SLOT – L39+40

VIT
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

CYCLE SHEET – II – WINTER SEMESTER 2021-2022

Programme Name & Branch: B.Tech (CSE)

Course Name: DATAWAREHOUSE AND DATAMINING LAB

Course Code: CSE3030

Upload Source Code with Output screen in VTOP

CYCLE SHEET -2

1. Download the employee churn dataset (15000 rows and 10 columns). Perform cluster using K-Means on this data. Identify and characterize the clusters in which the churn rate is higher. (Hint: Study the variables and select the useful ones for cluster analysis. Create clusters (how many? Analyze!). Examine the rate of churn in each cluster and discuss the characteristic of the clusters.)

2. Consider the any one suitable dataset available in UCI. Find clusters in this dataset using Hierarchical clustering technique having the following properties.

Number of clusters: 3
Distance measure: Euclidean distance.
List the members of each cluster.
Draw the Cluster tree diagram

3. Download the wholesale customer dataset from UCI ML. The data set refers to clients of a wholesale distributor. It includes the annual spending in monetary units on diverse product categories. Segment the customers (find clusters in the dataset using DBSCAN clustering algorithm) for possible offers. Analyse the noise points in the dataset

4. Use the Apriori technique to determine the association rules in the AdultUCI dataset. Choose education, marital-status, occupation, race, sex and native-country as the parameters of the given dataset for analysis. Perform several iterations to determine the min_support that can be chosen if the number of association rules that can be considered for further analysis is between 25 and 30.AdultUCI dataset can be downloaded from web or construct the dataset

5. From the given data, they wish to find the items that were purchased most frequently. They also wish to determine the item(s) which encouraged the customer to purchase additional item(s). Such analysis is commonly termed as **Market Basket Analysis**, where the interesting associations between various items are determined.

The analysis that leads to determining purchase behaviour of customers arises from the items attribute. The marketing team seeks to study the items attribute more closely to determine associations between various items.

From the given sample data set, the most frequently purchased item can be determined using a frequency table, as shown below

Transaction	Seat Cover	Audio system	Car cover	Steering Cover	Toolbox	Foot mats	Mud flaps	Window tint
1	Y		Y			Y	Y	
2	Y	Y		Y		Y		
3			Y					Y
4		Y					Y	Y
5					Y			
6	Y					Y		
7		Y			Y			Y
8							Y	
9				Y		Y		
10		Y			Y			Y
11			Y			Y		
12				Y				
13	Y				Y			
14						Y		
15			Y					
Total	**4**	**4**	**4**	**3**	**4**	**6**	**3**	**4**

^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^