

פרויקט קורס חלק א' - דיאגרמות ER ו-SQL

הנחיות להגשת התרגיל:

1. תאריך הגשה – 26.12.21 בשעה 11:55 (בצהריים).
2. מתרגל אחראי על התרגיל- בר גנוסר
3. הגשה לפי קובץ הזוגות המופיע במודל.
4. יש להגיש תיקיית ZIP ששמה הוא מספרי זהות של המגישים בפורמט: ID1_ID2 כאשר ID1, ID2 הם מספרי הזהות של הסטודנטים. התיקיה תכלול את הקבצים הבאים:
 - (i) קובץ pdf ובו פתרון לשאלה 1. קובץ ה-pdf צריך להכיל עמוד שער עם מספרי תעודות הזהות של המגישים ושמותיהם. שם הקובץ צריך להיות מספרי זהות של המגישים בפורמט: ID1_ID2.pdf
 - (ii) קובץ SQL עבור כל אחת מבין השאלות 2 (DDL), 3 (SQL) ו-4 (SQL). שם קבצי ה-SQL צריך להיות ID1_ID2_q<x>.sql כך ש- <x> הוא מספר השאלה. למשל, עבור זוג סטודנטים שמספרי הזהות שלהם הם 12345678 ו-987654321, התשובות לשאלה הראשונה יוגשו בקובץ SQL בשם 123456789_987654321_q1.sql. הנחיות מפורטות לגבי אופן יצירת הקבצים הללו מופיעות בסרטון הוידאו " הנחיות- תרגיל בית1" שהועלה למודל.
 - (iii) קובץ SQL בשם ID1_ID2_views.sql כאשר ID1 ו-ID2 הם מספרי הזהות של המגישים. הקובץ אמור להכיל את כל פקודות יצירת ה-views בהן השתמשתם עבור שאלות 3 ו-4.
 - (iv) קובץ SQL בשם ID1_ID2_delete_views.sql כאשר ID1 ו-ID2 הם מספרי הזהות של המגישים. הקובץ אמור להכיל את כל פקודות מחיקת ה-views שיצרתם עבור שאלות 3 ו-4. באחריותכם לוודא כי פקודות המחיקה אכן עובדות.
5. את הקובץ יש להגיש דרך אתר הקורס במקום המתאים ב-moodle על ידי **אחד** מבני הזוג.
6. איחור בהגשת התרגיל יגרור קנס בגובה 20% מהציון עבור כל יום איחור (פרט למקרים חריגים כגון מילואים, רמקריח אלה יש ליווה טורח הגשה החרגיל)

במסגרת הפרויקט בקורס תשלימו שני חלקים. מטרת הפרויקט היא לבנות מערכת מידע פיננסית השומרת מידע אודות משקיעים, חברות הנשכרות בבורסה וערכי המניות שלהן. בחלק א' (כלומר חלק זה) של הפרויקט תידרשו:

- א. ליצור דיאגרמת ER על סמך הדרישות המפורטות.
- ב. לכתוב את פקודות ה-CREATE TABLE הנחוצות ליצירת מסד הנתונים על סמך דיאגרמת ה-ER שבניתם.
- ג. ליצור views (יוסבר בהמשך) ולהשתמש בהם בתוך שאלות אותן תתבקשו ליצור.

תאור המסד:

בית ההשקעות "יותר תשואה" מעוניין למדל מחדש את מסד הנתונים המיושן בו נשמר המידע הנחוץ לצרכי החברה. "יותר תשואה" מספקת ללקוחותיה פלטפורמה **למסחר בבורסה** ולצורך כך עליה לשמור מידע אודות לקוחותיה (המשקיעים בבורסה), עובדיה, החברות הנשכרות וערכי המניות שלהן.

