

חלק ב' [60 נקודות]

נתונה טבלת סיכום של נתוני קניות. כל טרנזקציה מייצגת קניה אחת. העמודות הפריטים בטבלה מציינות את הנוכחות הפריטים בקניה. למשל, עפ"י השורה הראשונה בטבלה, בקניה/טרנזקציה אחת נקנו לחם, בירה, קטשופ, חלב וציפס.

ID	Bread (Br)	Beer (B)	Ketchup (K)	Milk (M)	Cola (C)	Chips (Ch)
1	1	1	1	1	0	1
2	0	1	0	0	1	1
3	1	0	0	0	1	1
4	0	0	0	1	1	0
5	1	0	1	0	1	1
6	1	1	1	0	0	1
7	1	1	0	0	1	0
8	1	1	0	1	1	1
9	1	0	1	1	1	0
10	1	0	1	0	0	0
11	0	0	1	1	0	0

א. יש לחשב את ה-support עבור זוג הפריטים הכי נדיר ואת ה-support עבור זוג הכי שכיח. יש לסמן תשובה אחת בלבד בכל שורה. 10 נקודות.

זוג הכי נדיר	36%	27%	18%	0%	תשובה ←
זוג הכי שכיח	73%	45%	36%	27%	תשובה ←

ב. בהינתן $\text{min support} = 30\%$ מהו גודל של קבוצת הפריטים השכיחה הכי גדולה? יש לסמן תשובה אחת בלבד. 5 נקודות.

תשובה ←	4	1	3	2
---------	---	---	---	---

ג. מהם ה-confidence וה-support של החוק הבא: $\text{Bread} \rightarrow \text{Beer} \wedge \text{Chips}$? יש לסמן תשובה אחת בלבד בכל שורה. 10 נקודות.

support	18%	27%	0%	100%	תשובה ←
confidence	38%	63%	75%	50%	תשובה ←

7. מהו העץ (FP-tree) הנכון עבור 3 טרנזקציות הראשונות? יש לסמן תשובה אחת בלבד. 5 נקודות.

תשובה	FP-tree
1	
2	
3	
4	