Name: Platham Yudau PRN: 2020 BTE ITO0006 Sub: Data Mining lab

Walchand College of Engineering, Sangli.
Experiment no: 3
For equal brownship in the
Title: Perform Binning of data
mid of the mine of a side of the first
Thiony: Data binining, also telemed to as data bucketing
is a proposition technique employed to mitigate the
influence of minor observational errors in data. The process
entails signishing original data a values into coma pact
intervals called bins, and subsequently, they att substitu-
-ted with a upresentative value computed for the specific
bin. This action imparts a smoothing effect on the
input, data potentially wroing the risk of overlitting
especially when alling with smaller datasets.
1 1
the categorization of data into bins follows two
primary methods.
where of entry in sedence a sould are
Equal Fit quency Binning! who is make & 416000 cm
The bins are disigned to contain an equivalent
number of obstructions. This ensures that the data is
distributed uniformly across the bins.
U 1
Equal width Binning
The bins are established to have uniform wight
Each bin's range is alined as [min + w], [min + ew]
The bins are established to have uniform wiath Each bin's range is affined as [min + w], [min + ew] and so forth, white 'w' stands for (man-min)
divided by the number of bins.
ı

Formula:	ngineering, Sangli.
For equal prequency binning:	
1-2 pg (1)	partial married mar
bin eix = mo data / no bins	3
if is to a state of other than it main.	promised and a special
For equal wiath binning	the formation of a state of the
trange = (man-ell - mintele) / in	
Hadre 40 met, principlation this	to , early hour simula
size of each bin	
[min+w] [min+2w]	
ratificate to that we make see	published when the
where effect entries at in	sailorn arrive pulling
binsize : size of the bin the	10 Addresion de do
nu-data: number of data	shout promin
no bins: number of bins to r	nake
w: width I tonge of data ear	ch bin can have
man ele: maximum element in de	atoms and will
min-ce: minimum element in	day looks a admir
तर १०३३ - द्वित विशिष्ट -	whatofine Baladelika
	g Signatura (Signatura
	Class might Binning
planting the property brighters	. HE INIC OF
RETURNED LOW THINK IN BURELLE	Cock bip's bear in
Cather - Meria : Act a plant in	o of ENUDY Afto os baro

TEME NAME OF STREET

COLLEGE OF	Walchand College of Engineering, Sangli
Algon	thm:
٦,	. ohrl
	The cost in custary libraries
3	Define function for Equal Frequency Binning and
	Equal unidth Minning.
4	Define a function to tead data from a CSU fill Define a function to write data from a CSU fill
5	Deline a function to write data from a CSU LIU
6	. The the main lunchion
	6.1 Read INM Vinput. COV
	6.2 prompt voir to enouse a binning method
	and the number of bins.
	6.8 Cl Equal Irequency is chosin:
	y: Persom equilteg binning and write
	y: Person equipmed binning and write output to "autput-equipmed. (sv"
21	6.4 [Equal width is chosen:
11.	y: perform equiwidth binning and with
	output to output equiwidth csu'
	6.5 Display an error mussage if an invalid
10	method is chosen.
Exan	iple:
	restruction of manifest telements is a manifest
Co	noide an array mother said the state annitano
	= [5, 10, 11, 13, 15, 35, 30]
NO	am will be binned using both method
	al Equal Frequency b) Equal Width.
	10 5 1 10 11 11