כל משקיע (investor) מזוהה באמצעות מספר תעודת הזהות שלו, ובנוסף נשמרים שמו, תאריך לידתו, כתובת המייל שלו ותאריך הרשמתו למערכת. "יותר תשואה" לא מאפשרת הרשמה למשקיעים שנולדו בשנת 2004 ואילך. כמו כן, לא ייתכנו שני משקיעים שונים המשתמשים באותה כתובת מייל. משקיע חדש המתחיל להשתמש בשירותי החברה מוגדר בשלושת החודשים הראשונים שלאחר הרשמתו כמשקיע מתחיל (beginner). שנה לאחר תאריך ההרשמה של משקיע הוא הופך להיות משקיע פרימיום (premium). עבור כל משקיע פרימיום יש לשמור את יעדיו הפיננסיים (שדה מילולי). במסד מתועד מידע אודות עובדי החברה. כל אחד מהעובדים הוא בהכרח גם משקיע פרימיום המשתמש בעצמו בשירותי החברה. קיימים שני סוגים של עובדים – כלכלן (economist) ועורך דין (lawyer), כאשר ייתכנו עובדים המשתמשים לשתי הקבוצות. הן עבור כלכלן והן עבור עורך דין יש לשמור את הסקטור בו הוא מתמחה. כל משקיע מתחיל מודרך על ידי (guided by) כלכלן אחד בדיוק, ולכל כלכלן יש לפחות משקיע מתחיל אחד אותו הוא מדריך.

במסד נשמר מידע אודות החברות (company) שבמניותיהן ניתן לשכור. חברה מזוהה באמצעות קוד זיהוי (symbol, שדה טקסטואלי) ובנוסף נשמר הסקטור אליו היא משתייכת (נדל"ן, אנרגיה וכד'), שנת הקמתה (founded) והמדינה בה יושב המטה הראשי שלה (location). חברה עשויה להיות ביריבות עם חברה אחרת. במקרים אלו יש לתעד את הסיבה ליריבות. כלכלן עשוי להתעניין ביריבות בין חברות ולדבר על כך בבלוג האישי שלו. במקרה שבו יריבות בין חברות מופיעה בבלוג של כלכלן יש לתעד במסד סיכום קצר של דבריו אודות היריבות. יריבות כלשהי בין חברות יכולה לכל היותר פעם אחת בבלוג (כלומר אם היא הופיעה בבלוג של כלכלן כלשהו, היא לא תופיע אצלו בבלוג פעם נוספת והיא גם לא תופיע בבלוג של כלכלן אחר).

משקיע יכול להעביר כסף מחשבון הבנק הפרטי שלו לפלטפורמת המסחר לצורך רכישת מניות. כל העברה (transaction) מזוהה באמצעות תאריך וזהות המשקיע שביצע אותה, ובנוסף נשמר סכום הכסף שהועבר. לא ניתן להעביר סכום כסף הקטן מ-1,000 דולר. לעיתים העברה כלשהי נראית למערכת חשודה, ובמקרים אלו ההעברה מועברת לבדיקה של עורך דין יחיד. יש לתעד את ההחלטה של עורך הדין בנושא.

מניה של חברה (stock) מזוהה באמצעות תאריך והחברה אליה היא מתייחסת (ניתן להניח לשם פשטות כי ערך המניה של חברה כלשהי לא משתנה במהלך היום). כמו כן, יש לשמור את ערך המניה באותו יום. עבור כל אחת מהחברות המופיעות במסד קיים תיעוד של לפחות מניה אחת ביום כלשהו. פלטפורמת המסחר של "יותר תשואה" מאפשרת למשקיעים לבצע קניית מניות (buying). עבור כל קניית מניות של חברה בידי משקיע יש לתעד את כמות המניות שנרכשו במסגרתה. משקיע לא יכול לבצע ביום כלשהו יותר מקניה אחד של מניות השייכות לאותה החברה (אך הוא כן יכול באותו היום לבצע קניות של מניות השייכות לחברות שונות).

מידול:

(1) (40 נק') צרו דיאגרמת ER של המערכת. ציינו את כל ההנחות עליהן אתם מתבססים. עבור דרישות שלא ניתנות למידול באמצעות הדיאגרמה, יש לציין מילולית מהי המגבלה ולהציע דרך התמודדות עימה (למשל ברמת ה-DDL). ניתן להיעזר לצורך יצירת התרשים באתר <https://www.draw.io/>. יש לבנות את הדיאגרמה באנגלית ולהשתמש במונחים המופיעים בהערות הסוגריים בתיאור המסד.

