

## תרגיל בית מספר 1

### הנחיות להגשת התרגיל:

1. תאריך הגשה - 14/11/22 בשעה 23:55

- הגשה לפי הזוגות המופיעים במודל.
- יש להגיש תיקיית zip הכוללת קבצי sql עבור כל אחת מהשאלות 1-3. שם תיקיית ה-zip צריך להיות מספרי זהות של המגישים בפורמט: ID1\_ID2 כאשר ID1, ID2 הם מספרי הזהות של הסטודנטים. שם הקובץ צריך להיות <א>\_ID1\_ID2\_q<א>.sql כך ש-<א> הוא מספר השאלה. למשל, עבור זוג סטודנטים שמספרי הזהות שלהם הם 123456789 ו-987654321, התשובות לשאלה הראשונה יוגשו בקובץ sql בשם 123456789\_987654321\_q1.sql. ראו נספח בסוף התרגיל לגבי אופן יצירת הקבצים הללו. את פתרוכם לשאלה 4 יש לכתוב בקובץ PDF ששמו ID1\_ID2\_q4.pdf כאשר ID1, ID2 הם מספרי הזהות של הסטודנטים (את קובץ זה יש לצרף **מחוץ** לתיקיית ה-zip).
- את הקובץ יש להגיש דרך אתר הקורס במקום המתאים ב-moodle על ידי **אחד** מבני הזוג.
- שאלות ייענו בפורום מיועד במודל בלבד.

### הקדמה:

מסד הנתונים שילווה אותנו בתרגיל בית זה מתאר ממשק (Real-Time Bidding) RTB. המטרה המרכזית של המטלה היא חזרה על SQL לצד שימוש בכלי זה לצורך אקספלורציה של דאטה חדש. הרלציות המתוארות בהמשך נוצרו מתוך קובץ גולמי וחלק מהמידע המופיע בהן שונה מעט ביחס למקור. הן לא בהכרח מייצגות את הדאטה עימו תתמודדו בהמשך הקורס. פירוט מלא של הדאטה (בצורתו הגולמית) מופיע במסמך שבמודל (בקישור הבא [RTB Specifications](#)). להלן המבנה הסכמטי של המסד:

**Bids** (ID, TimeStamp, DeviceID, ImpID)

**Cities** (City, RegionCode, CountryCode)

**Devices** (DeviceID, City, Make, Model)

**Imps** (ImpID, W, H, Pos, BidFloor)

עבור כל אחת מהרלציות, ה-key primary מסומן באמצעות קו תחתון.

### תיאור הרלציות:

#### **Bids – רלציה שמתעדת bids.**

- ID – מזהה ייחודי של ה-bid.
- TimeStamp – הזמן המדויק בו הוגש ה-bid.
- DeviceID – מזהה המכשיר שאליו כוון ה-bid. פרטים נוספים בפרק 3.2.18 במסמך שבמודל.
- ImpID – מזהה ה-impression של ה-bid. פרטים נוספים בפרק 3.2.4 במסמך שבמודל.

### **Cities – רלציה שמתעדת ערים.**

- City – שם העיר.
- RegionCode – מזהה המחוז.
- CountryCode – מזהה הארץ בה נמצאת העיר.

### **Devices – רלציה המתארת מכשירים אליהם מוגשים bids.** פרטים נוספים בפרק 3.2.18 במסמך שבמודל.

- DeviceID – מזהה המכשיר.
- City – העיר בה אוכן המכשיר.
- Make – היצרן של המכשיר.
- Model – הדגם של המכשיר.

### **Imps – רלציה המתארת impressions.** ניתן להניח כי לכל bid יש impression יחיד וכי כל ה-impressions המתועדים במסד מוצעים כבאנר. פרטים נוספים בפרק 3.2.4 במסמך שבמודל.

- ImpID – מזהה ה-impression.
- W – רוחב (פיקסלים).
- H – גובה (פיקסלים).
- Pos – מיקום הפרסומת על המסך.
- BidFloor – עלות מינימלית ל-bid (ב-CPM).

## מומלץ לעבור על הנספח לפני תחילת העבודה על השאילתות

ניתן להשתמש ב-VIEWS לצורך הפתרון (תזכורת בסוף הנספח) אך אין באמת צורך בכך למעט שאילתה 3 (ואולי) 4.

