מבחן בקורס: "מבוא למדעי הנתונים 1"

סמסטר א' תשפ"ו - מבחן בית
מרצה: ד"ר סלבה נובגורודוב
מתרגל: אחמד ריאן

תאריך הגשה: 21.02.2026
יש לתת תשובות קצרות ומנומקות היטב.

המבחן מורכב מ 10 שאלות רב-ברירתיות ("אמריקאיות") ו 4 שאלות פתוחות.

בהצלחה!

שאלה 1 (3 נק'):

נתון קובץ csv בשם cities.csv הכולל 205 שורות ו5 עמודות (ללא כותרות).
מה ידפיס קוד פייתון הבא:

```
import pandas as pd

df = pd.read_csv("cities.csv")
print(df.shape[1])
```

א. 204
ב. 205
ג. 4
ד. 5

שאלה 2 (3 נק'):

נתונה טבלה בשם T הכוללת מידע היסטורי על תחרות הארוויזיון ב30 שנה האחרונות. להלן חלק קטן מהטבלה הזאת:

Country	Artist	Language	Place	Year
Sweden	Loreen	English	1	2023
Finland	Käärijä	Finnish	2	2023
Israel	Noa Kirel	English	3	2023

איזו מהפקודות הבאות תוציא מהטבלה הנ״ל נתונים על כל הזמרים ששרים באנגלית וזכו במקומות 1 עד 8.

א. **SELECT** * **FROM** T **WHERE** Language = 'English' **AND** Place < 9

ב. **SELECT** * **FROM** T **WHERE** Language = 'English' **AND** Place > 9

ג. **SELECT** * **FROM** T **WHERE** Language = 'English' **AND** Place = 9

ד. **SELECT** Year **FROM** T **WHERE** Language = 'English' **AND** Place < 9

שאלה 3 (3 נק'):

מה ידפיס קוד פייתון הבא:

```
text = "San Francisco, officially the City and County of San Francisco, is a commercial, financial, and cultural center within Northern California. With a population of 808,988 residents as of 2023, San Francisco is the 17th-most populous in the U.S. Large numbers of entrepreneurs and computer application developers moved into the city, followed by marketing, design, and sales professionals"

words = text.lower().replace(“,”,””).split()
words_histo = {}
for w in words:
	if w not in words_histo:
		words_histo[w] = 0
	words_histo[w] += 1
print(words_histo["san francisco"])
```

א. שגיאה
ב. 1
ג. 2
ד. 3

שאלה 4 (3 נק'):

מה ידפיס קוד פייתון הבא:

```
mul = 1
for a in [1, 2, 4, 8] :
	mul *= a
print(mul)
```

א. 2
ב. 4
ג. 64
ד. 128

<u>שאלה 5 (3 נק'):</u>

להלן קטע של דף HTML, מה הוא מייצג?

```
<table>
<tr>
	<td>TEXT</td>
	<td><img src=’photo2.jpg’ /></td>
</tr>
<tr>
	<td>TEXT</td>
	<td><img src=’photo2.jpg’ /></td>
</tr>
</table>
```

א. טבלה עם 2 שורות ו2 עמודות ו3 תמונות שכולן שונות
ב. טבלה עם 2 שורות ו2 עמודות ו3 תמונות שחלקן זהות
ג. טבלה עם 2 שורות ו2 עמודות ו2 תמונות זהות
ד. טבלה עם 2 שורות ו2 עמודות ו3 תמונות שכולן זהות

<u>שאלה 6 (3 נק'):</u>

מה מוצא ה xPath הבא:

```
//a[@href = 'https://www.youtube.com/']
```

א. את הלינקים שמובילים לדף הראשי של YouTube
ב. את הלינקים שמובילים לכל הדפים של YouTube
ג. את התמונות שנמצאות ב YouTube
ד. הקוד לא מוצא כלום ויחזיר שגיאה

שאלה 7 (3 נק'):

נתונה טבלה של תוצאות של משחקי כדורגל הכוללת עמודות game_result (תוצאת המשחק), game_date (תאריך המשחק), team1 (שם הקבוצה שאירחה את המשחק) ו team2 (שם הקבוצה שהתארחה במשחק).

