

מבחן במסדי נתונים

עמוס עזריה ומירב שקרון

7029210-1,3,4,5

סמסטר ב' מועד א' ג' באב התשפ"א, 12.7.2021

הנחיות כלליות:

- משך הבחינה: 150 דקות.
- יש לענות בגוף השאלון! המחברת תשמש כטיטא בלבד, מענה במחברת עלול לגרור ציון 0.
- אין להכניס שום חומר עזר.
- השימוש במחשבון אסור.
- בשאלות האמריקאיות רק תשובה אחת נכונה. שימו לב שיש לנמק את הבחירה
- בסיום הבחינה - נא למסור את השאלון ואת המחברת.

	1	2	3	4	5	6	Total
Max points	28	19	10	15	15	18	105
Grade							

ב ה צ ל ח ה !

שאלה 1 - SQL (28 נק')

נתונות הטבלאות הבאות המתארות נתונים מקופת חולים מסוימת הקשורים למחלת הקורונה:

patients(id, first_name, last_name, birth_date, gender, first_dose_date, second_dose_date)

[טבלת מטופלים: תז, שם פרטי, שם משפחה, תאריך לידה, מגדר, תאריך מנת חיסון ראשונה, תאריך מנת חיסון שניה]

isolated(patient_id, start_date, end_date, isolation_type_id)

[טבלת מבודדים: תז, תאריך כניסה לבידוד, תאריך סיום הבידוד, מזהה סוג בידוד]

isolation_types(id, desc)

[טבלת סוג הבידוד: מזהה סוג בידוד, תיאור סוג בידוד]

הערכים בטבלה זו הם:

id	desc
1	חדר ושירותים נפרדים. בבית גרים אנשים נוספים
2	חדר נפרד בבית. שירותים עם שאר יושבי הבית
3	לבד בבית
4	מלונית

confirmed_positive_cases(patient_id, start_date, is_symptomatic, end_date)

[טבלת חולים מאומתים: תז, תאריך תחילת מחלה, האם יש סימפטומים, תאריך סיום המחלה]

- חלו לאחר שהתחסנו במנה ראשונה (ולא השנייה)
- כבר החלימו מהמחלה
- הם בקבוצת הגיל 20-33

עבור כל מטופל/ת יש להציג את הנתונים הבאים:

- שם פרטי
- שם משפחה
- תז
- מס' הבדיקות שעשה/תה אי פעם

(14 נק')

ב. כיתבו שאילתא שתציג את מס' האנשים שחלו בתקופת הבידוד שלהם לפי סוג הבידוד. כלומר, לכל סוג בידוד ייספרו רק מטופלים שאומתו בזמן שהותם בבידוד.

בכל רשומה יש להציג:

- מזהה סוג הבידוד
- תיאור סוג הבידוד
- מס' המטופלים שחלו בזמן שהותם בבידוד זה

(14 בק')

This image shows a single page of white paper with horizontal blue or grey ruling lines. The lines are evenly spaced and run across the width of the page, leaving small margins at the top and bottom. There is no handwriting or other markings on the paper.

$$R(A, B, C)$$

נתונות השורות (הרשומות) הבאות ברלציה:

 a_2, b_2, c_1

כמו כן, נתון שמהחוקיות של הרלציה, ומהשורות הנתונות נובעות גם השורות הבאות (ייתכן שנובעות גם שורות נוספות):

$$a_1, b_2, c_2$$

א. מה ניתן לומר על ה-NF של הרלציה? נמקו היטב. (10 נק')

[illegible]

ב. עבור כל אחת מהשורות להלן, ציינו האם **ייתכן** שהיא ברלציה (כיתבו "ייתכן"), לא ייתכן שהיא ברלציה (כיתבו "לא ייתכן"), או שהיא מוכרחת להיות ברלציה כיוון שהיא נובעת מכל הנתונים (כיתבו "נובעת"):

(9 נק')

$$a_2, b_1, c_1$$

3. XPath – XML (10 נק')

נתון מסמך xml המתאר נתונים על חולי קורונה ומבודדים.

להלן דוגמא למבנה ותכולת הxml-

```
<corona>
  <isolations>
    <isolated>
      <id>335598740</id>
      <full_name>"Tom Segev"</full_name>
      <age>36</age>
      <gender>male</gender>
      <start_date>01/10/2021</start_date>
      <end_date>14/01/2021</end_date>
      <place>"חדר ושירותים נפרדים. בבית גרים אנשים נוספים"</place>
    </isolated>
    .....
  </isolations>
  <confirmed_positive_cases>
    <case>
      <id>775521547</id>
      <name>"Jacky Levi"</name>
      <age>56</age>
      <gender>male</gender>
      <start_date>10/10/2021</start_date>
      <is_symptomatics>"True"</is_symptomatics>
    </case>
    .....
  </confirmed_positive_cases>
</corona>
```

כתבו שאילתא **אחת** ב-XPath שתחזיר את שמות החולים והמבודדים שגילם גדול מ-40 (מטופלים שמופיעים גם ברשימת החולים וגם ברשימת המבודדים יופיעו פעמיים בתוצאה)

שם ה-collection : Vaccinated.

להלן דוגמא למסמך ב-collection:

```
{
  _id: ObjectId("50a8240b927d5d8b5891743c"),
  id: 0335598750
  city: "Ariel",
  first_name: "Yael",
  last_name: "Cohen",
  gender: "female",
  age: "39"
  first_dose_date: new Date("May 04, 2021"),
  second_dose_date: new Date("May 25, 2021"),
  is_recovered: "True"
}
```

כתבו שאילתא ב-mongoDB שתחזיר לכל עיר את מס' המחסונים (חיסון ראשון) שהם גם מחלימים וגם מעל גיל 30.

