פרויקט קורס חלק א' - דיאגרמות ER פרויקט קורס

הנחיות להגשת התרגיל:

- .1. תאריך הגשה 11.08.24 בשעה 23:55.
- 2. מתרגלת אחראי על המטלה: קורל שרף
- .90. יש להגיש את תרגיל הבית בזוגות בלבד. ציון המטלה של סטודנט שיגיש לבד ללא אישור יתחיל מ-90.
- 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:
- קובץ pdf ובו פתרון לשאלה 1. קובץ ה-pdf צריך להכיל עמוד שער עם מספרי תעודות הזהות של pdf ובו פתרון לשאלה 1. קובץ הקובץ צריך להיות מספרי זהות של המגישים בפורמט המגישים ושמותיהם. שם הקובץ צריך להיות מספרי זהות של המגישים בפורמט ID1_ID2_ERD.pdf אין להגיש את הפתרון לחלק זה בכתב יד, אלא באמצעות תוכנת סרטוט (המלצה בהמשך).
 - .ID1_ID2_Q2 עבור שאלה 2 (בנושא DDL). שם הקובץ צריך להיות בפורמט SQL קובץ SQL (ii
- של views- קובץ פייתון עבור ה-ID1_ID2_Queries קובץ פייתון עבור ה-3-4 של (iii) שאילתות 3-4 בפורמט ID1_ID2_Views).
 - 5. את תיקיית ה-ZIP 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. במקרים אלה יש לפנות למתרגל האחראי טרם הגשת התרגיל).
- 7. פתרונות לשאילתות שלא יעברו את הבדיקה האוטומטית יקבלו את הציון 0. ניתן להגיש ערעור, אך בהנחה שיתברר כי הסיבה לכישלון בבדיקה היא אי מעקב אחרי הוראות ההגשה, ציון הסטודנטים יתחיל מ-90 והתרגיל ייבדק מחדש (ראו פירוט בסוף מסמך זה).

במסגרת הפרויקט בקורס תשלימו שני חלקים. מטרת הפרויקט היא לבנות מערכת מידע אודות חברת שירותי תוכן. בחלק א׳ (חלק זה) של הפרויקט תידרשו:

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

תיאור המסד:

חברת שירותי תוכן חדשה הוקמה, ונקראתם לסייע לה למדל את מסד הנתונים שלה בהתאם לדרישות המצורפות.

יש לשמור במסד מידע אודות כל התכנים (Content) המתועדים באתר. כל תוכן מזוהה באמצעות שמו, ובנוסף יש לשמור את השפה המדוברת בו ואת תאריך העלאתו. האתר מציע מגוון רחב של תכנים, ביניהם סרטים (Movie) ושירים (Song). עבור כל סרט נרצה לשמור את הז׳אנר שלו ועבור שיר את האורך שלו (בשניות). כמו כן, נרצה לשמור במסד מידע אודות אנשי תוכן. כל איש תוכן מזוהה באמצעות שמו והוא יכול להיות שחקן (Actor) או אמן (Artist). לא קיימים אנשי תוכן שלא שייכים ללפחות אחת מן הקבוצות, ויתכן ששחקן הוא גם אמן. נרצה לתעד לכל סרט את השחקנים ששיחקו בו ואת השכר שהם מקבלים עבור השתתפותם בסרט. יש לוודא כי לכל שחקן שהשתתף בסרט מתועד שכר שגבוה מ-10,000 דולר. בנוסף, לעתים נרצה לתעד מי האמן שביצע שיר כלשהו. לא ייתכן סרט שלא כולל אף שחקן, ולא ייתכן שמספר אמנים מבצעים שיר מסוים. כל אמן מבצע שיר אחד לפחות, ונרצה לתעד מהו השיר האהוב על האמן (לא בהכרח שיר שהוא מבצע). שחקנים עשויים לקנא בשחקנים אחרים. במקרים כאלו, נרצה לתעד את הדבר ולציין את הסיבה לקנאה. שימו לב כי אם שחקן ב׳, אין הכרח שגם שחקן ב׳ מקנא בשחקן ב׳, אין הכרח שגם שחקן ב׳ מקנא בשחקן ב׳, אין הכרח שגם שחקן ב׳ מקנא בשחקן ב׳.

