COMP20008 2020S1 workshop - week 7
Correlation

Consider the following hypothetical dataset providing measurements for *Average Steps per day* and *Average Resting Heart Rate*, across a sample of 12 people.

Part A

Person ID	**Average Steps per day**	**Average Resting Heart Rate**
1	1000	100
2	2500	105
3	3000	80
4	5000	77
5	6000	74
6	9000	70
7	11000	65
8	14000	63
9	18000	62
10	19000	61
11	19500	60.5
12	22000	55

1. Compute the Pearson correlation between *Average Steps per day* and *Average Resting Heart Rate*. Show your working. How would you interpret this correlation value?

2. Based on the Pearson correlation value, can one conclude that doing more steps per day will cause one's average resting heart rate to decrease? How else might it be interpreted?

3. Discretise the data as follows: Apply 3 bin equal frequency discretisation to *Average Steps per day* and 4 bin equal frequency discretisation to *Average Resting Heart Rate*. Show the values of the discretised features.

4. Using the discretised features, compute the entropies: *H(Average Steps per day)*, *H(Average Resting Heart Rate)*, *H(Average steps per day | Average Resting Heart Rate)*, *H(Average Resting Heart Rate | Average Steps per day)*.

5. Using the above information, compute the mutual information between *Average Steps per day* and *Average Resting Heart Rate*.

Part B

Use Python code to verify the calculations of (1) Pearson correlation, (4) entropy, and (5) mutual imformation in **Part A**.