

מכללת הדסה, החוג למדעי המחשב

אנליזה של ביג דאטה / חננאל פרל

מסטר ב', תשפ"ד

תרגיל 5 / 17.05.2024

תאריך הגשה:

יום ראשון 26.05.2024 בשעה 20:00

מטרות התרגיל:

תרגול Mongo DB

הוראות:

נא לפתוח חשבון חנים במונגו בענן Atlas

<https://account.mongodb.com/account/register>

לייצר שם את הטבלאות לדוגמא (Sample Dataset), לפי ההסברים כאן:

<https://www.mongodb.com/docs/charts/tutorial/order-data/prerequisites-setup/>

<https://www.mongodb.com/docs/atlas/sample-data/>

אפשר בממשק לשחק ולהריץ את השאילתות.
ההגשה צריכה להיות בפייטון.
התחברות ע"י הספרייה של מונגו:

<https://www.mongodb.com/docs/drivers/pymongo/>

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.

עבור כל שאלה יש להדפיס את:

- (1) מספר התרגיל.
- (2) את השאילתה עצמה.
- (3) יש להדפיס את ה-10 רשומות הראשונות של התוצאה.

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 השאילתה והתוצאה ריקים.

השאלות לא בהכרח ממוינות לפי סדר קושי..

[הערה: יתכן שבהמשך כדי לאפשר לבדוקת להריץ את הקוד אצלה מול המונגו שלכם, תצטרכו להוסיף את הכתובת IP שלה ולאפשר לה גישה. נפרסם פרטים מדויקים בהמשך.]

בהצלחה!

השאלות MONGO 5:

sample_airbnb (1)

- a. נא להציג את הרשומות מסוג בית (property_type) שיש להם מעל 50 ביקורות ובנוסף מתקיים תנאי אחד מאלו: אלו שיש 5 או יותר מיטות או שיש 2 או פחות מיטות.
- b. צור שדה חדש בשם "מחיר_לאדם" (לפי שדה accommodates). לאחר מכן, מצא את 5 הערים עם "מחיר_לאדם" הממוצע הגבוה ביותר עבור סוג החדר (room_type) "חדר פרטי", בתוצאה עבור כל עיר להראות את הממוצע הזה ואת כמות החדרים מהסוג הזה.
- c. עבור כל הביקורות (reviews) נא להראות את 5 המגיבים הכי הרבה, ברשימה ממוינת מהקטן לגדול.

sample_mflix (2)

- a. להציג את 5 הסרטים המובילים עם המספר הגבוה ביותר של תגובות, עבור כל סרט להציג את הכותרת, השנה וכמות המגיבים.
- b. למצוא את הבמאים שביימו לפחות 5 סרטים עם דירוג ממוצע גבוה מ-7. עבור כל במאי, כלול את שמם, מספר הסרטים המתאימים והדירוג הממוצע של אותם סרטים.

(3) שאלה כללית:

- יש לכתוב שתי (2) שאלות אינטלגנטיות ורלוונטיות על הדאטה.
- לבחור דאטהבייס אחד מה- Sample Dataset שאיננו מופיע בשאלות 1,2 (ז"א שאיננו sample_airbnb ואיננו sample_mflix).
- שתי השאלות צריכות להיות על אותו דאטה בייס (DataBase) שבחרתם.
- יש לכתוב את השאילתות שעונות על השאלות שיצרתם.
- יש להגיש את השאלות כתובות בעברית בקובץ וורד. ליד כל שאלה לרשום את השאילתה שעונה עליה.
- בנוסף בקובץ פייטון שאתם מגישים יש להוסיף את השאילתות שלכם עם ההדפסות כמו שמוסבר למעלה.
- יש להשתמש בשאילתות במגוון פעולות.