דניאל לייבנר - 208271775 קורס - כריית מידע – 20595

עמוד **1** מתוך **28** סמסטר 2022ב

ממ"ן 22 - כריית מידע

[שאלה 1: 3](#_Toc101739885)

[סעיף א' 3](#_Toc101739886)

[סעיף ב' 4](#_Toc101739887)

[סעיף ג' 7](#_Toc101739888)

[1. הגדרת המטרות של כריית המידע – 7](#_Toc101739889)

[2. איסוף ושמירת הנתונים- 7](#_Toc101739890)

[3. ניקוי הנתונים – 7](#_Toc101739891)

[4. ביצוע טרנספורמציות על הנתונים – 7](#_Toc101739892)

[5. בחירת שיטות לכריית מידע (כגון רגרסיה, עצי החלטה וכו) 7](#_Toc101739893)

[6. ביצוע דיסקרטיזציות וסיווג הנתונים 7](#_Toc101739894)

[7. הרצת שיטות לכריית מידע שנבחרו 7](#_Toc101739895)

[8. ניתוח התוצאות: 7](#_Toc101739896)

[9. הסקת מסקנות: 7](#_Toc101739897)

[סעיף ד' 8](#_Toc101739898)

[סעיף ה' 9](#_Toc101739899)

[ניקוי הנתונים – 9](#_Toc101739900)

[בדיקה האם קיימות רשומות בעלות ערכים חסרים: 9](#_Toc101739901)

[בדיקה האם קיימות רשומות בעלות ערכים חריגים: 9](#_Toc101739902)

[בדיקה של תכונות מיותרות: 13](#_Toc101739903)

[בדיקה האם קיימות רשומות בעלות ערכים לא הגיוניים: 17](#_Toc101739904)

[ביצוע טרנספורמציות על הנתונים – 18](#_Toc101739905)

[שינוי ערכי הרשומות לצורה אחרת 18](#_Toc101739906)

[שינוי דרך השמירה של המידע. 20](#_Toc101739907)

[יצירת מאפיינים חדשים בהתאם למטרת הכרייה: 21](#_Toc101739908)

[שאלה 2: 22](#_Toc101739909)

[סעיף א' 22](#_Toc101739910)

[סעיף ב' 23](#_Toc101739911)

[סעיף ג' + ד' : 24](#_Toc101739912)

[סעיף ה' 28](#_Toc101739913)

# שאלה 1:

## סעיף א'

בחרו אלגוריתם של חוקי הקשר. תארו ותחו את האלגוריתם שבחרתם. מקו את בחירתכם.