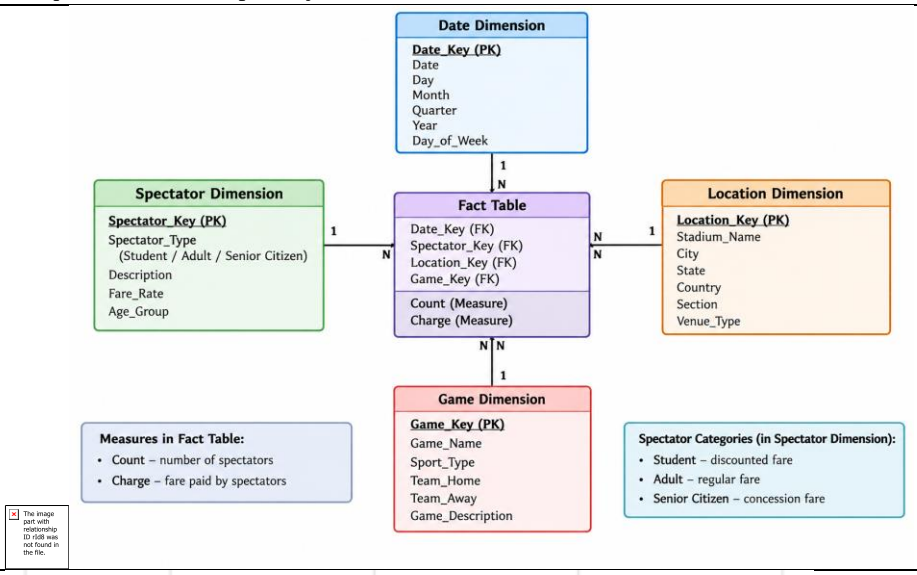
SIES Graduate School of Technology (Autonomous) NAAC A+ RISE WITH EDUCATION (Affiliated to University of Mumbai)	***End Semester Examination (R-24/R-25) SH 2026 Answer Key with marking scheme***
Branch:	Course:
Year/ Semester:	Course code:
Time: 03 hours	Marks: 80

		Marks
Q. 1	All questions are compulsory.	5
A.		
B.		5

	Height	Weight	Age
Jack	1.00	0.6	1.8
Jill	1.00	-1.9	-0.7
Hansel	-0.67	0.6	-0.5
Gretel	-1.00	-0.1	-0.3
Humpty	0.33	0.6	-0.4
Mean	**0.13**	**-0.1**	**0.0**
STD	**0.93**	**1.05**	**1.04**

$$Z - Score = \frac{(X - Mean)}{SD}$$

<table>
<tr><td>C.</td><td>

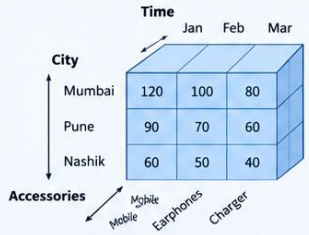
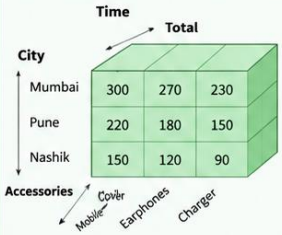
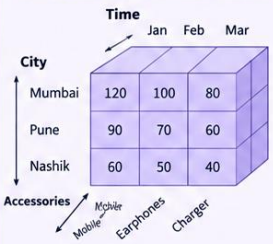
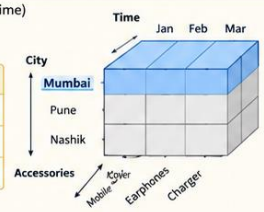
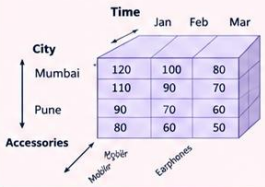
1. Confusion Matrix

Actual / Predicted	Predicted Positive	Predicted Negative	Total
Actual Positive	TP = 40	FN = 5	45
Actual Negative	FP = 10	TN = 50	60
Total	50	55	105

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{Precision} = \frac{TP}{TP+FP} \quad \text{Recall} = \frac{TP}{TP+FN}$$

$$= \frac{40+50}{40+50+10+5} = \frac{90}{105} \times 100 \quad = \frac{40}{40+10} \times 100 = \frac{40}{50} \times 100 \quad = \frac{40}{40+5} \times 100 = \frac{40}{45} \times 100$$