Al Fanal Estamentu	t on IT vogt
Al Equal Farquency	1 toda et
am size = 6 . n= 1	Tracis of the second of the se
bin = 8	to the morning miles
	married third lend
Bin-1: [5, 40]	a figure a product to the
Bin-2: [11,13] Bin-3: [15,85]	to में क्षिणों के मेरिया है
Bin-3: [15,85]	"ne'done" that it is a
were lun	Well forms from the
heriva parantal a signi c	
B) Equal wiath sand to	educted on
	or of the court from
amsize= 6 para bin=3	
man ele = 35 min ele =	5
w = (35-5)/3 = 30	18. = 10 1 1 h 9
sign kar printed Attento	
Bin1: (57.10) 11/13 10000	of legito
Gin 2 : 15 AC 1 1 1 1 2 W 100	* AD . R. C. C. C. C. C. C. C.
Bins: 35	VS 21 Ke Jilani
and the second second second second second	
Conclusion:	: sigmost
Binning is a usilal techni	nique for transforming
continous data unto disc	rde categories, making it
easily to analyze and v	isualiza. In all all all and
	W A
deput folk, Mithed	Aspard to the ma work
	i-ropat from in
	William invital

Code:

```
#include <iostream>
#include <fstream>
#include <vector>
#include <sstream>
#include <climits>
using namespace std;
// Equal Frequency Binning
vector<vector<int>> equifreq(vector<int> arr1, int m)
 int a = arr1.size();
 int n = a / m;
 vector<vector<int>> bins;
 for (int i = 0; i < m; i++)
  vector<int> bin;
  for (int j = i * n; j < (i + 1) * n; j++)
   if (j \ge a)
     break;
   bin.push_back(arr1[j]);
  bins.push back(bin);
 return bins;
// Equal Width Binning
vector<vector<int>> equiwidth(vector<int> arr1, int m)
 int a = arr1.size();
 int max ele = INT MIN;
 int min ele = INT MAX;
```

```
for (int i = 0; i < arr1.size(); i++)
  \max ele = \max(\max ele, arr1[i]);
  min ele = min(min ele, arr1[i]);
 int w = (max_ele - min_ele) / m;
 int min1 = min_ele;
 vector<int> arr;
 for (int i = 0; i < m + 1; i++)
  arr.push back(min1 + w * i);
 vector<vector<int>> arri;
 for (int i = 0; i < m; i++)
  vector<int> temp;
  for (int j : arr1)
   if (j \ge arr[i] & j \le arr[i + 1])
     temp.push_back(j);
  arri.push back(temp);
 return arri;
// Read data from CSV
vector<int> readCSV(string filename)
 ifstream inputFile(filename);
 vector<int> data;
 string line, value;
 while (getline(inputFile, line))
```

```
stringstream ss(line);
  while (getline(ss, value, ','))
   data.push back(stoi(value));
 inputFile.close();
 return data;
}
// Write binning outputs to CSV
void writeCSV(string filename, vector<vector<int>>> bins)
 ofstream outputFile(filename);
 for (int i = 0; i < bins.size(); i++)
  outputFile << "Bin " << i + 1 << ",";
  for (int num : bins[i])
   outputFile << num << ",";
  outputFile << "\n";
 outputFile.close();
int main()
 vector<int> data = readCSV("input.csv");
 int m;
 int method;
 cout << "Choose binning method: " << endl;</pre>
 cout << "1. Equal Frequency Binning" << endl;</pre>
 cout << "2. Equal Width Binning" << endl;
 cout << "\nEnter method number: ";</pre>
 cin >> method;
 cout << "\nEnter number of bins: ";</pre>
 cin >> m;
```

```
if (method == 1)
{
  vector<vector<int>>> freqBins = equifreq(data, m);
  writeCSV("output_equifreq.csv", freqBins);
}
else if (method == 2)
{
  vector<vector<int>>> widthBins = equiwidth(data, m);
  writeCSV("output_equiwidth.csv", widthBins);
}
else
{
  cout << "Invalid method choice." << endl;
}
return 0;
}</pre>
```

I/P

```
input.csv

1 5
2 10
3 11
4 13
5 15
6 35
7
```

a) Equal frequency

```
output_equifreq.csv

Bin 1,5,10,

Bin 2,11,13,

Bin 3,15,35,

4
```

b) Equal Width

```
output_equiwidth.csv

1    Bin 1,5,10,11,13,
2    Bin 2,15,
3    Bin 3,35,
4
```