



دانشگاه تهران  
پردیس دانشکده‌های فنی  
دانشکده برق و کامپیوتر



درس

# تحلیل داده و مصورسازی

تمرین شماره ۴

موضوع:

کار با پایگاه داده

نیمسال اول سال تحصیلی ۱۴۰۱ - ۱۴۰۲

## مقدمه

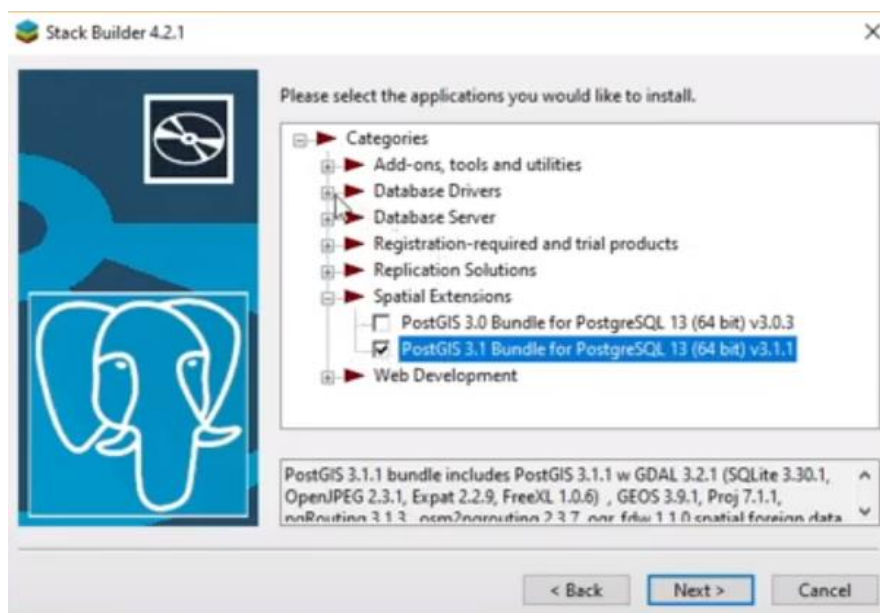
در این تمرین ما قصد داریم تا با databaseها کار کنیم و با تعدادی از ابزارهای معروف که به منظور ارتباط با databaseها استفاده می‌شود، آشنا شویم.

در این تمرین از PostgreSQL به عنوان یکی از RDBMSهای (دیتابیس‌های رابطه‌ای) رایج و حرفه‌ای استفاده می‌کنیم، و قصد بررسی و کار با آن و تعدادی از ابزارهای مرتبطش را داریم.

ابزارهایی که در این تمرین به بررسی آنها می‌پردازیم عبارتند از: pgAdmin، psycopg2 و همین‌طور سعی می‌کنیم تا با ابزار QGIS کار کنیم و یک آشنایی اولیه با آن داشته باشیم.

ابتدا شما باید PostgreSQL را بر روی سیستم خود نصب کنید و به منظور دانلود آن می‌توانید که از [این لینک](#) اقدام کنید. همچنین به منظور آشنایی با نحوه نصب PostgreSQL می‌توانید از [این ویدیو](#) استفاده کنید.

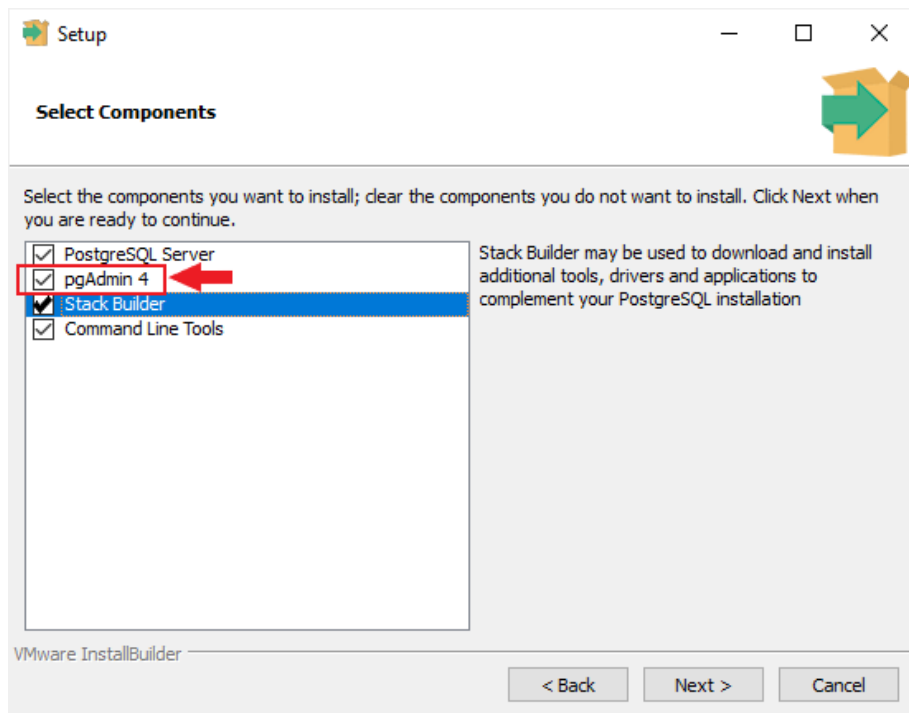
هنگام نصب PostgreSQL دقت کنید که مطابق تصویر زیر تیک مربوط به نصب ماژول PostGIS را نیز بزنید تا بتوانید در قسمت مربوط با این ماژول نیز کار کنید.



