Practical Data Science/Analytics (Preprocessing)

Write R Scripts or use R to perform any mathematical operations while solving the following problems.

Problem 1: Employees who have similar profiles

You are given a sample data set of employee details. Assume that two employees are similar if the Euclidian distance between any two employees is very less. You can represent each employee as a vector/point in 3-dimensional space.

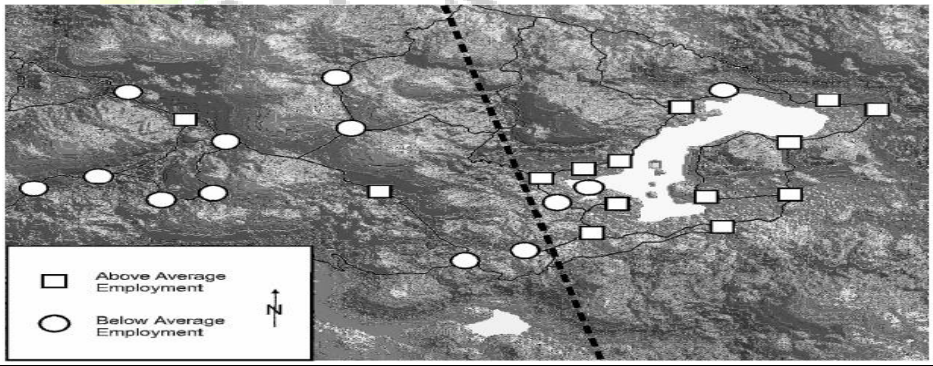
Empid	Salary	Age	Experience
1	25000	24	4
2	40000	27	5
3	55000	32	7
4	27000	25	5
5	53000	30	5
6	26000	35	10

Find the employees who have similar profiles,

a. without normalizing the data
b. with 0-1 normalization
c. with Z-score normalization

Problem 2: Living place vs Education

The map below shows employment levels for two groups of villages in India: those near the lake and those in the mountains (the dashed line separates these two groups). You have to determine whether there is a difference in employment between these two groups. In other words, you have to find whether there is significant relationship between the living area and education variables.



Practical Data Science/Analytics (Preprocessing)

Problem 3: Near Highway Property valuable or not

The map below depicts the value per square foot for 7 houses in the Globeville neighborhood of Denver, Colorado. The elevated portion of I-70(National highway) is shown as a heavy black line on the map. Each house is labeled with the value per square foot and the distance from I-70 in meters. Is there correlation between property value and distance variables?