Name:	Ayush Gut	pa					
Roll No:	12						
Class/Sem:	TE/V						
<b>Experiment No.:</b>	3						
Title:	Tutorial	on:	a)	Data	Exploration	b)	Data
	pre-proce	ssing					
<b>Date of Performance:</b>							
<b>Date of Submission:</b>							
Marks:							
Sign of Faculty:							



**Aim:** To solve problems in Data Exploration and Data Pre-processing.

**Objective:** To enable students to effectively identify sources of data and process it for data mining.

- 1. Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 36, 40, 45, 46, 52, 70.
- a. What is the mean of the data? What is the median?
- b. What is the mode of the data? Comment on the data's modality (i.e., unimodal, bimodal, trimodal, etc.).
- c. What is the midrange of the data?
- d. Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?
  - e. Give the five-number summary of the data.
  - f. Show a boxplot of the data.

#### **Solution:**

	Aim: To sake Roblems in Odo Exploration & Data Pre-goversing
	: /h . 1 II .
(i	Suppose that the data for analysis increasing arter) 13,15, 16,16,19,20
	volves for the data topies or ( ), 35, 35, 35, 35, 35, 35, 36, 36, 46, 52, 5
	guant's the mean of data curation the median ?
(	d Charles the mean of data to
	Hear, N=77  Mear Zz: 13+15+16+16+19+20+2+71+72+72+725+725+725+725  N 33+33+35+35+35+35+35+36+40+45+46+52+70
	Mean Ex: = 13+30+35+35+35+36+40+45+46+52+70
	/4
	= 809 = 29.96.
77114	27
	Me dian -> Middle value
	= 25,
	what is the mode of the data Comment on the data's modality?
$\Theta$	What is the mode of the data! Comment of the
7	Mude=most occured values In the above sequence 15 & 35 are most occured values
7	In the above sequence is 433 are most accordance
	Mode 25, 35 is bimodal
	(4.14)
(c)	what is the midrange of the data
<u>,</u>	Midrange: Ithe vale + Hax-value/2
	· 13+70/2
	,
	Midrange 41.5



	the dato
-	the dito!
	9. First Quartile = Middle value of 13,15, K, 16, 19, 20, 20, 21, 22, 27,26
	Q <sub>1</sub> : 20 ,,
	8, =1mrd Quartile= Hiddle value of 30,33,35,35,35,35,35,35,36,46,45,45,20
	930 35,,
. (	Give the Grumber summary of the data -
(	Minimum value = 13
	Frist Quantile=20
	Medfor Value 25 Third Quartik = 35
	Horimm volue- 70
A	Show a barplet of the data
9	Thou a property of
	/
	(3 25 35
	0 5 10 15 20 25 30 35 40 45 50 55 60 65 40 75 80
-	

2. Suppose that the values for a given set of data are grouped into intervals. The intervals and corresponding frequencies are as follows:

age	frequency
1–5	200
6–15	450
16–20	300
21–50	1500
51–80	700
81–110	44

Compute an approximate median value for the data.

#### **Solution:**



	n=3194
	n/2=1597
	The state of the politicania
-	This observation le between the Marie Internal
	21-60 which is the median klass.
	lower dose limit = 21 = (1)
-	class size (h) = 30
	facquercy of the median class (+)=1500
	median slaw (cr)-050
.0	median= 1+(1/2-cf) xh = 21+ (1597-950) X30 = 21+1294 = 33.9
	Median = 33.94
ram*	FOR EDUCATIONAL USE

3. Consider the data given below and compute the Euclidean distance between each point. P1 (0,2), P2(2,0), P3(3,1) and P4(5,1).

#### **Solution:**

	Consider of Guclidean of	, P2(s	2,0), F	3(3,1)	& P4(5,	D.	
7 00	io17:- Asiale						
	PI	0	2				
	P2	2	0				
	Р3	3	1				
	P4	5	1				

							-
1/4	w= (3)	(x:-4;)	2)/2 =	2.16	(1-yi) <sup>2</sup>		
acas	32-(81	(,,	12	1=17	2 - [10 -	2 (2-0)2	
:. e d	(P, P2)	=1(x'-	x2)+1	y,-ya)	= 110-3	1)2 (2-0)2	
	(000)	= 144	H = .	18 =	2.838		
	(P, F2)	- 4-1					0
: d(P.	P3)= V(x	1-X3)2+(y	1-407=	10-3	2+(2-1)2=	J9+1 = J10=	2.
		. 221	4	- 110-5	12+12-19=1	25+1=126=	5.
:.d(P	P4)= V(x	(1-X11)+1	4-40	= 4600			-
- ALP	R)= 1/x	-x <sub>2</sub> ) <sup>2</sup> +(1	12-43)2:	- 1(2-3)	7-(0-02=11	1+1= 12=1.41	4.
- dir	13/			0 -		Total To	3.
·. dle	Pw = 16	(2-X4)2+	(y2-y4)	)==1(	2-5)+(0-1)	= 19+1= 14=	
. 1/0	0) [	/w ~ \ <sup>2</sup>	· ( )	12 - [a	-5)2+(1-1)2=	J22 = 24	
- dla	Py JE V	(13-14)	143-94	) 40			
		PI	P2	P3	рч		-
	PI	0	2.828	3-16	5.09		
	P2	2.828	0	1-414	3.16		-
	РЗ	3.16	1-414	2	0		
	PH	5.09	3-16		0		
	Y A 5 +						
	4 +						
	2+	_					
	9	2/02)		-			
	3 + 2 + F	10000	010	(1.)	121)		
	2 + 6	1(0,5)	· 6 (3	9, 60	4(31)		
	2 + 5	1 +	B(2,0)	1 1	((5,1) →×		
	2 + 1	1 1	B(0,0)	1 1	ATIONAL USE		

4. Suppose that the minimum and maximum values for the attribute income are \$12,000 and



\$98,000 respectively. Normalize income value \$73,600 to the range [0.0, 1.0] using min-max normalization method.

5. Partition the given data into bins of size 3 using equi-depth binning method and perform smoothing by bin mean, bin median and bin boundaries. Consider the data: 2, 10, 18, 18, 19, 20, 22, 25, 28.

#### Solution:

6	fartition the given data ento bins of size 3 using equi-dept
	binning nuther & perform smoothing by bin mean, bin me
	and her boundances. Consider the data: 2,10,18,18,19,20,22,25,
->	Oata !- 2,10,18,18,19,20,22,25,28.
	BPO size = 3
	ids data is already souted in increasing order divide the
	data into bins of size 3.
	Bin 1:- 2,10,18
	Bin 2:- 18, 19, 20
	8°n 3!- 22,25,28
	D)
-	smoothing by bin mean.
	8°n 1 :-10, 10, 10
	Bin 2:-19, 19, 19
	Bin 3: - 25,25.
+	smoothing by bin median
	Bin 1:-10, 10, 10
	Bin 2:- 19, 19, 19
	8in 3:- 25, 25, 25
	the transfer of the second
	Imoothing by bin boundaries.
-	Bm 1:- 2,2,18
	Bin 21-18,18, \$20
	8 m 3: 22, 22, 28
	THE PERSON HERE THE TAX THE WAY OF THE PERSON HERE.