شکل ۱: نصب ماژول PostGIS (در سیستم عامل windows)

به هنگام نصب PostgreSQL نرم افزار pgAdmin نیز نصب می شود (باید component آن را هنگام نصب انتخاب کنید). اما اگر بخواهید به صورت جداگانه آخرین نسخه pgAdmin را نیز نصب کنید می توانید از طریق [این لینک](#) آن را دانلود کرده و نصب کنید.

به منظور نصب شدن pgAdmin هنگام نصب نرم افزار تیک زیر را بزنید.



شکل ۲: نصب کامپوننت های مختلف هنگام نصب PostgreSQL

همچنین به منظور نصب نرم افزار QGIS از طریق [این لینک](#) می توانید اقدام کنید.

## بخش اول – کار با psycopg2

در این قسمت ما قصد داریم تا با کتابخانه psycopg2 پایتون که به منظور تعامل با PostgreSQL ایجاد شده است کار کنیم.

در این قسمت از شما خواسته می شود که با یک دیتاست دیگر کار کنید و اقدام به طراحی یک پایگاه داده به منظور ذخیره داده های این جدول در PostgreSQL کنید.

۱. از شما به منظور ارزیابی جهت استخدام در مرکز ملی آمار ایران خواسته می‌شود که یک دیتابیس به منظور نظم‌دهی و همگام‌سازی اطلاعات زاد و ولد کشور طراحی کنید. از این پایگاه داده در مراحل بعدی به منظور مقایسه و یافتن نرخ زاد و ولد در کشور و به دست آوردن بینشی در مورد روند این تغییرات قرار است استفاده شود.
- a. بدین منظور یک مجموعه دیتاست در اختیار شما قرار داده شده است که مربوط به آمار زاد و ولد در ایران به صورت فصلی از سال ۱۳۹۱ تا ۱۴۰۰ می‌باشد.
- b. این اطلاعات به صورت استانی دسته‌بندی شده است و همچنین اطلاعات جنسیت افراد را نیز شامل می‌شود.
- c. شما باید از این دیتاست و با توجه به مفاهیمی که تاکنون از پایگاه داده فراگرفته‌اید اقدام به طراحی یک پایگاه داده به منظور ذخیره این اطلاعات کنید به صورتی که این پایگاه داده این قابلیت را داشته باشد تا توسط سازمان آمار بنابر اطلاعاتی که به صورت فصلی از استان‌های مختلف دریافت می‌کند، آن را به روز کند و ولادت‌های جدید را در آن ثبت نماید.
- d. به منظور طراحی این دیتابیس می‌توانید از سایت [DBDesigner](#) استفاده کنید. این وبسایت به شما این امکان را می‌دهد که نمودار ER طراحی کنید و سپس خروجی SQL مطابق هر RDBMS ای که با آن کار می‌کنید بگیرید. (این وبسایت قبلاً کاملاً رایگان بود اما مدتی است برای طراحی دیتابیس‌های بزرگ نیاز به پرداخت هزینه دارد اما برای پروژه شما میتوان از همان نسخه رایگان استفاده کرد).
- e. به منظور آشنایی بیشتر با نمودارهای ER پیشنهاد می‌کنم که [این پلی‌لیست](#) یوتیوب را مشاهده کنید.
- f. پس از طراحی نمودار ER و ساخت و طراحی پایگاه داده اقدام به وارد کردن داده‌ها به داخل پایگاه داده کنید. (طراحی ER اجباری نیست و صرفاً برای راحت‌تر شدن طراحی جداول است و می‌توانید که مستقیماً کد جداول را بنویسید).
- g. توجه: هدف ما در این درس با توجه به این که دوره تخصصی پایگاه داده نیست، این نیست که شما یک پایگاه داده را به صورت اصولی طراحی کنید و فرم‌های نرمال مختلف و سایر ویژگی‌های یک سیستم پایگاه داده را به صورت دقیق و کامل پیاده کنید به همین دلیل در زمینه طراحی پایگاه داده سخت‌گیری زیادی نخواهد شد.



