

מכללת הדסה, החוג למדעי המחשב

אנליזה של ביג דאטה / חננאל פרל

סמסטר ב', תשפ"ד

תרגיל 2 / 10.04.2024

תאריך הגשה:

יום שישי 19.04.2024 בשעה 18:00

מטרות התרגיל:

תרגול שפת SQL ביניים (intermediate)

הוראות (כמו בתרגיל 1 בצהוב התוספות):

הורד את הדאטה בייס:

<https://drive.google.com/file/d/1C4MNtoLAAwGWAjBHOUGL1fTn0zo-rkHL/view?usp=sharing>

לצורך משחק ובדיקות אפשר לפתוח אותו בקלינט הזה (כמובן שיש גם אחרים):

<https://sqlitebrowser.org/>

פתור התרגילים.

ההגשה צריכה להכיל:

- קובץ טקסט אחד (.py) שמכיל תוכנית פייטון שניתן להריצה.
- קובץ טקסט אחד (.txt) של ההדפסות של הרצה שאתם עשיתם אצלכם.

התוכנית פייטון צריכה לפתוח את הדאטהבייס מקובץ לוקלי בסיפרייה נוכחית (ללא תתי תיקיות וכד'), ולהריץ מולו את השאילתות שעונות על השאלות. עבור כל שאלה יש להדפיס את:

(1) מספר התרגיל.

(2) את השאילתא עצמה בעימוד לפי הקונבנציה שלמדנו.

(3) יש להדפיס גם את התוצאה באופן הבא:

a. יש להדפיס את כמות השורות שיצאו סה"כ (לכתוב את מספר השורות, לא כולל כותרות כמובן)

b. ויש להדפיס 5 שורות ראשונות וגם 5 שורות אחרונות של התוצאה (לא להדפיס את כל השורות, רק ראשונות ואחרונות).

אם יש בתוצאה פחות מ 10 שורות, אז להדפיס את כל השורות של התוצאה

פעם אחת.

הדגשה: השאילתא עצמה צריכה להחזיר את כל השורות לפי התרגיל. רק בהדפסה למסך בפייטון צריך להדפיס כאמור את השורות הראשונות והאחרונות.

השאילתא צריכה להיות כתובה בקונבנציה שלמדנו (בכמה שורות, אותיות גדולות איפה שצריך וכו) בפייטון אפשר בנוחות לכתוב מחרוזת פרוסה על פני כמה שורות עם שלוש גרשיים "" (ואפשר באופן כזה במחרוזת לכתוב גם " ' ועוד). ככה:

```
query = ""  
SELECT *  
FROM City  
""
```

יש לשים שורה ריקה וקו מפריד באורך 55 '=' (התו =) בין כל שאלה ושאלה.

הערה: אין לכתוב עברית בשאילתות ובקוד.

שאלה שאתם לא יודעים את התשובה עליה, עדיין לכתוב את הקו המפריד ומספר השאלה, אך להשאיר את השאילתא והתוצאה ריקים.

=====

Question: 1

The query:

```
SELECT *
FROM City
```

Num of rows: 4079

The results:

	ID	Name	CountryCode	District	Population
0	1	Kabul	AFG	Kabul	1780000
1	2	Qandahar	AFG	Qandahar	237500
2	3	Herat	AFG	Herat	186800
3	4	Mazar-e-Sharif	AFG	Balkh	127800
4	5	Amsterdam	NLD	Noord-Holland	731200
	ID	Name	CountryCode	District	Population
4074	4075	Khan Yunis	PSE	Khan Yunis	123175
4075	4076	Hebron	PSE	Hebron	119401
4076	4077	Jabaliya	PSE	North Gaza	113901
4077	4078	Nablus	PSE	Nablus	100231
4078	4079	Rafah	PSE	Rafah	92020

=====

Question: 2

The query:

=====

Question: 3

..

..

השאלות לא בהכרח ממוינות לפי סדר קושי..

בהצלחה!

השאלות SQL 2:

- (1) כמה מדינות ישנן בטבלה
- (2) כמה מדינות ישנן בטבלה שידועה התוחלת חיים שלהם
- (3) כמה יבשות מופיעות בטבלה
- (4) להראות את שנת העצמאות הכי מאוחרת
- (5) מהו ממוצע תוחלת החיים בכל המדינות
- (6) מהו ממוצע תוחלת החיים במדינות גדולות מעל מיליון שטח
- (7) מהו שם המדינה עם השטח הכי גדול (בנוסף: לתת 2 דרכים לפתרון)
- (8) מהו שם המדינה עם השטח הכי קטן
- (9) כל המדינות שהן באותו אזור (Region=) כמו Armenia ממזין בסדר יורד של אוכלוסיה
- (10) להציג 10 מדינות עליונות (ממזין לפי גודל אוכלוסיה) שיש להן אוכלוסיה קטנה מהממוצע
- (11) להציג בשורה אחת - את התמג הכי גבוה ואת התמג הכי נמוך (שאינו 0)
- (12) להציג 2 שורות בטבלה – שורה ראשונה עם התמג הכי גבוה ושורה שניה עם התמג הכי נמוך (שאינו 0)

	category	val
1	max_gnp	8510700.0
2	min_gnp	6.0

- (13) הצג את כל המדינות שאוכלוסייתן גדולה יותר מפי 5 מערך הממוצע באותו אזור (Region=)
- (14) הצג את מספר התושבים בישראל וגם את אחוז התושבים מתוך כלל היבשת (Continent=)
- (15) הצג את כל הערים במדינות ששנת עצמאותם לפני הספירה ממזין לפי מדינה, מחוז, שם עיר
- (16) להראות את כל השפות שמדברים אותן באירופה
- (17) להראות את כל השפות שמדברות גם באירופה וגם באסיה
- (השפה צריכה להיות מדוברות **בשתי** היבשות כדי שתופיע)
- (18) להראות את כל השפות שמדוברות או ב Antarctica או ב Oceania –
(מספיק שתדובר במדינה אחת כדי שתופיע) (בנוסף: לתת 2 דרכים לפתרון)
- (19) להראות את כל השפות **שלא** מדברים באירופה (בנוסף: לתת 2 דרכים לפתרון)
- (20) הראה את העיר בירה של כל המדינות
- (21) כל הערים שיש להם מעל 10% תושבים מתוך המדינה כולה
- (22) להראות את המדינה הרביעית בשטחה
- (23) הצג את שם המדינה ותוחלת החיים ובעמודה ליד תופיע אחת המילים:
HIGH MED LOW לפי הטוחים 0.40.70.100 באופן כזה: 0-low-40-med-70-high-100