</td><td>5</td></tr>
<tr><td>D.</td><td>Marking Scheme: Definition of outliers – 1 mark; explanation of any one outlier detection method – 2 marks; suitable example with identification of an outlier – 2 marks.</td><td>5</td></tr>
<tr><td>E.</td><td>Marking Scheme: Definition of multilevel association rules – 1 mark; explanation of different levels of concept hierarchy – 2 marks; suitable example demonstrating an association rule at different levels – 2 marks.</td><td>5</td></tr>
<tr><td>F.</td><td>Marking Scheme: Definition of the PageRank algorithm – 1 mark; explanation of its need or purpose – 1 mark; explanation of how PageRank evaluates incoming links and their importance – 2 marks; suitable example illustrating the ranking of web pages – 1 mark.</td><td>5</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td>Q.2
A</td><td>Neat and correctly labelled diagram of the three-tier architecture – 2 marks; explanation of the bottom tier (data warehouse database and data sources) – 1 mark; explanation of the middle tier (OLAP server) – 1 mark; explanation of the top tier (front-end tools for querying, reporting, and analysis) – 1 mark.
Definition of a data warehouse – 1 mark; definition of a data mart – 1 mark; differentiation based on scope, size, data sources, users, and implementation cost or complexity – 3 marks (any three relevant differences, 1 mark each).</td><td>10</td></tr>
<tr><td colspan="3">OR</td></tr>
</table>

Q.2B .	**Original 3-D Data Cube (Total Sales Quantity)**  **Original Data (for reference)** City | Accessories | Jan | Feb | Mar Mumbai | Mobile Cover | 120 | 100 | 80 Mumbai | Earphones | 110 | 90 | 70 Mumbai | Charger | 80 | 70 | 60 Pune | Mobile Cover | 90 | 70 | 60 Pune | Earphones | 80 | 60 | 50 Pune | Charger | 70 | 50 | 40 Nashik | Mobile Cover | 60 | 50 | 40 Nashik | Earphones | 50 | 40 | 30 Nashik | Charger | 40 | 30 | 20 **Dimensions** • **City:** Mumbai, Pune, Na • **Accessories:** Mobile Co Earphones, Charger • **Time:** January, February **Measures** • Total Sales Quantity **1. Roll-up** **Roll-up along Time** (sum of months → total per city and accessory)  **Result after Roll-up (Time → Total)** City | Accessories | Total Mumbai | Mobile Cover | 300 Mumbai | Earphones | 270 Mumbai | Charger | 230 Pune | Mobile Cover | 220 Pune | Earphones | 180 Pune | Charger | 150 Nashik | Mobile Cover | 150 Nashik | Earphones | 120 Nashik | Charger | 90 **2. Drill-down** **Drill-down on City** (from total to individual cities)  **Result after Drill-down** City | Accessories | Jan | F Mumbai | Mobile Cover | 120 | 1 Mumbai | Earphones | 110 | 9 Mumbai | Charger | 80 | 7 Pune | Mobile Cover | 90 | 7 Pune | Earphones | 80 | 6 Pune | Charger | 70 | 5 Nashik | Mobile Cover | 60 | 5 Nashik | Earphones | 50 | 4 Nashik | Charger | 40 | 3 **3. Slice** **Fix City = Mumbai** (2-D slice: Accessories × Time) **City = Mumbai** Accessories | Jan | Feb | Mar Mobile Cover | 120 | 100 | 80 Earphones | 110 | 90 | 70 Charger | 80 | 70 | 60  **4. Dice** **Subset:** Cities = {Mumbai, Pune} and Accessories = {Mobile Cover, Earphones}  **Result after Dice** City | Accessories | Jan | F Mumbai | Mobile Cover | 120 | 1 Mumbai | Earphones | 110 | 9 Pune | Mobile Cover | 90 | 7 Pune | Earphones | 80 | 6 **5. Pivot (Rotate)** Rotate the cube to make Accessories the rows and City the columns (Time as pages/levels)  **January** Accessories | Mumbai | Pune | Nashik Mobile Cover | 120 | 90 | 60 Earphones | 110 | 80 | 50 Charger | 80 | 70 | 40 **February** Accessories | Mumbai | Pune | Nashik Mobile Cover | 100 | 70 | 50 Earphones | 90 | 60 | 40 Charger | 70 | 50 | 30 **March** Accessories | Mumbai | Pune Mobile Cover | 80 | 60 Earphones | 70 | 50 Charger | 60 | 40	10
Q.3 A	Slope = 7.00 5 Marks Intercept = 1.33 5 Marks	10
	OR	
Q.3B	<table><tr><th>#No. of Pictures</th><th># No. of Paragraphs</th><th>Euclidean Dist</th><th>Rank minimum distance</th><th>Include in 3-nearest neighbors</th><th>Label</th></tr><tr><td>10</td><td>2</td><td>2.83</td><td>2</td><td>Yes</td><td>Y</td></tr><tr><td>12</td><td>3</td><td>4.12</td><td>3</td><td>Yes</td><td>Y</td></tr><tr><td>9</td><td>2</td><td>2.24</td><td>1</td><td>Yes</td><td>Y</td></tr><tr><td>0</td><td>10</td><td>10</td><td>6</td><td>No</td><td>N</td></tr><tr><td>4</td><td>8</td><td>5.66</td><td>4</td><td>Yes</td><td>N</td></tr><tr><td>3</td><td>11</td><td>8.60</td><td>5</td><td>No</td><td>N</td></tr></table> **Test Query:** Pictures=8, Paragraphs=4 Label =Y as out of 4, three labels are Y.	10