מה עושה קוד SQL הבא:

```
SELECT game_result, game_date, team1, team2
FROM games
WHERE team1= 'Juventus' OR team2 = 'Juventus'
```

א. מחזיר רשימה של כל המשחקים, כולל שמות של שתי הקבוצות, תוצאה ותאריך בהן יובנטוס נצחה

ב. מחזיר רשימה של כל המשחקים, כולל שמות של שתי הקבוצות, תוצאה ותאריך בהן יובנטוס לא שיחקה

ג. מחזיר רשימה של כל המשחקים, כולל שמות של שתי הקבוצות, תוצאה ותאריך בהן יובנטוס אירחה או התארחה

ד. מחזיר שגיאה

שאלה 8 (3 נק'):

מה ידפיס קוד פייתון הבא:

```
names = [“Alice”, “Bob”, “Chen”, “Dennis”, “Eugene”, “Francesca”]

for name in names:
    if len(name) < 3:
        print(name)
```

א. Alice
Dennis

ב. Francesca

ג. Francesca
Eugene

ד. לא ידפיס כלום

שאלה 9 (3 נק'):

נתון קובץ XML הבא:

```
<players>
   <player>
      <name>Alice</name>
      <city>Paris</city>
      <country>France</country>
   </player>
   <player>
      <name>Bob</name>
      <city>Porto</city>
      <country>Portugal</country>
   </player>
</players>
```

אם נרצה לשמור את המידע הזה במבנה זהה בקובץ JSON, איך הקובץ הזה יראה?

א.

```
{"players": { {"name":"Alice", "city":"Paris", "country":"France" },
{"name":"Bob", "city":"Porto", "country":"Portugal"} } }
```

ב.

```
{"players": [ {"name":"Alice", "city":"Paris", "country":"France" },
{"name":"Bob", "city":"Porto", "country":"Portugal"} ] }
```

ג.

```
["players": [ ["name":"Alice", "city":"Paris", "country":"France" },
{"name":"Bob", "city":"Porto", "country":"Portugal"} ] ]
```

ד. לא ניתן לייצג סוג כזה של נתונים בעזרת JSON

שאלה 10 (3 נק'):

מה ידפיס קוד פייתון הבא:

```
print(len("This is simple".replace("simple text","simple").strip().split()))
```

א. 1

ב. 2

ג. 3

ד. 4

שאלה 11 (20 נק'):

סעיף א׳ (10 נק׳)

נתון קובץ CSV בשם transactions_01_2026.csv המתאר פעולות פיננסיות של בן אדם מסויים בחשבונות בנק שונים שלו, כגון הכנסות (הסכום יופיע עם סימן +) או הוצאות (הסכום יופיע עם סימן מינוס) בחודש ינואר של 2026. הקובץ כולל 4 עמודות - תאריך הפעולה, מס חשבון הבנק, סכום ומטבע.

דוגמה לקובץ כזה:

```
2026-01-03, 770-60055, 430, ILS
2026-01-08, 770-98540, -50, EUR
2026-01-09, 770-98540, 109, ILS
2026-01-09, 770-50938, 220, ILS
2026-01-10, 770-98540, -90, EUR
2026-01-11, 770-50938, 100, USD
2026-01-12, 770-50938, 240, EUR
```

כתבו תוכנית פייתון פותחת את הקובץ הנ״ל ומחשבת את היתרה של הבן אדם בסוף החודש (בכל החשבונות ביחד). ניתן להניח שקיימת פונקציה get_USD_rate המקבלת קוד של מטבע (למשל ILS) ומחזירה את שער ההמרה של דולר למטבע הזה (למשל 3.13 במקרה של דולר לשקל).

סעיף ב׳ (5 נק׳)

כתבו קוד שעבור כל מטבע מדפיס את ההוצאה הכי גבוה וההכנסה הכי גבוה שהייתה במטבע הזה.