בנוסף, במסד נשמר מידע אודות כל המשתמשים במערכת. כל משתמש מזוהה על ידי תעודת זהות, ונשמר עבורו שמו, כתובת מייל והמדינה ממנה הוא מגיע. לכל משתמש יש מייל ייחודי. החברה מעוניינת לתעד את כל התכנים שמשתמש צרך (בין אם בצפייה או האזנה). משתמש יכול לצורך תוכן במספר ימים שונים, אך לא יתכן שצרך את אותו התוכן באותו היום יותר מפעם אחת. בנוסף, משתמש יכול לתת ציון (מספר שלם בין 1 ל-5) לתוכן אותו צרך שמבטא את מידת ההנאה שלו מהתוכן במועד בו צרך אותו.

לחברת התוכן מאגר של פרסומות (Advertisement) שהיא עשויה להציג למשתמש לפני הצגת התוכן אותו בחר. כדי למשוך את תשומת הלב של משתמשים, החברה מקפידה להציג בכל פרסומת אמן אחד בדיוק. לכל פרסומת יש מזהה ייחודי ונשמר תיאור קצר של תוכן הפרסומת. במקרה שהוצגה פרסומת למשתמש בתחילת צריכת התוכן, נרצה לתעד מהי. לא יתכן שתופיע יותר מפרסומת אחת לפני הצגת התוכן.

משתמשים יכולים ליצור רשימת השמעה (Playlist) שכוללת מספר שירים. קיים לפחות שיר אחד בכל רשימת השמעה. לכל רשימת השמעה יש שם, והיא מזוהה על ידי השילוב של שמה ומזהה המשתמש שיצר אותה. משתמשים ביקורתיים במיוחד עשויים לבדוק רשימות השמעה של משתמשים אחרים, ולסמן שירים בהם כלא מתאימים לרשימת ההשמעה. למשל, אם רשימת ההשמעה נקראת "שירים רגועים" ויש שיר קצבי, הוא עשוי להיות מסומן כלא מתאים על ידי משתמש כלשהו. נרצה לתעד כל מקרה בו משתמש סימן שיר ברשימת השמעה כשיר לא מתאים. אם שיר סומן כלא מתאים ברשימת השמעה יותר מ-5 פעמים, נרצה למחוק שיר זה מרשימת ההשמעה של המשתמש.

מידול:

- (40 נקי) צרו דיאגרמת ER של המערכת. ציינו את כל ההנחות עליהן אתם מתבססים. עבור דרישות שלא ניתנות למידול באמצעות הדיאגרמה, יש לציין מילולית מהי הדרישה ולהציע דרך התמודדות עימה (למשל ברמת ה-DDL). ניתן להיעזר לצורך יצירת התרשים באתר /https://www.draw.io. יש לבנות את הדיאגרמה באנגלית כך ששמות הישויות והתכונות יתאמו לאלו שבהם תשתמשו בשאלה 2.
- (2) (20 נקי) צרו סקריפט DDL עם פקודות Create Table מתאימות ליצירת מסד הנתונים. השתמשו בטיפוסים מתאימים לפי הערכים הנשמרים בכל שדה. בשדות בהם לדעתכם יש להשתמש בשדה מסוג מחרוזת אתם יכולים להניח כי לא ייתכן ערך שאורכו גדול מ-50 תווים, וכי בכל פעם שיש בסיפור התייחסות לתאריך הכוונה היא לפורמט DATE (YYYY-MM-DD) DATE). זכרו יש חשיבות לסדר יצירת הטבלאות! הקפידו עליו. כמו כן, יש לוודא כי שמות הטבלאות והשדות תואמים לאלו שבהם השתמשתם בתרשים (שאלה 1).