### שאלה 1 (25 נק')

כיתבו שאילתה המחזירה את מזהי כל הארצות אשר יש בהן **יותר** מ-3 ערים המתועדות במסד ששמן (של הערים) מופיע במילון אחרי המילה 'WOW'

### שאלה 2 (25 נק')

כיתבו שאילתה המחזירה את שמות יצרני המכשירים אשר מקיימים את התנאים הבאים:

- מספר המכשירים המתועדים במסד ששייכים לאותו יצרן אשר אוכנו בארה"ב (USA) **גבוה** מ-100.

- מספר הדגמים **השונים** ששייכים לאותו יצרן **גבוה** מ-20.

יש להחזיר את המידע בצורה ממוינת בסדר יורד לפי מספר המכשירים המתועדים במסד ששייכים לאותו יצרן אשר אוכנו בארה"ב, וכמו כן יש להחזיר גם את מספר זה. שימו לב כי אין להחזיר ערכי NULL תחת השדה של Make.

### שאלה 3 (25 נק')

"עיר לגיטימית" היא עיר שבשטחה אוכנו **יותר** מ-10 מכשירים, ושמוצע ערכי העלות המינימלית ל-bid של impressions שהוגשו עבור מכשירים בעיר זו גבוה מהערך **הממוצע** של העלות המינימלית ל-bid של כל ה-impressions המתועדים במסד.

"מדינה מעודדת bids" היא מדינה שיש בה **לפחות** 2 ערים לגיטימיות שונות.

עליכם להחזיר רשימה ממוינת של מדינות מעודדות bids (בסדר אלפביתי עולה), ועבור כל מדינה להחזיר את ערך העלות המינימלית ל-bid ה**גבוה** ביותר באותה המדינה (כלומר להחזיר את ערך ה-BidFloor המקסימלי באותה המדינה מבין כל ה-bids שהוגשו למכשירים הנמצאים בה).

**הערה:** אתם יכולים להתעלם מכך שיתכן שהמסד עשוי להתייחס לערך הנקוב לפי המדינה הרלוונטית ולהניח שכולם משלמים לפי אותו מטבע.

### שאלה 4 (25 נק') – אקספלורציה חופשית (להגיש כקובץ PDF)

כיתבו שאילתה לבחירתכם אשר מדגימה תופעה מעניינת בנתונים שקיבלתם.

עליכם להגיש בקובץ ה-PDF את השאילתה עצמה, צילום מסך של התוצאה המתקבלת ולהוסיף הסבר בן 5 משפטים לכל היותר המפרט את מה שעושה השאילתה ומסקנה הנובעת מהתוצאה.

זכרו- נושא הקורס השנה הוא Anomaly detection. לכן, נסו למצוא תופעות אשר עלולות להצביע על חריגים בדאטא.

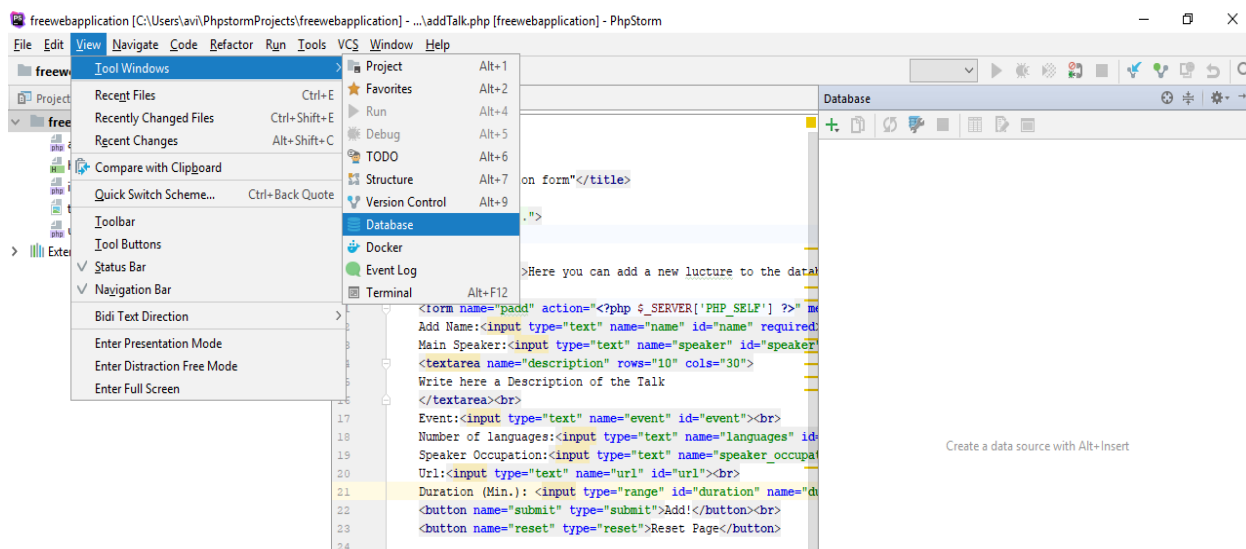
אתם מוזמנים להשתמש בידע אישי שלכם (למשל על גודל מדינות ופופולריות חברות) על מנת להסביר מדוע אתם סבורים שהתופעה מעידה על חריגים בדאטא.