(2) (30 נק') צרו סקריפט עם פקודות Create Table מתאימות ליצירת מסד הנתונים. השתמשו בטיפוסים מתאימים (https://www.w3schools.com/sql/sql_datatypes.asp) לפי הערכים הנשמרים בכל שדה. בשדות בהם לדעתכם יש להשתמש בשדה מסוג מחרוזת אתם יכולים להניח כי לא ייתכן ערך שאורכו גדול מ-100 תווים. זכרו – יש חשיבות לסדר יצירת הטבלאות! הקפידו עליו. כמו כן, יש לוודא כי שמות הטבלאות והשדות תואמים לאלו שבהם השתמשתם בתרשים.

טיפ: במקביל ליצירת סקריפט המקיים את הטבלאות, צרו סקריפט המוחק אותן (שייכתב בסדר הפוך לסדר בו הן נוצרו במסד), כך במקרה שתהיה לכם טעות ביצירת הטבלאות תוכלו למחוק בצורה מהירה את המסד ולהתחיל מחדש.

שאלות (Views):

בעולם מסדי הנתונים view הינו טבלה וירטואלית הנוצרת כתוצאה מהפעלת שאלתה על טבלאות רגילות במסד הנתונים. מכיוון ש-view דומה לטבלאות הרגילות במסד הנתונים בכך שגם הוא מורכב משורות ועמודות, ניתן לשלוף ממנו מידע ולעדכן אותו בדיוק כמו טבלה רגילה. במסד הנתונים, view מוגדר על ידי שאלת SQL. כשהמידע בטבלאות עליהן ה-view בנוי משתנה, המידע ב-view משתנה אף הוא בהתאם. נדגים את יתרונות ה-view ואופן הגדרתו בעזרת הדוגמא הבאה:

נניח ובמסד הנתונים שלנו קיימת טבלה בשם Order Details המוגדרת כדלקמן:

OrderDetails: (OrderNumber, ProductNum, QuantityOrdered, PriceEach)

כאשר:

– OrderNumber : מספר הזמנה

– ProductNum : מק"ט המוצר שהוזמן

– QuantityOrdered : כמות שהוזמנה מאותו מוצר

– PriceEach : מחיר ליחידה

כעת, נוכל לבנות view מעל טבלה זו בו יוצג לכל מוצר סכום תשלומי כלל ההזמנות עבורו:

```
CREATE VIEW SalesPerProduct
```

```
AS
```

```
SELECT ProductNum, SUM (QuantityOrdered * PriceEach) as TotalPrice
```

```
FROM OrderDetails
```

```
GROUP by ProductNum
```

יצרנו טבלה וירטואלית בשם SalesPerProduct, ובכל פעם שנרצה לדעת מהו סך ההכנסות ממוצר מסוים נוכל להריץ את השאלתה הבאה (עבור מזהה המוצר המתאים, למשל 102 בדוגמה המופיעה כאן):

```
SELECT TotalPrice
```

```
FROM SalesPerProduct
```

```
WHERE ProductNum = 102
```

כדי להגדיר view נשתמש בפקודה CREATE VIEW ואחריה נרשום את שם טבלת ה-view (הטבלה הוירטואלית).

לאחר מכן נקבע איך view זה צריך להיראות באמצעות שאלת SQL רגילה המופיעה לאחר האופרטור AS. * מחיקת view מתבצעת בדומה למחיקת טבלה:

```
DROP VIEW viewname;
```

כאשר viewname הוא שם ה-view שברצוננו למחוק.

מידע נוסף (כולל syntax והסברים מפורטים) ניתן למצוא בכתובת הבאה:

http://www.w3schools.com/sql/sql_view.asp

בשתי השאלות הבאות עליכם להשתמש אך ורק בשתי הרלציות המוגדרות כדלקמן :

Company (Symbol, Sector, Founded, Location)

Stock (tDate, Symbol, Price)

עבור כל אחת מהרלציות, ה-**primary key** מסומן באמצעות קו תחתון.