טיפ: במקביל ליצירת סקריפט המקים 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.

:(Views):

בעולם מסדי הנתונים view הינו טבלה וירטואלית הנוצרת כתוצאה מהפעלת שאילתה 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 שאילתת SQL.

כשהמידע בטבלאות עליהן ה-view בנוי משתנה, המידע ב-view משתנה אף הוא בהתאם. נדגים את יתרונות

: יאופן הגדרתו בעזרת הדוגמא הבאה view-ה

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

OrderDetails: (OrderNumber, ProductNum, QuantityOrdered, PriceEach)

: כאשר

מספר הזמנה : OrderNumber –

מקייט המוצר שהוזמן : ProductNum –

- כמות שהוזמנה מאותו מוצר : QuantityOrdered

: PriceEach –

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 השאילתא הבאה (עבור מזהה המוצר המתאים, למשל 102 בדוגמה המופיעה כאן):

SELECT TotalPrice

FROM SalesPerProduct

WHERE ProductNum = 102

view משם טבלת ה-view ואחריה נרשום את שם טבלת CREATE VIEW נשתמש בפקודה view כדי להגדיר

הוירטואלית).

.AS זה צריך להיראות באמצעות שאילתת SQL איך להיראות באמצעות באמצעות לאחר האופרטור

* מחיקת view מתבצעת בדומה למחיקת טבלה

DROP VIEW viewname;

כאשר viewname הוא שם ה-viewname כאשר

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

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

בשתי השאלות הבאות עליכם להשתמש אך ורק בשלוש הרלציות המוגדרות כדלקמן (שימו לב כי הרלציות הללו לא בהכרח תואמות בדיוק לאלו שמוגדרות בסיפור שניתן עבור שאלות 1 ו-2):

Movies (<u>title</u>, language, releaseDate)

ActorsInMovies (aName, title, salary)

Jealous (aName1, aName2)

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

<u>תיאור הרלציות:</u>

את הסרטים. – רלציה המתעדת את הסרטים.

- שם הסרט. title •
- שפה המדוברת בסרט. language ●
- .תאריך העלאת הסרט releaseDate •

בסרט. – ActorsInMovies – רלציה המתעדת השתתפות של שחקן בסרט.

- ם השחקן. aName
 - .שם הסרט title •
- השכר על השתתפות השחקן בסרט.salary •

שחקן אחד באחר. – רלציה המתעדת רגשי קנאות של שחקן אחד באחר.

- שם השחקן שמפגין קנאות לשחקן .2 aName1
 - שם השחקן ששחקן 1 מקנא בו. aName2 •

: (זהי) (3

ישחקן על זמנייי הוא שחקן שהשתתף בסרט כלשהו בכל שנה, מבין השנים בהם הועלו סרטים.

״סרט משתלם״ הוא סרט שהשכר **שכל** שחקן קיבל בו גדול או שווה מהשכר הממוצע של אותו שחקן שמחושב על פני כל הסרטים בהם השתתף.

מצאו את הסרטים המשתלמים ששפתם היא אנגלית (English) וכוללים **לפחות** שני שחקנים על זמניים. עבור סרטים אלו, החזירו את שם הסרט, מספר השחקנים בסרט והשכר הממוצע שניתן לשחקן בסרט (מעוגל כ-2 ספרות אחרי הנקודה). החזירו את התוצאה ממוינת בסדר עולה לפי השכר הממוצע. במקרה של שוויון בשכר הממוצע בין סרטים שונים, יש להדפיסם בסדר עולה לפי שם הסרט.

