



## مقدمه

این تمرین از دو بخش تشکیل شده است:

در بخش اول 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.

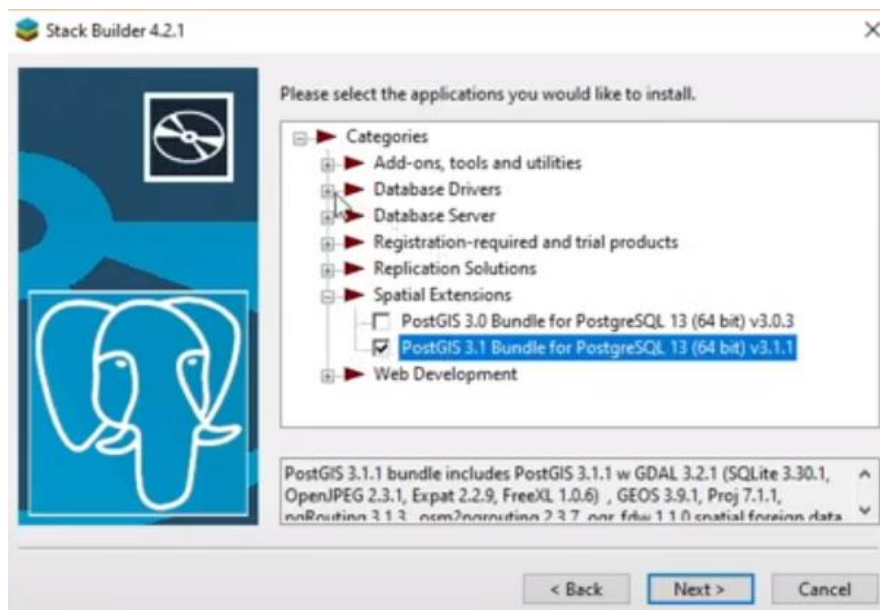
در بخش دوم تمرین شما هم هدف این است که یک پروژه‌ای به صورت ترکیبی از کار تحلیل داده با استفاده از پایگاه داده انجام بدهید و خروجی‌های مربوطه را با استفاده از Visualizationهای مناسب با استفاده از کتابخانه‌های مربوطه در پایتون نمایش و گزارش دهید تا بتوانید به خوبی تحلیل را انجام دهید. به این صورت که برای دیتاست داده شده به شما یک پایگاه داده طراحی می‌کنید و سپس اقدام به نوشتن Query با استفاده از SQL کرده و سپس اقدام به گرفتن خروجی‌ها مختلف متناسب با خواسته‌های قسمت‌های مختلف سوال می‌کنید.

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

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

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

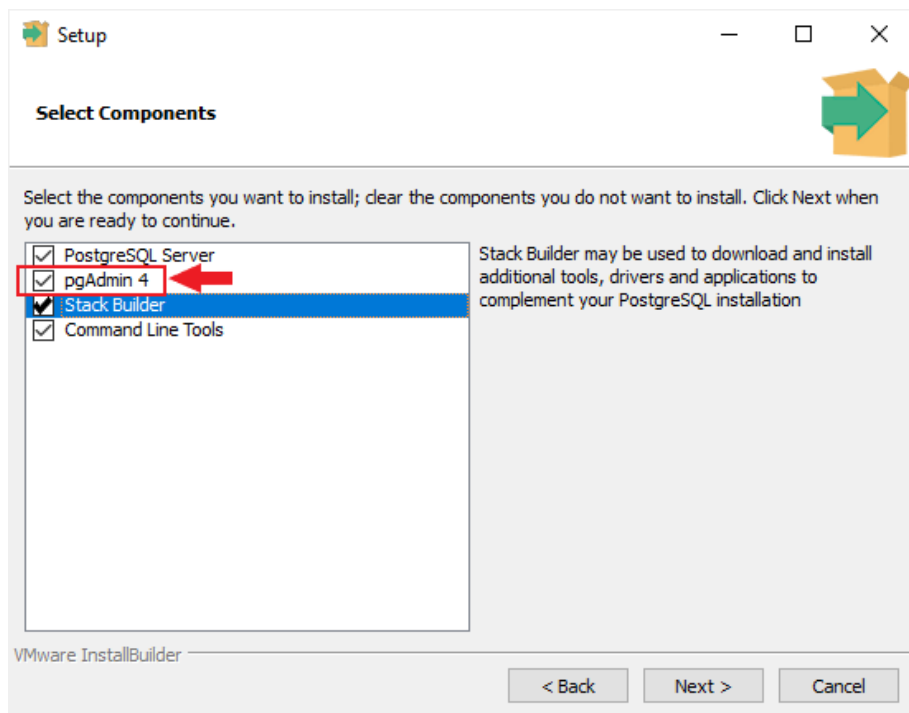
هنگام نصب PostgreSQL دقت کنید مطابق تصویر زیر تیک مربوط به نصب ماژول PostGIS را نیز بزنید تا بتوانید که در قسمت مربوط با این ماژول نیز کار کنید.



