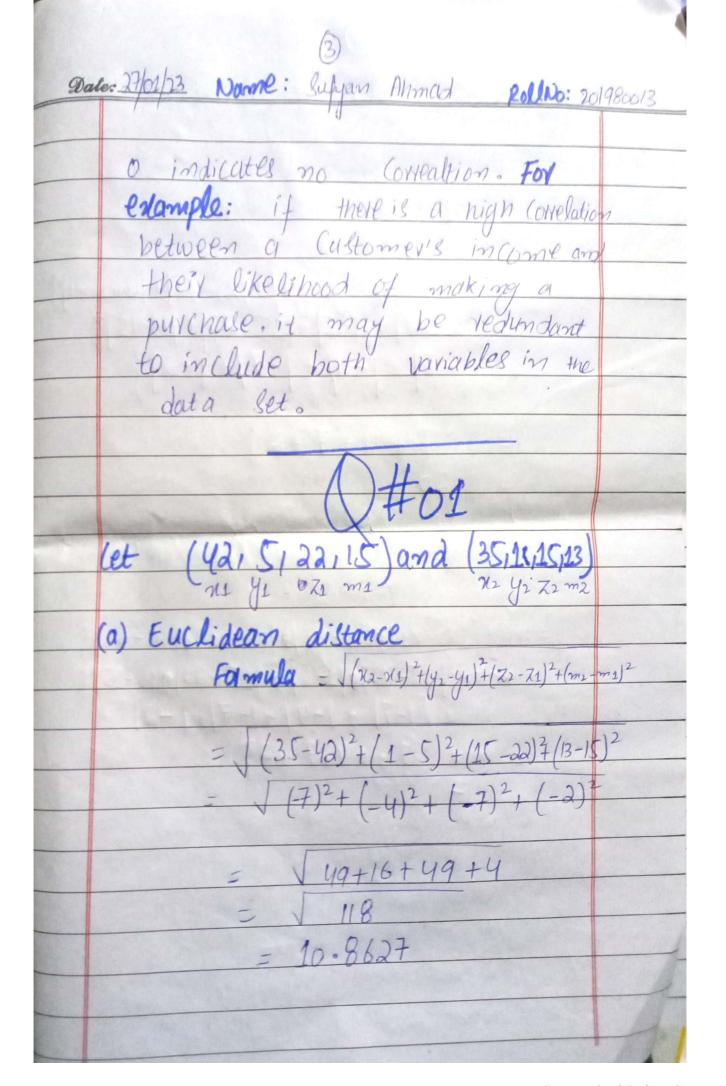
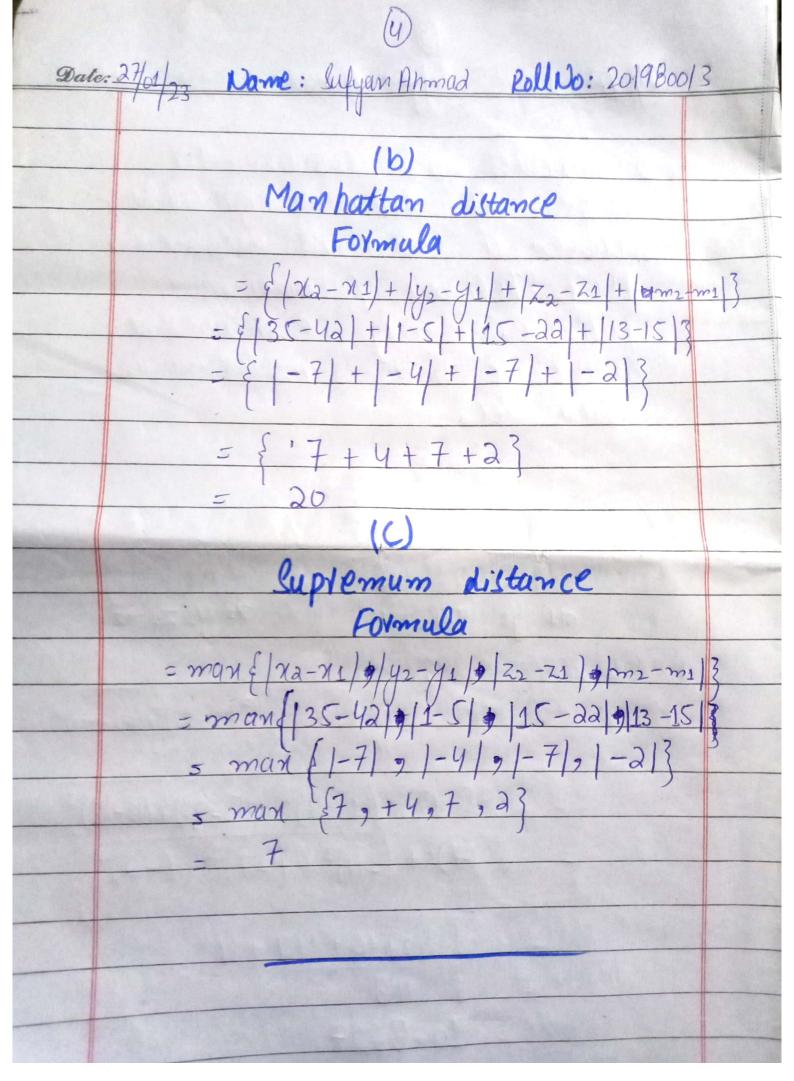
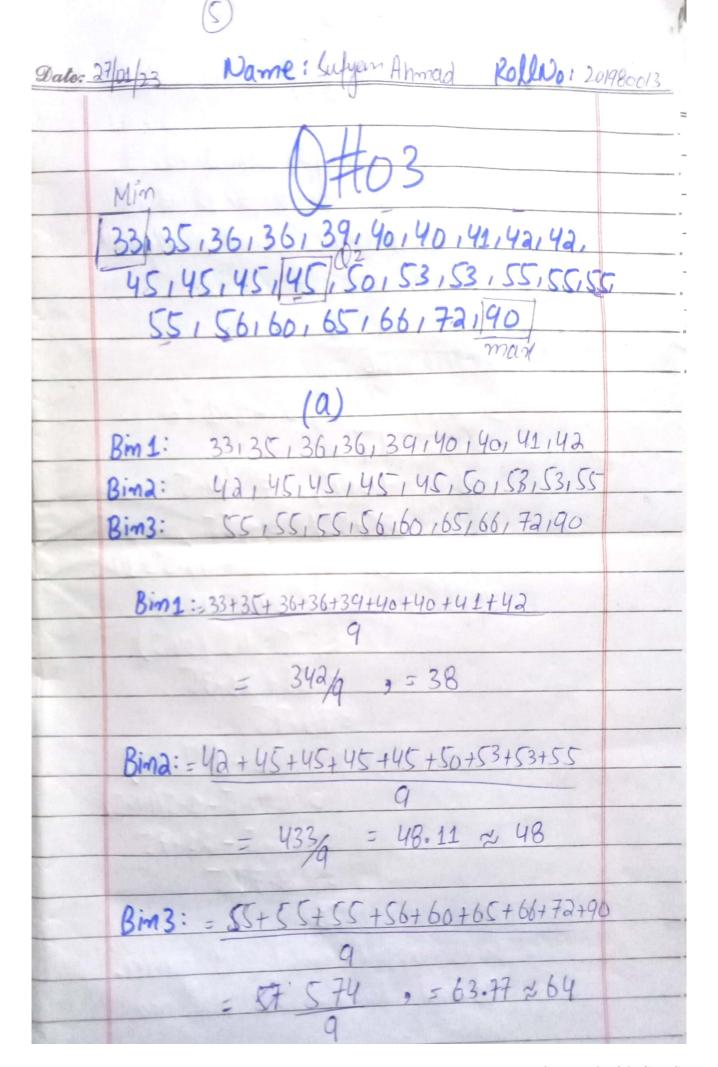
Date: 27/01/73	
Assignment#02	
Name Sufyan Ahmad	,
Roll00 201980013	
Subject Data Mining	
Section	
(1)#02	
(a)	
The entities (iden) identification	•
problem in data integration regers	
to the process of identifying and	
resolving inconsistencies in the naming	
and representation of entitles across	
different data souvces for example: In	
à data integration scenario involving	
a Company's customer information	
Stove a multiple databases the entity	
"Customer" may be represented	
differently in each database (e.g.	
"patron" in another). I dentifying	
and resolving these inconsistencies.	
anto le souring	







	6	
Dales 27	101/23 Name: Sufyan Ahmed Rollio: 201980013	
	The state of the s	
	Bin1: 38,38,38,38,38,38,38,38,38	
