computer\ Science$" : 90, "$Science$" : 50},
{"$student\_id$" : 7, "$Math$" : 90, "$Computer\ Science$" : 95, "$Science$" : 55},
{"$student\_id$" : 8, "$Math$" : 89, "$Computer\ Science$" : 93, "$Science$" : 53},
{"$student\_id$" : 9, "$Math$" : 88, "$Computer\ Science$" : 92, "$Science$" : 58},
{"$student\_id$" : 10, "$Math$" : 90, "$Computer\ Science$" : 95, "$Science$" : 55},
{"$student\_id$" : 11, "$Math$" : 70, "$Computer\ Science$" : 65, "$Science$" : 39},
{"$student\_id$" : 12, "$Math$" : 65, "$Computer\ Science$" : 60, "$Science$" : 35},
{"$student\_id$" : 13, "$Math$" : 60, "$Computer\ Science$" : 55, "$Science$" : 30},
{"$student\_id$" : 14, "$Math$" : 55, "$Computer\ Science$" : 57, "$Science$" : 25},
{"$student\_id$" : 15, "$Math$" : 49, "$Computer\ Science$" : 54, "$Science$" : 22},
{"$student\_id$" : 16, "$Math$" : 10, "$Computer\ Science$" : 30, "$Science$" : 11},
{"$student\_id$" : 17, "$Math$" : 50, "$Computer\ Science$" : 40, "$Science$" : 16},
{"$student\_id$" : 18, "$Math$" : 90, "$Computer\ Science$" : 45, "$Science$" : 80},
{"$student\_id$" : 19, "$Math$" : 70, "$Computer\ Science$" : 50, "$Science$" : 39},
{"$student\_id$" : 20, "$Math$" : 70, "$Computer\ Science$" : 80, "$Science$" : 75}]

1. Write a Python program to group the student ids corresponding to the following $Science$ mark.
   - less or equal to 30.
   - between 30 and 70.
   - more than 70.
2. Create a line chart for student ids and $Computer\ Science$ marks. Student id on the x-axis and $Computer\ Science$ mark on the y-axis.
3. Write a Python program to create a list of 20 vectors by taking student's $Math$, $Computer\ Science$ and $Science$ marks. Find the sum of all these 20 vectors. After that, find the average marks for $Math$, $Computer\ Science$ and $Science$.
   Hint: [50,60,73] will be one vector.
4. Write a Python program to find the mean, median and mode of $Computer\ Science$ marks.

5. Write a Python program to find the covariance of $Math$ and $Science$ marks.
6. Write a Python program to find the correlation between $Computer\ Science$ and $Math$ marks.

3 Mark Distribution

- Problem-1 – [2 marks]
- Problem-2 – [2 marks]
- Problem-3 – [2 marks]
- Problem-4 – [2 marks]
- Problem-5 – [2 marks]
- Problem-6 – [3 marks]