סעיף ג׳ (5 נק׳)

כתבו קוד שמדפיס את הממוצע הוצאות והכנסות יומי (על סמך כמות הימים שהייתה בהם פעילות כלשהו ולאו דווקא על סמך 31 ימים שיש בחודש ינואר).

שאלה 12 (20 נק'):

נתונה טבלה בשם songs הכוללת מידע על שירים של טיילור סוויפט. להלן חלק קטן מהטבלה הזאת:

Name	Album	Year
I Knew You Were Trouble	Red	2012
Shake It Off	1989	2013
Cruel Summer	Lover	2019
Love Story	Fearless	2008

סעיף א׳ (4 נק׳)

כתבו פקודת SQL המחזירה את כל השירים שיצאו לאור אחרי שנת 2013.

סעיף ב׳ (4 נק׳)

כתבו פקודת SQL את הכמות של שירים בכל אלבום.

סעיף ג׳ (4 נק׳)

כתבו פקודת SQL המחזירה ל3 שנים שבהן יצאו הכי הרבה שירים.

סעיף ד׳ (4 נק׳)

כתבו פקודת SQL המחזירה את שמות כל האלבומים שכוללים ספרה בתוך השם שלהם.

סעיף ה׳ (4 נק׳)

כתבו פקודת SQL המחזירה לכל אלבום את השנה שבה הוא יצא.

שאלה 13 (10 נק'):

סעיף א׳ (5 נק׳)

נתון קובץ CSV בשם grased.csv הכולל את מספר הסטודנט (מתחיל באות S),
מספר הקורס (מתחיל באות C) ואת הציון שהסטודנט קיבל בקורס.
דוגמה לקובץ כזה:

```
S5478201, C301701, 95
S3708443, C301701, 85
S8061216, C301701, 90
S5478201, C402301, 90
S3708443, C402301, 100
S5478201, C510801, 95
```

כתבו תוכנית פייתון פותחת את הקובץ הנ״ל ומחשבת את הממוצע של כל קורס,
ומדפיסה אותו.
למשל, עבור הקובץ לדוגמה הקוד ידפיס:

```
C301701, 90
C402301, 95
C510801, 95
```

סעיף ב׳ (3 נק׳)

הוסיפו לקוד שכתבת בסעיף א׳ הדפסה של הציון הכי נמוך והכי גבוה שהיה בקורס.

סעיף ג׳ (2 נק׳)

כתבו קוד המוצא סטודנט שלקח הכי הרבה קורסים מבין כל הסטודנטים.

שאלה 14 (20 נק'):

נתונה טבלה בשם eurovision הכוללת מידע היסטורי על תחרות הארוויזיון ב30 שנה האחרונות. להלן חלק קטן מהטבלה הזאת:

Country	Artist	Language	Place	Year
Sweden	Loreen	English	1	2023
Finland	Käärijä	Finnish	2	2023
Israel	Noa Kirel	English	3	2023
Italy	Marco Mengoni	Italian	4	2023
Norway	Alessandra	English	5	2023

סעיף א׳ (4 נק׳)

כתבו פקודת SQL המחזירה את כל המדינות שהופיעו עם שיר באנגלית החל מ2015. (כל מדינה צריכה להופיע פעם אחת)

סעיף ב׳ (4 נק׳)

כתבו פקודת SQL המחזירה את כל המדינות שזכו במקומות 10 עד 20 (כולל) בשנת 2018.

סעיף ג׳ (4 נק׳)

כתבו פקודת SQL המחזירה לכל שפה את כמות המדינות שהופיעו עם שיר בשפה הזאת עד שנת 2010.

סעיף ד׳ (4 נק׳)

כתבו פקודת SQL המחזירה את המקום הכי גבוה שזכה שיר בעברית אי פעם.

סעיף ה׳ (4 נק׳)

כתבו פקודת SQL המחזירה את כל הזמרים (והשנים שבהן הם הופיעו) שהופיעו עם שיר באיטלקית ולקחו לפחות מקום עשירי.