מומלץ להוסיף גרפים/אמצעים ויזואליים אחרים אשר יתרמו להצגת התופעה.

## נספח

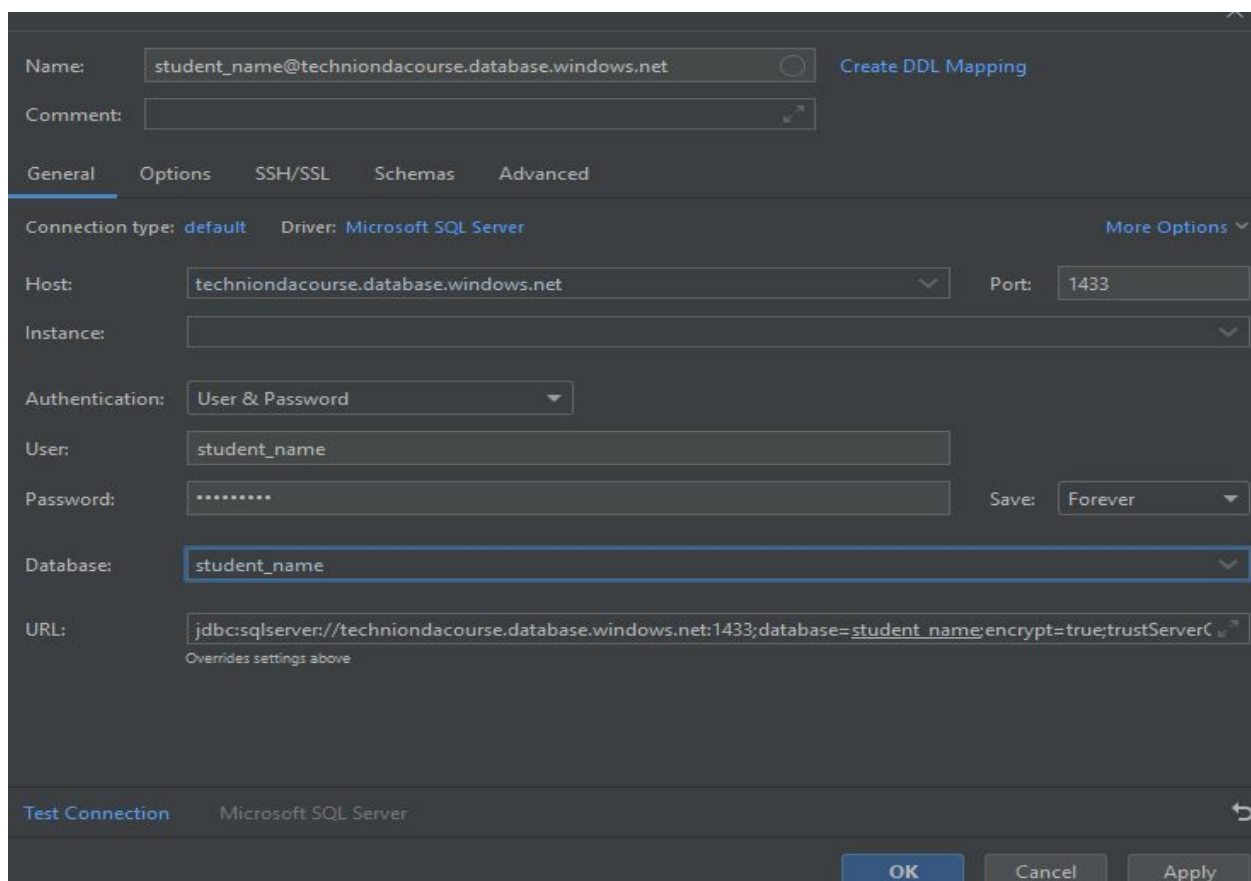
לפני תחילת העבודה מומלץ לבצע את השלבים הבאים כדי להכניס את הנתונים למסד.

### 1. היכנסו ל-PyCharm והתחברו למסד הנתונים שהוקצה עבורכם לפי סדר הפעולות הבא:



כעת בחרו Azure SQL database.

