

MySQL-Python Connectivity

בתרגיל זה נתרגל את כל השלבים של חיבור בין מסד נתונים ב-MYSQL לבין קוד בפייתון. בתרגיל שלושה חלקים, כל חלק בנוי על קודמו.

ניתן למצוא את המידע לתרגיל בגיטהאב - <https://github.com/shaytavor/MySQL-Python>

חלק 1 – התקנת MySQL

בחלק זה נתקין את מסד הנתונים MySQL.

גשו לכתובת <https://dev.mysql.com/downloads/mysql/> והורידו את גרסת ה-community של MySQL. התקינו את הקובץ. אם אתם מתבקשים לייצר סיסמה, אל תשכחו לשמור אותה במקום בטוח, תצטרכו אותה בהמשך.

לאחר ההתקנה – פתחו את חלון ה-command line והעתיקו את הפקודה הבאה:

`usr/local/mysql/bin/mysql -uroot -p` (עבור מחשבי מאק)

`C:\Program Files\MySQL\MySQL Server 8.0\bin` (עבור windows).

אם ההתקנה עברה בהצלחה, תתבקשו להזין את הסיסמה שבחרתם בעת ההתקנה ולאחר הזנת הסיסמה יפתח prompt של MySQL:

Mysql>

הקלידו את הפקודה:

`show databases;`

הפלט אמור להראות כמה טבלאות ברירת מחדל שקיימות בשרת.

חלק 2 – יצירת מסד הנתונים

בחלק זה ניצור את מסד הנתונים ואת הטבלאות שנעבוד איתן. אנו נעבוד עם נתונים שנלקחו עבור מחקר על הפרעות שינה.

צרו מסד נתונים חדש ע"י הפקודה

`Create database sleep_health;`

והכנסו למסד הנתונים ע"י הפקודה

`Use sleep_health;`

בתיקיית התרגיל תמצאו את הקובץ `sleep_health.csv`. הקובץ מכיל מידע על 374 נבדקים, כאשר לכל נבדק נשמרים מספר נתונים פיזיולוגיים לגבי הרגלי השינה והחיים שלו, לצורך מחקר על הפרעות שינה. פתחו את הקובץ והתרשמו מסוגי הנתונים.

לצורך התרגיל, נשתמש רק בחלק מהנתונים:

- מספר מזהה הנבדק (מסוג integer)
- מין הנבדק (מסוג varchar(10))
- גיל (מסוג integer)
- מקצוע (מסוג varchar(20))
- שעות שינה (מסוג real)
- קטגוריית BMI (מסוג varchar(10))
- הפרעת שינה (מסוג varchar(10))

במסד הנתונים שיצרנו, צרו טבלה חדשה בשם patients שתכיל את העמודות האלה (לפי הטיפוסים המפורטים). מספר מזהה הנבדק יהיה מפתח ראשי.

כתבו פונקציה בפייתון שתקרא את קובץ ה-csv ותייצר ממנו קובץ חדש שבתוכו יהיו פקודות insert שיכניסו לטבלה החדשה את הנתונים.

על הפונקציה לעבור שורה אחת שורה בקובץ, לקרוא את הנתונים הרלבנטיים עבור כל שורה ולשרשר אותם למחרוזת.

כך למשל אמור להיראות קובץ הפלט עבור שתי השורות הראשונות:

```
Insert into patients(id, gender, age, occupation, sleep_duration, bmi_category, sleep_disorder)
```

Values

```
(1, 'Male', 27, 'Software Engineer', 6.1, 'Overweight', 'None'),
```

```
(2, 'Male', 28, 'Doctor', 6.2, 'Normal', 'None'),
```

....

לא לשכוח לשים נקודה פסיק בסוף המחרוזת.

הריצו את הפונקציה והעתיקו את הטקסט שנוצר בקובץ לתוך ה-shell של mysql. אם הפקודה נכתבה נכון, 374 שורות אמורות להכנס לטבלה patients.

בדקו את הטבלה – הריצו את השאילתה –

```
Select * from patients;
```

הנחיות –

- קובץ csv בנוי משורות, כאשר בכל שורה הערכים מופרדים בפסיקים. ניתן להשתמש בפקודה split של פייתון כדי להפריד מחרוזת לרשימה של תתי מחרוזות בין פסיק לפסיק.

- ניתן לפתוח קובץ ע"י הפקודה

```
f = open("filename", "r")
```

הפקודה פותחת את הקובץ filename לקריאה.

- אפשר לעבור על כל שורות הקובץ באמצעות הלולאה –

```
for line in f:
```

לולאה זו תקרא למשתנה line שורה אחרי שורה. כל שורה תהיה מחרוזת (לא לשכוח שכל שורה מסתיימת ב-enter).

- אפשר לפתוח קובץ לכתבה ע"י הפקודה

```
f = open("filename", "w")
```

ולכתוב לתוכו טקסט ע"י הפקודה

```
f.write("abcde")
```

חלק 3 – קישור למסד הנתונים מפייתון

בחלק זה נכתוב קוד בפייתון שיתחבר למסד הנתונים ויתקשר איתו.

כתבו מחלקה בשם Patient שתייצג נבדק. המחלקה תכיל את כל פרטי הנבדק כפי שמופיעים בטבלה

חוץ ממספר המזהה.

כתבו את הפונקציות הבאות:

def insertPatient(conn, p) – הפונקציה תקבל אובייקט connection למסד הנתונים, ואובייקט מסוג

Patient. הפונקציה תוסיף שורה חדשה עם פרטי הנבדק. שימו לב שכיוון שמספר המזהה חייב להיות

יחודי, חשבו על דרך בה תוכלו לוודא שהנבדק החדש שהפונקציה מוסיפה יקבל ערך יחודי במספר המזהה.

def getByOccupation(conn, occupation) – הפונקציה תקבל אובייקט connection למסד הנתונים, ומחרוזת שמייצגת מקצוע. הפונקציה תחזיר רשימה של אובייקטים מסוג patient של כל הנבדקים שהמקצוע שלהם שווה ל-occupation. אם אין אף נבדק כזה, הפונקציה תחזיר רשימה ריקה.

def getMFStatistics(conn) – הפונקציה תקבל אובייקט connection למסד הנתונים. הפונקציה תחזיר מילון (dictionary) שיכיל את אחוז הגברים מסך כל הנבדקים ואחוז הנשים. המילון יראה כך:
{‘Male’: 59.2, ‘Female’: 60.8}

def getDisorders(conn, disorder = “”) – הפונקציה תקבל אובייקט connection למסד הנתונים ופרמטר דיפולטיבי בשם disorder. הפונקציה תחזיר רשימה של כל הנבדקים שחוו הפרעת שינה שמצויינת בפרמטר disorder. אם הפרמטר מכיל מחרוזת ריקה, הפונקציה תחזיר רשימה של כל הנבדקים שחוו הפרעת שינה כלשהי.