Q.4 A — 10

L2

Itemset	Sup. Count
Buns, Hotdogs	2
Chips, Coke	3
Chips, Hotdogs	2
Coke, Hotdogs	2

C3

Itemset	Sup. Count
Chips, Coke, Hotdogs	2

Generate C3 from L2 and scan D for count of each candidate

L3

Itemset	Sup. Count
Chips, Coke, Hotdogs	2

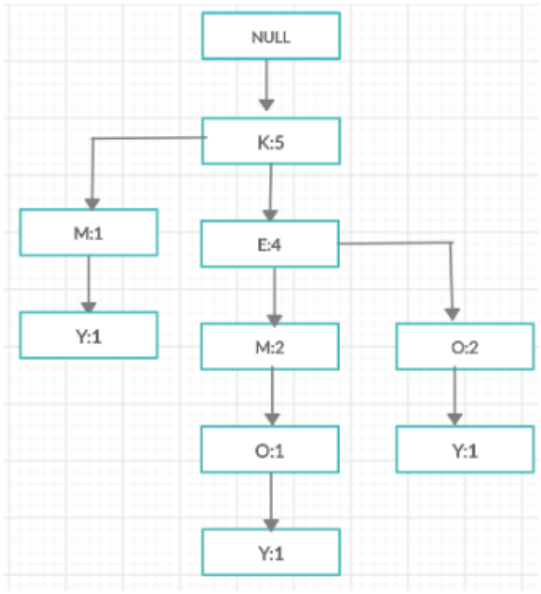
Compare candidate sup. count with min. Sup. count

Association Rule	Confidence	Confidence (%)
Chips ^ Coke→ Hotdogs	C(Chips, Coke, Hotdogs)/C(Chips, Coke)=2/3	67%
Chips^ Hotdogs →Coke	C(Chips, Coke, Hotdogs)/C(Chips, Hotdogs)=2/2	100%
Coke ^ Hotdogs→ Chips	C(Chips, Coke, Hotdogs)/C(Coke, Hotdogs)=2/2	100%
Chips → Coke ^ Hotdogs	C(Chips, Coke, Hotdogs)/C(Chips)=2/4	50%
Coke → Chips ^ Hotdogs	C(Chips, Coke, Hotdogs)/C(Coke)=2/3	67%
Hotdogs → Chips ^ Coke	C(Chips, Coke, Hotdogs)/C(Hotdogs)=2/4	50%

OR

Q.4B — 10

Final FP Tree



Q.5 A — 10

Definition of clustering – 2 marks; explanation of applications of clustering with suitable examples – 3 marks; classification of clustering methods based on different parameters, such as partitioning, hierarchical, density-based, grid-based, and model-based methods – 3 marks; suitable examples for the different clustering methods – 2 marks.

OR

Q.5B — 10

A	0						
B	1	0					
C	2	1	0				
D	3	2	1	0			
E	6	5	4	3	0		
F	7	6	5	4	1	0	
G	14	13	12	11	8	7	0
	A	**B**	**C**	**D**	**E**	**F**	**G**

A	A,B	Border	D	C,D	Border
B	A,B	Border	E	E,F	Noise
C	B,C,D	Core	F	E,F	Noise
			G	G	Noise

Q.6 A — 10

a) Data Cleaning – Definition and explanation with a suitable example: 2 marks.
b) Data Integration – Definition and explanation with a suitable example: 2 marks.
c) Data Reduction – Definition and explanation with a suitable example: 2 marks.
d) Data Transformation – Explanation with a suitable example: 2 marks, including feature scaling and normalization.
e) Data Discretization – Definition and explanation with a suitable example: 2 marks.

OR

Q.6B — 10

Binary Variables: Explanation of similarity and dissimilarity measures with a suitable example – 2 marks.
Nominal Variables: Explanation of similarity and dissimilarity measures with a suitable example – 2 marks.
Ordinal Variables: Explanation of similarity and dissimilarity measures with a suitable example – 2 marks.
Interval-Scaled Variables: Explanation of similarity and dissimilarity measures with a suitable example – 2 marks.
Ratio-Scaled Variables: Explanation of similarity and dissimilarity measures with a suitable example – 2 marks.

*********** All the Best***********