הערה- בחלק מהמקרים החיבור למסד לא מתאפשר דרך הרשת של TechPublic. אם אתם עובדים מהטכניון מומלץ להתחבר דרך TechSec או באמצעות נתב אלחוטי נייד.



במסך שנפתח עליכם למלא את פרטי החיבור שלכם.

שם המשתמש: אחד מבני הזוג ללא @campus.technion.ac.il

סיסמא: Qwerty12!

שם ה-database אחד מבני הזוג ללא @campus.technion.ac.il

שימו לב, יש להפוך את הסימנים {, -, .} ל-0.

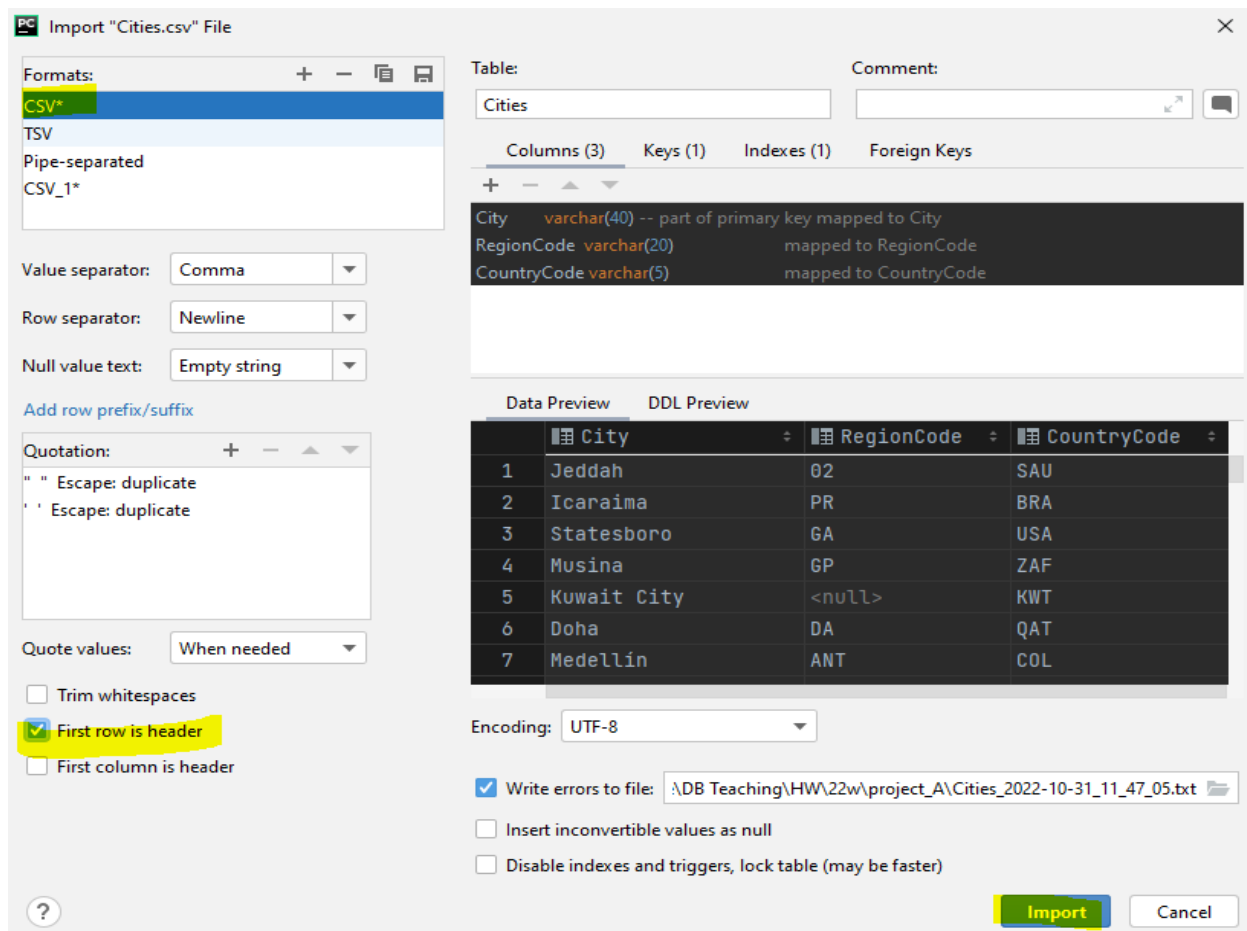
2. צרו את הטבלאות במסד הנתונים על ידי הרצת הפקודות בקובץ Create Table Commands.sql.