תיאור הרלציות:

Company – רלציה שכוללת מידע אודות חברות.

- **Symbol** – מזהה החברה.
- **Sector** – הסקטור אליו משתייכת החברה.
- **Founded** – השנה בה הוקמה החברה.
- **Location** – המדינה בה יושב המטה הראשי של החברה.

Stock – רלציה שכוללת מידע אודות מניות של חברות.

- **tDate** – תאריך.
- **Symbol** – מזהה החברה.
- **Price** – ערך המניה של החברה.

ניתן ואף מומלץ להשתמש ב-**views** לצורך פתרון שתי השאלות.

את כל פקודות יצירת ה-**views** (של שתי השאלות) יש להכניס לקובץ המתאים (הנחית הגשה iii.4). את השאילתה האחרונה שמחזירה את התוצאה המבוקשת אין לכתוב בתור **view** אלא בתור שאילתה רגילה ואותה להכניס לקובץ השאלה המתאים (הנחית הגשה ii.4).
לא לשכוח לסיים כל שאילתה ופקודת יצירת **view** ב-**;**.

(3) (10 נק'):

"מדינה מתועשת" מוגדרת כמדינה שבשטחה יושב המטה של יותר מ-5 חברות שנוסדו לפני שנת 1990.
"חברה מובילה" מוגדרת כחברה שערך השיא של מחיר מנייתה המתועד במסד גבוה יותר מערכי השיא של כל שאר החברות האחרות שהמטה שלהן יושב באותה המדינה.
עבור כל "מדינה מתועשת" יש להחזיר את מזהה "החברה המובילה" (או חברות ברבים אם יש כאלו) ואת ערך השיא של מניית החברה (או חברות) המתועד במסד. יש להחזיר את המידע בצורה ממוינת בסדר עולה לפי שם המדינה.

(4) (20 נק'):

"חברה משתפרת" מוגדרת כחברה שבכל יום המתועד במסד ערך המניה שלה היה גבוה מהערך ביום הקודם (אם עבור חברה כלשהי קיים תיעוד עבור יום אחד בלבד אזי התנאי מתקיים באופן הריק).
"חברה זוהרת" מוגדרת כ"חברה משתפרת" אשר לא קיימת אף חברה נוספת המשתייכת לסקטור שלה המוגדרת גם כן כ"חברה משתפרת".
עבור כל "חברה זוהרת" יש להחזיר את מזהה החברה שלה, את הסקטור אליו היא משתייכת ואת תשואתה בתקופה המתועדת במסד (השינוי באחוזים בין ערך המניה שלה ביום האחרון המתועד במסד לערך המתועד ביום הראשון) בדיוק של 3 ספרות אחרי הנקודה. יש להחזיר את המידע בצורה ממוינת בסדר יורד לפי התשואה.
הערה: האופרטור $\text{ROUND}(x, 3)$ מעגל את הערך של x לכדי דיוק של 3 ספרות אחרי הנקודה.

בהצלחה,

צוות הקורס

נספח

בדיקת נכונות אל מול הקלטים והפלטים

בתיקת zip ששמה test_files נתונים שלושה קבצי קלטים שונים עבור הרלציה Stock בשם Stock<i>.csv</i> כך ש-<i>i</i> מסמן את מספר הקלט (מספר שנע מ-1 עד 3). עבור כל קובץ שכזה מצורפים פלטם נכונים עבור שתי השאילתות. כדי לבדוק את נכונות הפתרונות שלכם מול השאילתות, עליכם לעקוב אחר השלבים הבאים :

1. ליצור את הרלציות במסד הנתונים שלכם על ידי הרצת הפקודות שבקובץ create_tables_commands.sql

2. הכנסת כל הרשומות לרלציה Company (הסבר על אופן טעינת הרשומות לרלציה מופיע בעמוד הבא) מהקובץ Company.csv.

