

תיאור התוכנית:

יצרתי מטריצה בגודל 5000x6 הכוללת 0 ו 1.

לעמודה הראשונה נתתי 60% להיות 1, לעמודה השנייה 50%, ל 3,4,5 40%, ולעמודה 6 30%.

בחרתי מתוך המטריצה רק את 10 השישיות השכיחות (סה"כ 1,500 שורות מתוך המטריצה) וזאת מתוך כוונה לצמצם את כמות החישובים (ראה הערה למטה).

ראיתי שבכל 10 השישיות השכיחות יש 1 בעמודה הראשונה וברובן יש 1 בשתי העמודות הראשונות ולכן בדקתי רק תלות של שאר השישיות (עמודות 2-6 ו 3-6) במקרים אלו (שיש 1 בעמודה הראשונה ושיש 1 בעמודות הראשונה והשנייה).

התלות הכי גבוהה שנמצאה ($lift=1.1113$) היא לקבוצה הכוללת 1 בעמודות 3,5 בקבוצה שכוללת 1 בעמודה 1.

חישבתי support, confidence, lift לכל אחת מ 10 השישיות, על פי נוסחאות המדדים.

התקבלו סה"כ 11 תוצאות של מדד lift גדול מ 1. אני מעריך שאלו אכן התלויות הבולטות.

להלן התוצאות:

```
Top 10 most frequent tuples with at least two 1s:
Dependency of the combination in columns 2-6 in case the first column is 1
Empirical probability of 1 in first column: 0.5934

(1, 1, 0, 0, 0, 0): count = 209, support = 0.0418, confidence = 0.0704, lift = 1.0092
(1, 1, 0, 0, 1, 0): count = 168, support = 0.0336, confidence = 0.0566, lift = 1.0931
(1, 0, 0, 1, 0, 0): count = 165, support = 0.0330, confidence = 0.0556, lift = 1.0736
(1, 1, 0, 1, 0, 0): count = 164, support = 0.0328, confidence = 0.0553, lift = 1.0312
(1, 0, 0, 0, 1, 0): count = 162, support = 0.0324, confidence = 0.0546, lift = 0.9750
(1, 0, 1, 0, 0, 0): count = 142, support = 0.0284, confidence = 0.0479, lift = 0.9971
(1, 1, 1, 0, 0, 0): count = 137, support = 0.0274, confidence = 0.0462, lift = 0.9540
(1, 0, 1, 0, 1, 0): count = 122, support = 0.0244, confidence = 0.0411, lift = 1.1113
(1, 1, 0, 1, 1, 0): count = 122, support = 0.0244, confidence = 0.0411, lift = 1.0436
(1, 0, 0, 1, 1, 0): count = 119, support = 0.0238, confidence = 0.0401, lift = 1.0445

Dependency of the combination in columns 3-6 in case the first and the second columns are 1
Tuples with 1 in both first and second columns:

(1, 1, 0, 0, 0, 0): count = 209, support = 0.0418, confidence = 0.1465, lift = 1.0129
(1, 1, 0, 0, 1, 0): count = 168, support = 0.0336, confidence = 0.1177, lift = 1.0921
(1, 1, 0, 1, 0, 0): count = 164, support = 0.0328, confidence = 0.1149, lift = 1.0904
(1, 1, 1, 0, 0, 0): count = 137, support = 0.0274, confidence = 0.0960, lift = 0.9959
(1, 1, 0, 1, 1, 0): count = 122, support = 0.0244, confidence = 0.0855, lift = 1.0989
boaz@LAPTOP-VOQ33HA0:/mnt/c/courseAI/lesson10$ python3 aliens.py
```

הערה:

אני חייב לציין, שבהתחלה ניסיתי לבדוק את כל האפשרויות השונות, ונוצרה תוכנית הכוללת כינון של 4 לולאות, כל אחת כנראה בת 5000 מופעים (סה"כ כתריליארד מעברים) וכשהרצתי אותה במהלך השעור במשך כשעתיים היא לא גמרה לרוץ. לכן צמצמתי את כמות הבדיקות רק ל 10 הקומבינציות השכיחות (ניתן גם ללבחור 100 במקום 10) ושילבתי גם הנחות שנבעו מהסתכלות ב 10 הקומבינציות השכיחות שהתקבלו (ששתי העמודות הראשונות מככבות בהן). זמן הריצה היה פחות משנייה.