۲. پس از ساخت دیتابیس و آماده کردن آن اقدام به نوشتن query های زیر با استفاده از Psycopg2 و PostgreSQL کنید و سپس خروجی مربوطه را به صورت خواسته شده نمایش دهید.

a. مجموع ولادت کل کشور در سال های مختلف را بدست آورید و با استفاده از bar-chart آن را نمایش دهید. سپس اقدام به تحلیل نمودارهای رسم شده نمایید. (مواردی از قبیل این که در کدام سال بیشترین میزان ولادت وجود داشته است و روند کلی کشور به چه سمتی می رود. همچنین به منظور مقایسه بهتر اگر روش یا تکنیکی به نظرتان می رسد پیاده کنید).

b. نمودار قسمت قبل را مجدد به صورت stack bar-chart و به تفکیک زن و مرد رسم کنید. مجدد در این قسمت نیز سعی کنید تا تحلیل های قسمت قبل را براساس جنسیت انجام دهید.

c. در پنج سال پایانی داده های داده شده به شما تعداد ۵ استان دارای بیشترین میزان ولادت را در نمودارهای جداگانه نمایش دهید و این کار را براساس تفکیک جنسیت نیز انجام دهید. و سپس تحلیل و مقایسه کنید.

d. مانند قسمت قبل در این قسمت کارهای گفته شده را برای کمترین میزان ولادت انجام دهید. سپس تحلیل و مقایسه کنید.

e. اکنون در این قسمت با استفاده از داده های موجود در فایل Iran\_population.csv که اطلاعات مربوط به جمعیت هر استان در آن قرار دارد، اقدام به انجام مجدد قسمت C کرده با این تفاوت که این بار به جای مرتب سازی بر اساس تعداد ولادت، براساس نرخ ولادت (نسبت تعداد متولدین بر جمعیت استان مورد نظر) مرتب سازی کنید. در انتها نیز اقدام به تحلیل و مقایسه کنید.

i. توجه کنید که به منظور استفاده از این داده ها دو راه مختلف برای شما وجود دارد. راه اول: این است که دیتای مربوطه را در پایگاه داده و در جدول مربوط به هر استاد به عنوان یک فیلد جدید جمعیت اضافه کنید. راه دوم: دیتا را در پایگاه داده ذخیره نکرده و صرفا اقدام به استفاده از اطلاعات موجود در آن



براساس نیاز خود در لحظه کنید. می‌توانید از هر یک از دو روش بدین منظور استفاده کنید.

f. همانند قسمت قبل کارهای خواسته شده را برای قسمت d نیز انجام دهید. در انتها نیز اقدام به تحلیل و مقایسه بپردازید.

g. اکنون تعداد و سپس نرخ ولادت در سال‌های مختلف را براساس فصل و همچنین جنسیت بدست آورده و با نمودار مناسب نمایش دهید. چه نتیجه‌ای می‌گیرید؟ در چه فصلی معمولاً بیشترین میزان ولادت وجود دارد؟ (انجام این کار تنها برای پنج سال دلخواه نیز کافی است)

- توجه شود که تحلیل و مقایسه‌های خواسته شده تاثیر بسزایی در ارزیابی نهایی این تمرین شما دارد، بنابراین به آن‌ها توجه کنید.

۳. تحقیق کنید که مشکل استفاده زیاد و متوالی از `cursor.connect(**params)` چیست؟ یکی از راه‌های رفع این مشکل را بیان کنید.

## بخش دوم – کار با QGIS (امتیازی\*)

در این سوال هدف یک آشنایی اولیه با نرم‌افزار QGIS و همچنین اتصال یک پایگاه‌داده PostgreSQL به این نرم‌افزار از طریق ماژول PostGIS است.