: (זקי) **(4**

ייסרט נטול קנאהיי הוא סרט **שאף** שחקן בו לא מקנא בשחקן אחר, אך **לכל** שחקן קיים שחקן אחר שמקנא בו. מצאו את השחקנים שמקיימים את שלוש התנאים הבאים :

- השחקן שיחק בסרט נטול קנאה אחד לפחות.
- השחקן שיחק בלפחות 2 סרטים בעלי שפה שונה.
- **לא קיים** סרט בו השחקן שיחק וקיבל את השכר הכי גבוה מבין השחקנים שמשחקים בסרט. עבור שחקנים אלו, החזירו את שם הסרט האחרון ששיחקו בו (הסרט שתאריך העלאתו הוא המאוחר ביותר)

ותאריך העלאת הסרט. אם יש כמה סרטים שתאריך העלאתם הוא המאוחר ביותר, יש להחזיר את כולם.

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

את כל פקודות יצירת ה-views (של שתי השאלות) יש להכניס למקום המתאים בקובץ שפורמט שמו הוא view את כל פקודות אין לכתוב בתור ID1_ID2_Views.py. את השאילתה האחרונה שמחזירה את התוצאה המבוקשת אין לכתוב בתור אלא בתור שאילתה רגילה ואותה להכניס לקובץ שפורמט שמו הוא ID1_ID2_Queries.py. ההנחיות מפורטות בהמשך מסמך זה.

לא לשכוח לסיים כל שאילתה ופקודת יצירת view ב-;.

עבור כל אחת משתי השאילתות **מותר להשתמש ב-4 שאילתות VIEWS לכל היותר** (בנוסף לשאילתה המרכזית המחזירה את התשובה הסופית). כמובן שניתן להשתמש גם בפחות.

בהצלחה, צוות הקורס

הערות חשובות

באחריותכם לוודא שמבנה הרלציה (כולל סדר השדות) בשאילתות שתגישו תואם את המבנה הנדרש כפי שמופיע בפלטים לדוגמה (יש משמעות לסדר התכונות בתוצאה).

הבדיקה בחלק זה תורץ באופן אוטומטי ותתבצע עבור מספר קלטים שונים. לתרגיל מצורף קובץ בשם הבדיקה בחלק זה תורץ באופן אוטומטי ותתבצע עבור מספר קלטים הנתונות. בנוסף, מצורפים קבצי create_table_commands.sql קלט ופלט לדוגמה. תיאור הקלטים, אופן העלאת הנתונים לטבלאות והפלטים הנכונים מפורטים בהמשך. מומלץ ליצור את הטבלאות במסד הנתונים שלכם ולבדוק את נכונות הפתרון מול הקלטים והפלטים הנתונים בעצמכם ורק אז לוודא את תקינות ההגשה במנגנון הבדיקה האוטומטית (מוסבר עוד מעט). שימו לב כי בבדיקה האוטומטית שנריץ אצלנו נשתמש בקלטים ופלטים נוספים ולכן אנו ממליצים לכם לחשוב על מקרים נוספים ולבחון את עצמכם עליהם, שכן, ייתכן כי הפתרון שלכם יחזיר תוצאה נכונה עבור קלט מסוים אך לא עבור קלטים אחרים.

לתרגיל מצורף קובץ בשם Automatic_Grader.py אשר מריץ בדיקה אוטומטית על קבצי הקלט והפלט. עליכם לבצע מספר פעולות כדי להריץ את הבדיקה.

1) הזינו את מספרי הזהות של שני השותפים בפורמט ID1_ID2 תחת המשתנה IDS שבשורה 242. למשל, עבור זוג סטודנטים שמספרי הזהות שלהם הם 123456789 ו-987654321, יש לכתוב בשורה 242 (שימו לב כי קו תחתון מפריד בין המספרים):

IDS = "123456789 987654321**'**

- שנו את שם את הקובץ ID1_ID2_Views.py בהתאם למספרי הזהות שלכם. למשל, עבור זוג סטודנטים ID1_ID2_Views.py שמספרי הזהות שלהם הם 987654321 ו-987654321 יש לקרוא לקובץ Views.py שלכם ואותו תגישו.
- סטודנטים למשל, עבור אוג סטודנטים בהתאם למספרי הזהות שלכם. למשל, עבור אוג סטודנטים $ID1_ID2_Queries.py$ שמספרי הזהות שלהם הם 987654321 ו-123456789 יש לקרוא לקובץ בקובץ את השאילתות שלכם ואותו תגישו.
- בשורות NAME_USER ו-NAME בשורות NAME_USER בשורות את שם המסד של אחד מבני הזוג תחת השדות 245 ו-245.