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 که از طریق [این لینک](#) آن را دانلود کرده و نصب کنید.

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



شکل ۲: نصب کامپوننت‌های مختلف هنگام نصب PostgreSQL

همچنین به منظور نصب نرم‌افزار QGIS از طریق [این لینک](#) می‌توانید اقدام کنید.

## بخش اول – آشنایی با SQL

در این بخش از تمرین آشنایی با دستورات پایه SQL هست. و همچنین اگر قصد کار تخصصی در حوزه نرم‌افزار و داده را دارید، آشنایی با SQL به صورت تخصصی جزء ملزومات اصلی این کار خواهد بود و در این تمرین هم برآنیم که تا حد امکان به این هدف برسیم.

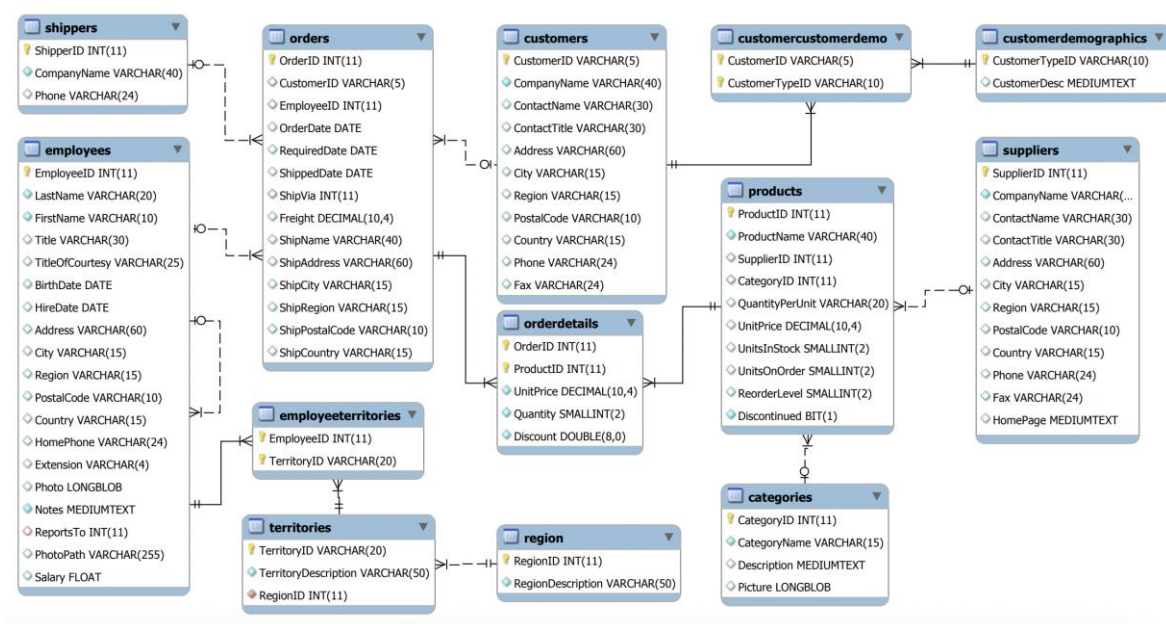
دیتابسی که برای این تمرین در نظر گرفته شده است، دیتابیس معروف آموزشی میکروسافت با نام 'Northwind' است که هر چند کمی قدیمی است اما برای اهداف آموزشی ما، بسیار مناسب است و نسخه پستگرس آن را در این تمرین استفاده خواهیم کرد.

این دیتابیس که یک فروشگاه بزرگ را مدل‌سازی می‌کند، از جداول اصلی زیر تشکیل شده است:

<sup>۱</sup> <https://github.com/Microsoft/sql-server-samples/tree/master/samples/databases/northwind-pubs>

- **Customer**: مشتریان فروشگاه
- **Order**: سفارشات هر مشتری – اطلاعات کلی
- **OrderDetails**: جزییات هر سفارش شامل آیتم‌های خریداری شده
- **Products**: محصولات فروشگاه
- **Suppliers**: تامین کنندگان کالا که محصولات از آن‌ها تهیه می‌شوند.
- **Categories**: گروه‌بندی کالاها
- **Shippers**: شرکت‌های حمل و نقلی که وظیفه ارسال کالا به دست مشتری را بر عهده دارند.
- **Employee**: کارمندان فروشگاه که هر سفارش، توسط یک کارمند ثبت یا تایید می‌شود.
- **Territory & Region**: مناطق و نواحی‌ای که این فروشگاه در آن‌ها شعبه دارد. هر منطقه شامل چندین ناحیه است و هر کارمند، می‌تواند به سفارشات چندین ناحیه رسیدگی کند.
- **CustomerDemographics**: برای رتبه‌بندی مشتریان استفاده می‌شود که در این تمرین با آن سر و کار نخواهیم داشت.