QGIS یک نرم‌افزار سیستم اطلاعات جغرافیایی متن‌باز و cross-platform می‌باشد که به ما امکان مشاهده، تحلیل و ویرایش داده‌های جغرافیایی را می‌دهد.

به منظور آشنایی با این نرم‌افزار و همچنین قابلیت‌های مختلف آن در امور جغرافیایی و نقشه‌برداری و... پیشنهاد می‌کنم تا [این پلی‌لیست](#) که مربوط به یک آموزش نسبتاً کامل از این نرم‌افزار توسط خانم دکتر کفاش استاد دانشگاه تبریز است را در آپارات مشاهده کنید.

A: سه عدد فایل SQL به نام‌های `Iran_States.sql`، `Iran_County.sql` و نیز `Iran_Country.sql` به همراه پروژه شما آپلود شده است. شما باید با استفاده از این فایل‌ها مشابه سوال اول یک پایگاه داده بسازید و دستورات مورد نیاز را برای وارد کردن جدول‌های مربوط به مختصات‌های جغرافیایی استان‌ها،

شهرستان‌ها و کل کشور ایران وارد کنید تا جداول مربوطه ساخته شوند. سپس در نرم‌افزار QGIS از طریق ماژول PostGIS اقدام به ایجاد یک ارتباط جدید با پایگاه‌داده ساخته شده در PostgreSQL کنید و به این طریق داده‌ها را داخل نرم‌افزار QGIS بارگزاری کنید. پس از انجام این موارد می‌توانید با استفاده از داده‌های بارگزاری شده لایه‌های جدید ایجاد کنید و این موارد را بر روی نقشه مشاهده کنید.

برای مشاهده نقشه‌ها پیشنهاد می‌کنم که لایه OpenStreetMap را نیز در نرم‌افزار QGIS فعال کنید.

توجه: بعد از ساختن پایگاه‌داده باید حتماً با استفاده از دستور `CREATE EXTENSION postgis` اقدام به نصب ماژول مربوطه در pgAdmin کنید تا بتوانید با نرم‌افزار QGIS تعامل داشته باشید.

B: پیدا کردن موقعیت: در این قسمت به شما یک موقعیت داده می‌شود و شما باید با نوشتن query مورد نظر محل (مانند اسم استان، شهرستان و...) را بدست آورید.

توجه: سیستم CRS (Project Coordinate Reference System) استفاده شده در داده‌هایی که در اختیار شما قرار گرفته شده است، EPSG:4326 می‌باشد.

a. مشخص کنید که موقعیت (51.3380650 35.6997300) کدام استان و شهرستان و محل را مشخص می‌کند. (یک تصویر از محل مشخص شده را نیز ترجیحاً بر روی نقشه QGIS نمایش دهید.)

b. موارد قبل را برای مختصات (52.8904157 29.9351594) تکرار کنید.

c. اکنون به همان شیوه قبل فایل `Iran_inland_waters.sql` را نیز وارد پایگاه‌داده خود کنید. این فایل شامل داده‌های مربوط به آب‌های درون سرزمینی ایران است. این پایگاه داده را نیز در QGIS مشابه قبل بارگزاری کنید و آن را مشاهده کنید.

d. دو مختصات (51.270365 35.728517) و (45.4533 37.6186) مشخص کننده چه دریاچه‌هایی هستند و همچنین در کدام استان و شهرستان قرار دارند؟ (ترجیحاً محل دریاچه را بر روی نقشه QGIS نیز نمایش دهید.)



## نکات تحویل

- مهلت ارسال این تمرین تا پایان روز شنبه ۵ آذر ماه خواهد بود.
- انجام این تمرین به صورت یک نفره می باشد.
- فایل Jupyter Notebook ای به عنوان تمپلیت تمرین تحت عنوان DA\_HW4.ipynb به همراه تمرین شما قرار داده شده است که می توانید از آن به عنوان گزارش کار خود استفاده کنید.
- لطفا هر گونه فرض در حل سوالات را در گزارش خود ذکر کنید.
- لطفا گزارش، فایل کدها و سایر ضmannم مورد نیاز را با فرمت زیر در سامانه مدیریت دروس بارگذاری نمائید.

HW4\_[Lastname]\_[StudentNumber].zip

- در صورت وجود سوال و یا ابهام می توانید از طریق رایانامه زیر با دستیار آموزشی در ارتباط باشید:  
[p.zilouchian@ut.ac.ir](mailto:p.zilouchian@ut.ac.ir)