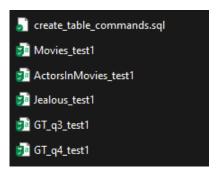
SERVER_NAME = "sglserverdds094241course.database.windows.net"

DATABASE_NAME = "Fname_Lname"

USER_NAME = "Fname_Lname"

PASSWORD = "Qwerty12!"

ל) ודאו כי הקבצים המצורפים מופיעים בתיקיה בשם data, אשר יושב בתיקיה המרכזית בה ממוקם הקובץ (ראו Automatic_Grader.py . ש לעשות לתיקיה ולשים אותה כתיקיה רגילה עם כל הקבצים (ראו תמונה בנקודה 8 שבעמוד הבא) תוכן התיקיה הוא:



(6) בקובץ ששמו (לפני שתשנו אותו בהתאם למספרי הזהות שלכם) ID1_ID2_Views.py מוגדר מילון בשם VIEWS_DICT למילון זה שני מפתחות – Q3 ו-Q4. הערך עבור כל אחד משני מפתחות אלו הוא רשימה VIEWS_DICT שלכם. כל פקודת יצירת VIEW היא איבר ברשימה (סדר הופעת האיברים ברשימה חייב להיות כסדר יצירת ה-VIEWS הרצוי) המופיע בין ייייי ל-ייייי. מזכירים שוב כי ניתן להשתמש ב-VIEWS 4 לכל היותר. אמנם גם אם תחרגו ממגבלה זו הטסט עשוי לעבור, אך על המסך תופיעה הודעה המודיעה על החריגה (ובבדיקה שלנו הטסט ייכשל). מבנה הזנת המידע הוא:

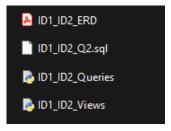
7) בקובץ ששמו (**לפני שתשנו אותו בהתאם למספרי הזהות שלכם**) ID1_ID1_Queries.py מוגדר מילון בשם OUERY_ANSWERS. למילון זה שני מפתחות – Q3 ו-Q4. הערך אותו יש להזין עבור כל אחד ממפתחות אלו הוא השאילתה המתאימה. שימו לב כי עבור שתי השאלות, יש להחזיר את התשובה הסופית דרך שאילתה (זהו התוכן שמוזן בקובץ זה), ולא דרך VIEW. מבנה הזנת המידע הוא:

```
QUERY_ANSWERS = {
    "Q3":
    """
    'Q4":
    """
}
```

8) התיקיה שבה יושב קובץ ההרצאה האוטומטית צריכה להכיל את הקבצים הבאים (עם החלפת השם ID1 ID2 במספרי הזהות שלכם):



כאשר תיקיית ה-zip היא התיקיה אותה אתם מגישים שמכילה את הקבצים הבאים (עם החלפת השם ID1 ID2 במספרי הזהות שלכם):



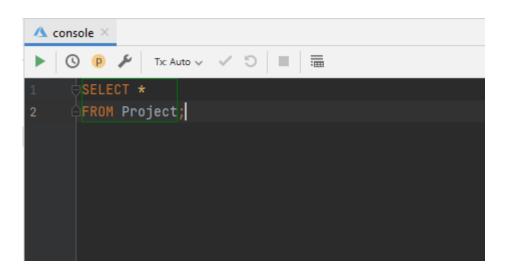
- 9) אנו ממליצים לכם לעבוד קודם כל דרך ה-console, לוודא שאתם בטוחים לגבי תשובתכם ורק לאחר מכן להעתיק את השאילתה למקום המתאים ולהריץ את Automatic_Grader.py.
 - 10) כעת, הריצו את הקובץ. אם אכן פתרונכם נכון עבור הקלט הנתון, הפלט אותו תראו יהיה:

