VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

Digital Assignment - I

Programme	:	**B.Tech (CSE)**	Semester	:	**WIN:2020-2021**
Course	:	**Large Scale Data Processing**	Code	:	**CSE3025**
Faculty	:	**Dr. Bharadwaja Kumar & Dr. Ramesh Ragala**	Slot	:	**G2**
Class Number	:	**CH2020215001364 & CH2020215001362**		:	

Slove the following assignment questions manually. Write (Hand written) your reg.no and name on every sheet of the answer script.

1. In this Assignment you need to understand the concept of clustering. For simplicity, you need to work in two dimensions. You have dataset shown in Table - 1 that describes the age and the total amount spend of each customers.

Age	Amount Spend
18	10
21	11
22	22
24	15
26	12
26	13
27	14
30	33
31	39
35	37
39	44
40	27
41	29
42	20
44	28
46	21
47	30
48	31
49	23
54	24

Table 1: Dataset for K-means Clustering

To improve advertising, the marketing team wants to send more targeted emails to their customers. Hence they draw the scatter plot to explore the data, which in shown in
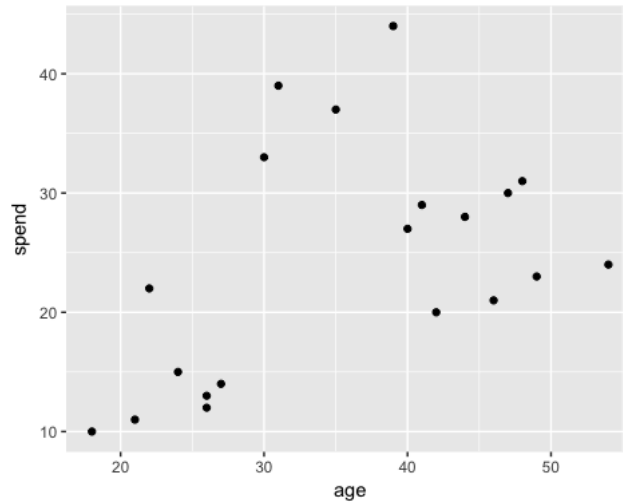
Figure 1: Scatter Plot of the dataset

Figure - 1.
A observed pattern from the scatter plot as follows:

- At the bottom-left, you can see young people with a lower purchasing power
- Upper-middle reflects people with a job that they can afford spend more
- Finally, older people with a lower budget.

Hence, you need to find the these three cluster using K-means clustering algorithm.

2. Induce a decision tree for given dataset shown in Table - 2 of learning examples (perform manual calculations supported by some tools like calculator)

Hangover	Exam	Weekend	Party
no	easy	no	yes
no	hard	no	no
no	no	no	yes
no	no	yes	yes
yes	easy	no	no
yes	hard	no	no
yes	no	no	no
yes	no	yes	no

Table 2: Dataset for Decision Tree algorithm