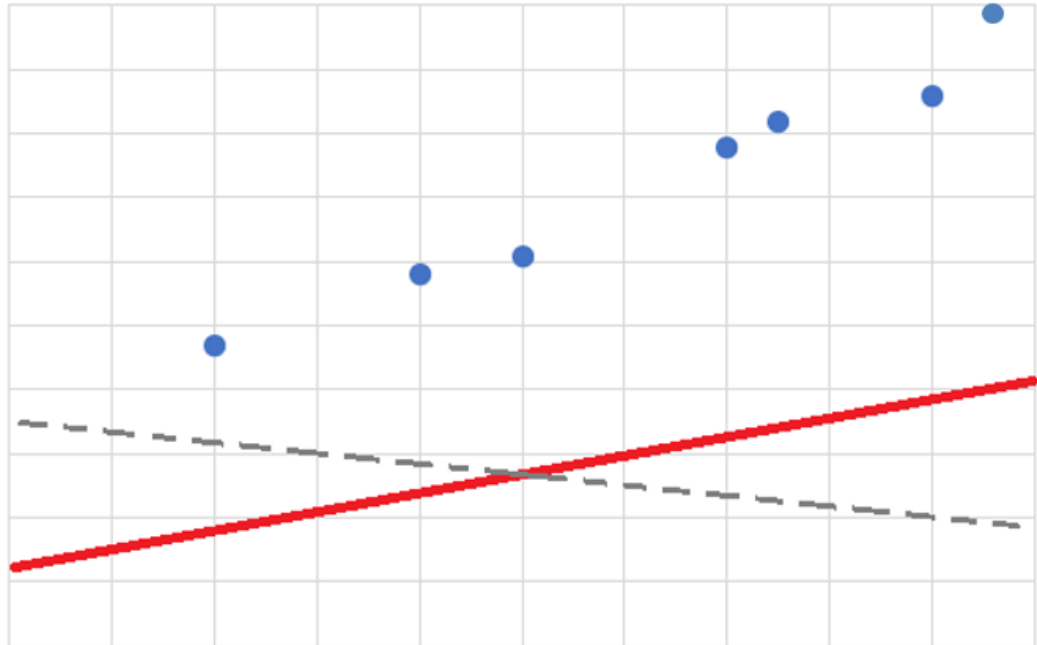
This image shows a blank sheet of white paper with horizontal ruling lines. The lines are evenly spaced and extend across the width of the page. There are no margins, text, or other markings on the paper.

כיתבו באמצעות שימוש ב-Java Streams בלבד פקודה המדפיסה את כל המספרים הראשוניים מ-2 ועד X נתון. **אין** לבנות פונקציית עזר, **אין** להשתמש בפונקציה קיימת לבדיקת ראשוניות (כגון: isProbablePrime) אין צורך להתחשב ביעילות. (רמז: העזרו ב-IntStream.rangeClosed(a,b) המחזירה stream המכיל את כל המספרים מ-a ועד b).

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

6. Linear regression (18 נק')

נתון הגרף הבא המתאר דאטה (באמצעות הנקודות) עם פיצ'ר בודד (בציר ה-x) וה- target label (בציר ה-y), וכן שני מודלים לינאריים: הקו הרציף והקו המרוסק. הניחו שהקצה השמאלי למטה הוא ראשית הצירים (0,0).



ענו על השאלות הבאות. נמקו את תשובותיכם!

א. איזה מהמשפטים הבאים נכון בנוגע ל-Loss (פונקציית הטעות) של שני המודלים: (6 נק')

- ה-Loss של המודל של הקו המרוסק נמוך יותר מזו של הקו הרציף.
- ה-Loss של המודל של הקו הרציף נמוך יותר מזו של הקו המרוסק.
- לא ניתן לדעת איזה Loss נמוך יותר כיוון שאין מספרים על הצירים.
- ה-Loss של שני המודלים זהה, שכן כל הנקודות נמצאות מעל הקווים.

ב. נסתכל על המודל של הקו המרוסק: מה ניתן לומר על w של המודל של הקו המרוסק, ומה יקרה לה באיטרציה הבאה של Gradient descent? (6 נק')

- a. w הנוכחית שלילית והיא תגדל (כלומר תהיה פחות שלילית או תהפוך חיובית).
- b. w הנוכחית שלילית והיא תקטן (כלומר תהיה יותר שלילית).
- c. w הנוכחית חיובית והיא תגדל.
- d. w הנוכחית חיובית והיא תקטן (כלומר תהיה קטנה יותר או תהפוך שלילית).
- e. לא ניתן לדעת האם w תקטן או תגדל לפי הנתונים בשאלה, הדבר תלוי במה שקרה באיטרציה הקודמת.

ג. נסתכל על המודל של הקו הרציף: מה ניתן לומר על הנגזרת של Loss בנקודה b , w התואמים את אלו של המודל של הקו הרציף. זכרו שב Gradient descent אנו תמיד מתקנים בצעד בכיוון הפוך לנגזרת. (6 נק')

- a. לא ניתן לומר דבר כיוון שאין מספרים על הצירים.
- b. הנגזרת החלקית לפי w בנקודה w, b תהיה חיובית, ואילו הנגזרת החלקית לפי b בנקודה w, b תהיה שלילית.
- c. הנגזרת החלקית לפי w בנקודה w, b תהיה שלילית, ואילו הנגזרת החלקית לפי b בנקודה w, b תהיה חיובית.
- d. הנגזרת החלקית לפי w בנקודה w, b תהיה חיובית, וכן הנגזרת החלקית לפי b בנקודה w, b תהיה חיובית.
- e. הנגזרת החלקית לפי w בנקודה w, b תהיה שלילית, וכן הנגזרת החלקית לפי b בנקודה w, b תהיה שלילית.