הערה חשובה- פתרונכם בשאלות 3 ו-4 ייבדק בצורה אוטומטית. באחריותכם לוודא כי הקפדתם על הנחיות ההגשה המפורטות בקובץ זה. אם הבדיקה האוטומטית נכשלה ציונכם בתרגיל יהיה 0. כמובן שתוכלו להגיש ערעור והתרגיל ייבדק מחדש אך ציונכם יתחיל מ-90 (אלא אם יתברר שלא חרגתם מאף אחת מההנחיות). אנא ודאו כי ההגשה שלכם מקיימת את כל סעיפי ה-checklist הבא:

- הגשת קובץ PDF שפורמט שמו ID1_ID2_ERD (יש להחליף את ID1 ו-ID2 במספרי הזהות של הגשת קובץ PDF עבור סרטוט התרשים ופירוט ההנחות
- הגשת קובץ ID1 ו-ID1 (יש להחליף את ID1_ID2_q2 שפורמט שמו SQL הגשת הגשת הגשת קובץ את SQL המבוקשות שפורמט שמו אבור פקודות ה-DDL המבוקשות בשאלה ב
 - לצורך הבדיקה הזנת מספרי הזהות של המגישים בשורה 202 בקובץ Automatic_Grader.py. •
- לצורך הבדיקה הזנת שם המסד של אחד מבני הזוג בשורות 245 ו-246 בקובץ Automatic_Grader.py
- הגשת שני קבצי הפייתון שפורמט שמם הוא ID1_ID2_Views.py ו-ID1_ID2_Queries.py (יש להחליף את ID1_ID2_Queries.py במספרי הזהות של הסטודנטים המגישים) כאשר את שני הקבצים הללו יש למלא בהתאם להנחיות הניתנות בנקודות 6 ו-7 שבעמוד הקודם.
- יש לשמור על הסדר הפנימי בין ID1 ל-ID2 בכל המקומות בהם אתם נדרשים לכתוב את מספר הזהות. למשל אין לכתוב בשורה 5 בקובץ 987654321_123456789 ולהגיש את התרגיל כולו תחת השם 123456789 987654321.

נספח 1 – שמירת תוכן שנכתב ב-Console

נניח כי ברצוננו לשמור כקובץ sql את השאילתה הבאה אותה כתבנו ב-console



- 1) לחיצה עם הכפתור הימני של העכבר על כותרת הלשונית (ליד הלוגו הכחול).
 - 2) ריחוף על האפשרות של Open in.
 - 3) בחירה באפשרות של Explorer (במחשבי מאק הפקודה היא
 - 4) כעת תיפתח התיקיה בה נשמר הקובץ. שנו את שמו בהתאם לצרכיכם.

\mathbf{CSV} נספח $\mathbf{-2}$ העלאת רשומות לרלציה מקובץ

- מתוך create_table_commands.sql מתוך בקובץ בידי הרצת הפקודות על ידי הרצת הפקודות במסד הנתונים על ידי הרצת הפקודות בקובץ.
- 2. הזינו את הערכים המופיעים בקבצי ה-CSV של הטסטים באמצעות טעינה ישירה שלהם למסד הנתונים. עליכם לעשות זאת בהתאם לסדר יצירת הטבלאות בקובץ יצירת הטבלאות.

: ניתן לעשות זאת באמצעות באופן הבא

- ובה Import/Export א) לחיצה (עם הלחצן הימני של העכבר) על שם הרלציה, בחירה באפשרות של Import/Export באפשרות. Import Data from file
 - ב) בחירת קובץ ה-csv הרלוונטי מהתיקייה בה הוא יושב על מחשבכם.
 - ג) וידוא המיפוי של השדות המופיעים בקובץ לשדות ברלציה.
 - ד) לחיצה על OK ווידוא קבלת הודעת הצלחה על העלאת כל הרשומות.
 - 3. לאחר שטענתם את כל הרלציות ניתן להתחיל בעבודה.