	Sina: 48,48,48,48,48,48,48,48,48	4
	Sin 3: 64,64,64,64,64,64,64,64,64	
	(b)	
	min = 33	
	max = 90	
	02: 45	
1	$\frac{01: 40+41/2 = 40.5}{03: 55+55/2 = 55}$	
	10R: Q3-Q1 = 55-40.5 = 14.5	Mary or
	40901-limit = Q3+1.5(10R)	
	= 55+1.05 (14.5)	
	= 55+21.75	
	=76.75	
	lower-limit: Q1-1.5(14.5)	
	= \$40.5-1.5(14.5)	
	= 40.5-21.75 = 18.735	
	Outliner = 90	
	. It is upper-limit is greater.  go and lower-limit has no	
	oullinear.	

	244	
Dale:	27/04/22 Name: Sufyan Ahmad Rollino: 20148001.	
	10 Holling: 20148001	3
	(C)	
	min-max normalization	
	min=33,  max=90, v=55	
	V'= V-mina (never man - new mina) + new many - mina	
	P	
	= 55 - 33 (1.0 - 0.0) + 0.0	
	90 - 33	
	= 22 (1.0)+0.0	
	57	
	= 23 (1.0) +0.0	
	= 0.3859 (1.0)	
	min-max = 0.3859	
	(d)	
	V= 35	
	V'= V - PA 6 A	
	6 A	
	V= 33+35+36+36+39+40+40+412+42+42 + 45+45+45+45+50+53+53+55+55755+55	
	+ 45+ 45+45+45+490	B
	2+	