حال در ادامه می‌توانید که نمودار ERD این database را مشاهده کنید:



شکل ۳: نمودار ERD پایگاه داده Northwind

در ابتدا باید که جداول این دیتاست به همراه دیتای موجود در آن را وارد database ساخته شده در PostgreSQL کنید.



برای شما دو فایل به نام‌های northwind.sql و northwind\_data.sql فراهم شده است که فایل اول دستورات مربوط به ایجاد جداول و فایل دوم دستورات مربوط به وارد کردن دیتاهای اولیه به داخل دیتابیس هست. شما باید که این دو فایل را محتوایشان را وارد یک دیتابیس PostgreSQL می‌سازید کنید. (ابتدا باید با فایل اول جداول را ساخته و سپس با فایل دوم دیتا را وارد کنید). در این قسمت می‌توانید از آموزش ویدیویی که برای شما فراهم شده است استفاده کنید.

به منظور اجرای queryها در این بخش از تمرین می‌توانید که از دو روش مختلف استفاده کنید. روش اول این است که ماژول مربوط به اجرای دستورات SQL در Jupyter Notebook را نصب کنید و سپس اقدام به اجرای دستورات SQL در Jupyter Notebook کنید. و نیز می‌توانید که مطابق روش گفته شده در ویدیو queryهای خود را در قسمت مربوطه در pgAdmin اجرا کنید. هر دو روش یکسان هست و فقط باید در گزارش کد مربوطه به همراه قسمتی از خروجی را قرار دهید. به منظور آشنایی با نحوه نصب PostgreSQL می‌توانید که از [این ویدیو](#) استفاده کنید.

A: ابتدا queryها و کدهای نوشته شده در فیلم HandsOn مربوطه را خودتان در قسمت‌های مربوطه در فایل نوتبوک بنویسید. (یا این که می‌توانید در PgAdmin نوشته و خروجی به همراه دستور نوشته شده را در گزارش خود بیاورید).

B: در ادامه Query های زیر را در یک نوتبوک پایتون یا این که PgAdmin نوشته و خروجی مربوطه به همراه query نوشته شده را در گزارش خود بیاورید.

۱. چه مناطقی در فروشگاه داریم. (region)
۲. می‌خواهیم ببینیم به ازای هر منطقه (region)، چه نواحی‌ای (territories) تعریف شده است. فقط نام منطقه و نام ناحیه در خروجی باید ظاهر شود.
۳. نام و نام خانوادگی تمام کارمندان ناحیه Chicago (حرف اول آن بزرگ است) را به دست آورید.
۴. می‌خواهیم سفارش جدیدی برای محصولات که تعداد آنها (UnitsInStock) از حد مجاز (ReorderLevel) کمتر شده است، بدهیم. این محصولات را بیابید و آنها بر اساس میزان موجودی (به صورت صعودی) نمایش دهید.
۵. می‌خواهیم براساس میزان عملکرد کارکنان، به آنها پاداش بدهیم. سه کارمندی را بیابید که بیشترین تعداد سفارش را تایید یا ثبت کرده‌اند.

۶. می‌خواهیم کارکنان Chicago و Orlando را به New York منتقل کنیم. دستور آپدیت متناظر را بنویسید. دقت کنید که تنها نام این شهرها را می‌دانیم و در دستور آپدیت هم از این سه نام استفاده کنید.

۷. به دلیل بالا رفتن بهای شیر ما باید محصولاتی را که Category آن‌ها Dairy Products هست را به میزان ۱۵٪ افزایش قیمت دهیم. (این تغییر را در جدول مربوطه اعمال کنید)

۸. همچنین تامین کننده‌ی مالک شرکت Mayumi's اقدام به افزایش ۱۰٪ ای محصولات خود کرده است. این تغییر را نیز در جدول مربوطه اعمال کنید. همچنین این تامین کننده به تعداد ۱۵ عدد از محصولات خود را پس از این افزایش قیمت به انبار ما ارسال کرده است. (موجودی مربوطه را نیز به روز کنید.) (امتیازی)

## بخش ۲: کار با psycopg2 و Visualization

در این قسمت ما قصد داریم تا با کتابخانه psycopg2 پایتون که به منظور تعامل با PostgreSQL ایجاد شده است کار کنیم. ابتدا مطابق راهنمایی‌های موجود در فایل PostgreSQL\_HandsOn.ipynb اقدام به نصب کتابخانه‌های موجود و همین‌طور آشنایی با روال کارکرد این کتابخانه کنید.

