

1. עבור סטטיסטיקה לפרטי קורסים- קיבלת את השאלות , ראה לינק-
https://github.com/pythonai170624/12.09.2024/blob/master/lectures_courses.sql
 הרץ את השאלות כדי לייצר ולאכלס את בסיס הנתונים.

טבלה 1- טבלת קורסים

course_id	course_name	lecturer_id
1	Mathematics 101	1
2	Physics 101	2
3	Chemistry 101	3
4	Biology 101	4
5	History 101	5
6	Philosophy 101	[NULL]
7	Art History 101	7

טבלה 2- טבלת מרצים

lecturer_id	first_name	last_name	email
1	John	Doe	john.doe@example.com
2	Jane	Smith	jane.smith@example.com
3	Mary	Johnson	mary.johnson@example.com
4	James	Williams	james.williams@example.com
5	Patricia	Brown	patricia.brown@example.com
6	Robert	Jones	robert.jones@example.com
7	Michael	Garcia	michael.garcia@example.com

שים לב:

- בחלק מהקורסים עדיין לא שובץ מרצה
- חלק מהמרצים עדיין לא שובצו לשום קורס

פתור באמצעות שאלות SQL (השתמש בכל סוגי ה- joins שלמדנו):

- הצג את רשימת הקורסים והמרצה המלמד בקורס, בהם יש מרצה המשובץ לקורס
- הצג את רשימת הקורסים בהם אין מרצה המשובץ לקורס. רמז: IS NULL
- הצג את רשימת כל הקורסים והמרצה המשובץ (היכן שאין מרצה משובץ, יופיע NULL בפרטי המרצה)
- הצג את רשימת המרצים והקורס שאותם הם מלמדים, רק עבור מרצים המשובצים
- הצג את רשימת המרצים שאינם משובצים לשום קורס. רמז: IS NULL
- הצג את רשימת כל המרצים והקורס שאותם הם מלמדים (היכן שהמרצה איננו משובץ לאף קורס, יופיע NULL בפרטי הקורס)
- הצג את רשימת כל הקורסים והמרצה המשובץ (היכן שלא משובץ מרצה, יופיע NULL בפרטי המרצה) ביחד עם כל המרצים והקורס שאותם הם מלמדים (היכן שהמרצה איננו משובץ לקורס, יופיע NULL בפרטי הקורס). רמז: FULL OUTER JOIN
- הצג רשימה בה כל מרצה מלמד את כל אחד מהקורסים

המשך בעמוד הבא...

2. הורד מאתר Keggale את רשימת e-commerce מהלינק:

<https://www.kaggle.com/datasets/uom190346a/e-commerce-customer-behavior-dataset>

כתוב תוכנית פייטון ההופכת את הקובץ CSV ל-SQL בשם **ecom.db**

ראה קוד לדוגמא מהשיעור-

https://github.com/pythonai170624/12.09.2024/blob/master/imdb/create_db_sql.py

סקיצה של הטבלה:

Customer ID	Gender	Age	City	Membership Type	Total Spend	Items Purchased	Ave
101	Female	29	New York	Gold	1,120.2	14	
102	Male	34	Los Angeles	Silver	780.5	11	
103	Female	43	Chicago	Bronze	510.75	9	
104	Male	30	San Francisco	Gold	1,480.3	19	
105	Male	27	Miami	Silver	720.4	13	
106	Female	37	Houston	Bronze	440.8	8	
107	Female	31	New York	Gold	1,150.6	15	
108	Male	35	Los Angeles	Silver	800.9	12	
109	Female	41	Chicago	Bronze	495.25	10	
110	Male	28	San Francisco	Gold	1,520.1	21	

הרץ את השאילתות הבאות דרך הפייטון:

השתמש בקובץ `sqlite_lib` מהשיעור הקודם, ובפונקציה `sqlite_lib.run_query_select`

לינק: https://github.com/pythonai170624/12.09.2024/blob/master/sqlite_lib.py

שים לב שהיכן שיש רווח יש לשים גרשיים לדוגמא-

```
select count(DISTINCT e."Membership Type") from ecomm e;
```

- כמה לקוחות יש בטבלה ?
- מהו גיל הלקוח הממוצע ?
- כמה לקוחות גברים יש ? וכמה נשים ?
- כמה מוצרים בממוצע רוכשים הגברים ? כמה הנשים ?
- כמה סוגי membership קיימים בטבלה ?
- כמה חברים יש בכל סוג של-membership ?
- כמה לקוחות גרים ב- new York ?
- כמה לקוחות יש בכל עיר (בסדר יורד מהגדול לקטן) ?
- כמה סה"כ שלמו הגברים (Total Spend) ? כמה הנשים ?
- מי הלקוח שקנה הכי הרבה מוצרים ? מי הלקוח שקנה הכי פחות מוצרים ?

3. ****בנוס/רשות/לא-חובה לאוהבי טסטים:**

כתוב פונקציה `members_count` המקבלת כפרטמטר מחרוזת של `membership type`

לדוגמא: Gold, Silver, Bronze ← ומחזירה כמה חברים יש באותה הקטגוריה

כעת, כתוב טסט לכל קטגוריה (3 טסטים סה"כ)

בתחילת כל טסט חשב את ה- `expected` באמצעות שאילתת SQL

השתמש ב- `pytest.fixture` לטובת חיבור וניתוק מהדאטא בייס לכל טסט

לדוגמא: https://github.com/pythonai170624/12.09.2024/blob/master/test_db.py

בהצלחה!



את שיעורי הבית יש לשלוח ל- pythonai170624+HW17@gmail.com