

מכללת הדסה, החוג למדעי המחשב

אנליזה של ביג דאטה / חננאל פרל

סמסטר ב', תשפ"ד

תרגיל 3 / 18.04.2024

תאריך הגשה:

יום שישי 3.05.2024 בשעה 18:00

מטרות התרגיל:

תרגול שפת SQL ביניים עוד (intermediate more)

הוראות (כמו בתרגיל 2):

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:

- א. קובץ טקסט אחד (.py) שמכיל תוכנית פייטון שניתן להריצה.
- ב. 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 3:

- 1) להציג רשימת זוגות מדינות שיש להם שפה משותפת (הערה1: אפשר לתת את שם הקוד של המדינה, לא צריך שם מלא) (הערה2: זה בסדר שכל זוג יופיע פעמיים למשל: אנגליה-ישראל, ישראל-אנגליה)
- 2) שמות המדינות עם יותר מ-3 שפות
- 3) כמה מדינות עם יותר מ-3 שפות
- 4) כמה מדינות יש להם שפה רשמית אחת
- 5) כמה מדינות יש להם רק שפות רשמיות
- 6) מה מדינות שמדברים בהם עברית
- 7) מדינות שמדברים בהם ערבית ואנגלית - מדברים את שניהם (ואפשר גם שידברו שפות אחרות) (Arabic,English)
- 8) מדינות שמדברים בהם או ערבית או אנגלית או שניהם (ואפשר גם שידברו שפות אחרות)
- 9) מדינות שמדברים בהם גם ערבית וגם אנגלית ורק אותם (שניהם) ואין שום שפה אחרת
- 10) מדינות שמדברות או ערבית או אנגלית אבל לא את שניהם
- 11) מדינות שאין להם אנגלית כשפה
- 12) מדינות שיש ערבית אבל אין אנגלית
- 13) להראות השפות שמדברים בשני מדינות או יותר
- 14) להראות השפות שמדברים בדיוק בשני מדינות
- 15) כמה ערים יש בכל אות (האות התחילית) "היסטוגרמה"
- 16) מדינות שבהם כל השפות מתחילות באותה אות (ויש יותר משפה אחת)
- 17) עבור כל מדינה להראות שפה אחת כלשהי (איזה שפה אתם מראים? רנדומלי או משהו מסוים?)
- 18) עבור כל מדינה להראות את השפות הרשמיות
- 19) להראות את 3 המדינות עם כמות השפות הכי גבוהה
- 20) עבור כל מדינה להראות את אחוז השטח שלה מתוך כלל שטח היבשת בה היא נמצאת
- 21) למצוא ממוצע כמות שפות לכל המדינות
- 22) להראות בשאלתה אחת שלושה דברים: א. את כמות המדינות ששנת עצמאות אחרי 1950. ב. את ממוצע תוחלת החיים באפריקה. ג. את ההפרש בין כמות התושבים במדינה המקסימלית באירופה לבין כמות התושבים במדינה המינימלית באירופה
- 23) להראות עבור כל הערים עם יותר מחצי מיליון תושבים: את שם העיר, שם המדינה המלא, בשורה אחת את כל השפות שבאותה עיר (ההנחה היא שבכל עיר מדברים את כל השפות של המדינה)
- 24) רשימת זוגות מדינות שיש להם לפחות 2 שפות משותפות (הערה1: הזכרו מה עשיתם בשאלה 1 כאן למעלה) (הערה2: אפשר לתת את שם הקוד של המדינה, לא צריך שם מלא)
- 25) עבור כל מדינה אני רוצה לראות בתוצאה שלוש שורות, בעמודה הראשונה: בכל שורה יופיע שם המדינה, בעמודה השנייה: בשורה הראשונה יופיע התמג מוכפל ב 3 בשנייה התמג מוכפל ב 11 ובשלישית התמג מוכפל ב 20. דוגמא:

Name	GNP_orig	v	GNP_new
------	----------	---	---------

Afghanistan	5976.0	3	17928.0
Afghanistan	5976.0	11	65736.0
Afghanistan	5976.0	20	119520.0
Albania	3205.0	3	9615.0
Albania	3205.0	11	35255.0
Albania	3205.0	20	64100.0

26) עבור כל קוד מדינה ושפה להראות את האחוז המצטבר (סכום של האחוזים) של השפות במדינה ממוין

מהאחוזים הקטנים לגדולים. דוגמא:

CountryCode	Language	percent	percent_agg_sum
ABW	Dutch	5.3	5.3
ABW	Spanish	7.4	12.7
ABW	English	9.5	22.2
ABW	Papiamentu	76.7	98.9
AFG	Balochi	0.9	0.9
AFG	Turkmenian	1.9	2.8
AFG	Uzbek	8.8	11.6
AFG	Dari	32.1	43.7
AFG	Pashto	52.4	96.1

27) עבור כל מדינה להראות את השפה עם האחוז הכי גבוה.