3. הזינו את הערכים המופיעים בקבצי ה-CSV המצורפים לתרגיל באמצעות טעינה ישירה שלהם למסד הנתונים. עליכם לעשות זאת בהתאם לסדר יצירת הטבלאות בקובץ Create Table Commands. ניתן לעשות זאת באמצעות באופן הבא:

א) לחיצה (עם הלחצן הימני של העכבר) על שם הרלציה, בחירה באפשרות של Import Data from File.

ב) בחירת קובץ ה-csv הרלוונטי מהתיקיה בה הוא יושב על מחשבכם.

ג) בתפריט שנבחר בחרו באפשרות של פורמט \*CSV, סמנו את האפשרות של First row is header ולחצו על Import (כפי שמודגש בתמונה הבאה).



The screenshot shows the 'Import "Cities.csv" File' dialog box. The 'Formats' list on the left has 'CSV\*' selected. The 'Table' field is set to 'Cities'. The 'Columns (3)' tab is active, showing the following columns:

Column	Mapping
City	varchar(40) -- part of primary key mapped to City
RegionCode	varchar(20) mapped to RegionCode
CountryCode	varchar(5) mapped to CountryCode

The 'Data Preview' tab shows the following data:

	City	RegionCode	CountryCode
1	Jeddah	02	SAU
2	Icaraima	PR	BRA
3	Statesboro	GA	USA
4	Musina	GP	ZAF
5	Kuwait City	<null>	KWT
6	Doha	DA	QAT
7	Medellin	ANT	COL

The 'First row is header' checkbox is checked. The 'Import' button is highlighted in green.

4. לאחר שטענתם את כל 4 הרלציות ניתן להתחיל בעבודה.



## שאלות (Views):

בעולם מסדי הנתונים view הינו טבלה וירטואלית הנוצרת כתוצאה מהפעלת שאלתה על טבלאות רגילות במסד הנתונים. מכיוון ש-view דומה לטבלאות הרגילות במסד הנתונים בכך שגם הוא מורכב משורות ועמודות, ניתן לשלוף ממנו מידע ולעדכן אותו בדיוק כמו טבלה רגילה. במסד הנתונים, view מוגדר על ידי שאלתת SQL. כשהמידע בטבלאות עליהן ה-view בנוי משתנה, המידע ב-view משתנה אף הוא בהתאם. נדגים את יתרונות ה-view ואופן הגדרתו בעזרת הדוגמא הבאה :

נניח ובמסד הנתונים שלנו קיימת טבלה בשם Order Details המוגדרת כדלקמן :

OrderDetails: (OrderNumber, ProductNum, QuantityOrdered, PriceEach)

כאשר :

– OrderNumber : מספר הזמנה

– ProductNum : מק"ט המוצר שהוזמן

– QuantityOrdered : כמות שהוזמנה מאותו מוצר

– PriceEach : מחיר ליחידה

כעת, נוכל לבנות view מעל טבלה זו בו יוצג לכל מוצר סכום תשלומי כלל ההזמנות עבורו :

```
CREATE VIEW SalesPerProduct
```

```
AS
```

```
SELECT ProductNum, SUM (QuantityOrdered * PriceEach) as TotalPrice
```

```
FROM OrderDetails
```

```
GROUP by ProductNum
```

יצרנו טבלה וירטואלית בשם SalesPerProduct, ובכל פעם שנרצה לדעת מהו סך ההכנסות ממוצר מסוים נוכל להריץ את השאלתה הבאה (עבור מזהה המוצר המתאים, למשל 102 בדוגמה המופיעה כאן) :

```
SELECT TotalPrice
```

```
FROM SalesPerProduct
```

```
WHERE ProductNum = 102
```

כדי להגדיר view נשתמש בפקודה CREATE VIEW ואחריה נרשום את שם טבלת ה-view (הטבלה הוירטואלית).

לאחר מכן נקבע איך view זה צריך להיראות באמצעות שאלתת SQL רגילה המופיעה לאחר האופרטור AS. \* מחיקת view מתבצעת בדומה למחיקת טבלה :

```
DROP VIEW viewname;
```

כאשר viewname הוא שם ה-view שברצוננו למחוק.

מידע נוסף (כולל syntax והסברים מפורטים) ניתן למצוא בכתובת הבאה :

[http://www.w3schools.com/sql/sql\\_view.asp](http://www.w3schools.com/sql/sql_view.asp)