3. עבור כל בדיקה 1-3 :

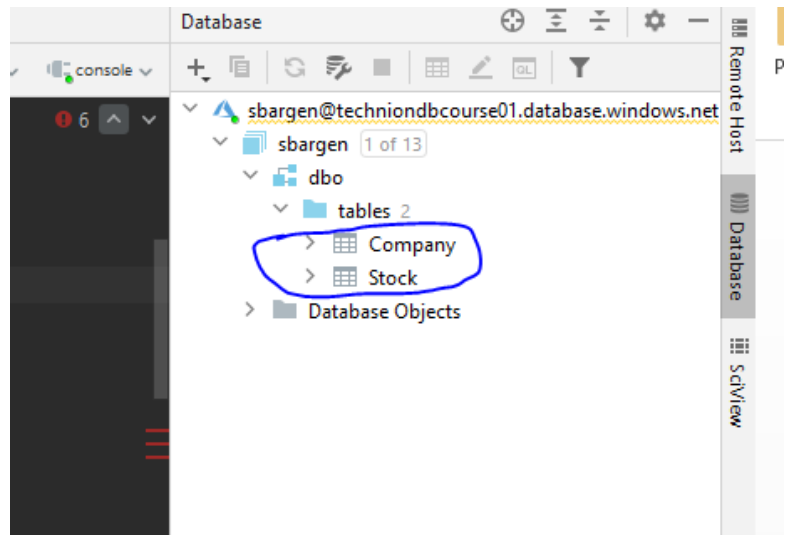
a. הכנסת הרשומות לרלציה Stock מהקובץ Stock<i>.csv</i> כך ש-*i* מסמן את מספר הבדיקה שאתם מריצים כעת.

b. להריץ כל שאילתה ולבדוק את נכונות הפתרון אל מול הפלטם הנתונים בעמודים הבאים.

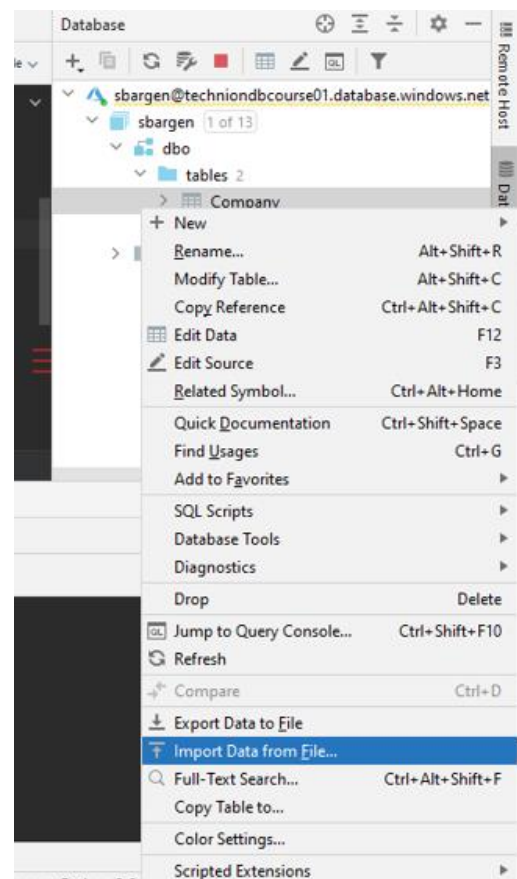
c. למחוק את כל הרשומות מהרלציה Stock באמצעות הרצת הפקודה DELETE FROM Stock.

כיצד טוענים רשומות לרלציה?

1. לאחר יצירת הרלציות (שלב 1) שמותיהן יופיעו בצד ימין, תחת הלשונית Database.



2. לחצו באמצעות הכפתור הימני של העכבר של שם הרלציה שאליה תרצו לטעון נתונים ובחרו באפשרות Import Data from File.



3. בחרו את הקובץ שברצונכם לטעון מתוך מיקומו על מחשבכם.

4. במסך שנפתח סמנו את האפשרות First row is header (צד שמאל למטה) ולחצו Import.

Import "Company.csv" File

Formats: CSV*
TSV
Pipe-separated

Value separator: Comma
Row separator: Newline
Null value text: Empty string

Add row prefix/suffix

Quotation: " " Escape: duplicate
' ' Escape: duplicate

Quote values: When needed

☐ Trim whitespaces
☒ First row is header
☐ First column is header

Table: Company

Comment:

Columns (4) Keys (1) Indexes (1) Foreign Keys

Symbol varchar(10) -- part of primary key mapped to Symbol
Sector varchar(40) mapped to Sector
Location varchar(40) mapped to Location
Founded int mapped to Founded

Data Preview DDL Preview

	Symbol	Sector	Location	Founded
1	MMM	Industrials	Minnesota	1902
2	ABT	Health Care	Illinois	1888
3	ABBV	Health Care	Illinois	2013
4	ABMD	Health Care	Massachusetts	1981
5	ACN	Information Technol...	Ireland	1989
6	ATVI	Communication Servi...	California	2008
7	ADBE	Information Technol...	California	1982

Encoding: UTF-8

☒ Write errors to file: Teaching\HW\22w\project_A\Company_2021-12-10_00_15_11.txt

☐ Insert inconvertible values as null
☐ Disable indexes and triggers, lock table (may be faster)

Import Cancel

בדיקה 1:

שאלתה 3:

	Location	Symbol	Price
1	California	GOOG	3014.18
2	Colorado	DVA	106.47
3	Connecticut	BKNG	2405.01
4	Florida	ROP	498.43
5	Georgia	HD	407.007
6	Illinois	ZBRA	606.96
7	Ireland	ACN	372.78
8	Massachusetts	TMO	641.48
9	Michigan	DPZ	536.83
10	Minnesota	TECH	504.49
11	Missouri	ORLY	652
12	New Jersey	BDX	248.978
13	New York	BLK	948.864
14	North Carolina	MLM	422.942
15	Ohio	MTD	1552.07
16	Pennsylvania	WST	449.36
17	Tennessee	AZO	1911.84
18	Texas	TYL	552.14
19	Virginia	NVR	5277.98
20	Washington	AMZN	3696.06
21	Wisconsin	GNRC	438.68

שאלתה 4:

	Symbol	Sector	Yield
1	ALB	Materials	3.738
2	DLR	Real Estate	2.717

בדיקה 2:

שאלתה 3:

	Location	Symbol	Price
1	California	G00G	2934.35
2	Colorado	DVA	102
3	Connecticut	BKNG	2323.12
4	Florida	ROP	484.87
5	Georgia	HD	410.413
6	Illinois	ZBRA	607.2
7	Ireland	ACN	365.82
8	Massachusetts	TMO	648.77
9	Michigan	DPZ	533.98
10	Minnesota	TECH	479.69
11	Missouri	ORLY	659.18
12	New Jersey	BDX	244.225
13	New York	BLK	926.447
14	North Carolina	MLM	422.343
15	Ohio	MTD	1547.6
16	Pennsylvania	WST	447.38
17	Tennessee	AZO	1868.29
18	Texas	TYL	521.28
19	Virginia	NVR	5549.18
20	Washington	AMZN	3580.41
21	Wisconsin	GNRC	437.1

שאלתה 4:

Symbol	Sector	Yield

בדיקה 3:

שאיילתה 3:

	Location	Symbol	Price
1	California	GOOG	2875.53
2	Colorado	DVA	98.84
3	Connecticut	BKNG	2171.93
4	Florida	ROP	467.27
5	Georgia	HD	407.81
6	Illinois	ZBRA	596.96
7	Ireland	ACN	365.3
8	Massachusetts	TMO	636.11
9	Michigan	DPZ	525
10	Minnesota	TECH	472.03
11	Missouri	ORLY	659.18
12	New Jersey	BDX	243.258
13	New York	BLK	914.631
14	North Carolina	MLM	413.91
15	Ohio	MTD	1547.6
16	Pennsylvania	WST	442.66
17	Tennessee	AZO	1868.29
18	Texas	TYL	518.98
19	Virginia	NVR	5549.18
20	Washington	AMZN	3507.07
21	Wisconsin	GNRC	421.24

שאיילתה 4:

	Symbol	Sector	Yield
1	PLD	Real Estate	2.594
2	HD	Consumer Discretionary	2.218