A: ابتدا اقدام به بازنویسی موارد جای خالی در فایل HandsOn مطابق با فیلم مربوطه کنید.

B: در این قسمت از شما خواسته می‌شود که با یک دیتاست دیگر کار کنید و اقدام به طراحی یک پایگاه داده به منظور ذخیره داده‌های این جدول در PostgreSQL کنید.

- از شما به منظور ارزیابی نحوه رویارویی و نیز عملکرد ایران در مقابله با ویروس کرونا و واکسیناسیون در مرکز ملی آمار ایران خواسته می‌شود که یک دیتابیس به منظور نظم‌دهی و همگام‌سازی اطلاعات مرگ و میر کشور طراحی کنید.

a. با توجه به شیوع ویروس کرونا سعی داریم که داده‌های کرونا در جهان را بررسی کنیم.

شما می‌توانید از [این لینک](#) داده‌های مورد نظر را دانلود کنید. توضیحات درباره feature

های مختلف این داده را نیز می‌توانید از [اینجا](#) مطالعه کنید. سپس ستون‌هایی را که اطلاعات در هر میلیون دارند را حذف کنید.

- b. 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 پیوسته در حال دریافت آن است، بتواند که آن را به روزرسانی کند و از آن بهره‌برداری لازم را انجام دهد.
- c. به منظور طراحی این دیتابیس می‌توانید که از سایت [DBDesigner](http://DBDesigner.org) استفاده کنید. این وبسایت به شما این امکان را می‌دهد که نمودار ER را 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 این پروژه شما همان نسخه رایگان آن کافی است و نیاز را برطرف می‌کند).
- d. به منظور آشنایی بیشتر با نمودارهای ER پیشنهاد می‌کنم که [این پلی لیست](#) یوتیوب را مشاهده کنید.
- e. 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).
- f. 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 سعی بر این است که سخت‌گیری زیادی نشود.
- C: حال در ادامه تمرین در بخش بعدی قصد داریم تا از این پایگاه داده طراحی شده استفاده کرده و سپس با استفاده از روش‌های مختلف مصورسازی داده به تحلیل داده بپردازیم.

## بخش ۲-۱: بررسی کرونا در ایران و مقایسه با کشورهای مختلف

۱. نمودار تعداد مبتلایان، تعداد فوتی‌ها و تعداد واکسینه شده‌های ایران را بصورت هفتگی در طول زمان در یک نمودار رسم کنید. در این نمودار تاریخ‌ها را به فرمت شمسی بنویسید و axis grid را هم به

ازای هر ماه رسم کنید. اگر داده‌های مربوط به این هفته هنوز کامل نشده است، باید این هفته آخر را بصورت متمایزی نمایش دهید. سپس نمودار را تحلیل کنید.

۲. فرض کنید ما زمانی که تعداد مبتلایان به کرونا افزایش یابد را به عنوان شروع یک پیک در نظر بگیریم. بررسی کنید که در زمان پیک کرونا و در شرایط عادی تر میزان سختگیری دولت چه تفاوتی داشته است؟ آیا میزان سختگیری کم باعث ایجاد موج جدید می‌شود؟

۳. تاثیر ویژگی‌های تراکم جمعیت، میانگین سنی، وجود امکانات بهداشتی، تعداد تخت بیمارستان‌ها و شاخص پیشرفت انسانی را بر تعداد فوتی‌ها با رسم correlation heatmap بررسی کنید. در این نمودار باید color bar مشخص باشد و min آن -1 به رنگ قرمز و max آن +1 به رنگ آبی باشد. همچنین بر روی color bar عنوان داشته باشد.

۴. می‌خواهیم تاثیر واکسیناسیون بر تعداد فوتی‌ها را بررسی کنیم. برای این کار فرض کنید لازمست که اطلاعات ۱۰ کشور را بررسی کنیم. شما کدام کشورها را برای مقایسه انتخاب می‌کنید؟ چرا؟ نمودارهای مرتبط برای رسیدن به این نتیجه را رسم و تحلیل کنید. (در این نوع نمودارها توجه کنید که بهتر است از تعداد فوتی‌ها برای تحلیل استفاده کنیم و یا تعداد فوتی‌ها در جمعیت)

۵. تعداد فوتی‌های شش ماه اخیر کشورهای مختلف را به نسبت جمعیت آن‌ها بر روی نقشه نمایش دهید. Color bar هم به نمودار اضافه کنید. (می‌توانید از لینک‌های ۱ و ۲ برای رسم این نمودار کمک بگیرید)

۶. برای سوالهای مشابه با سوال ۴ و ۵ استفاده از coropleth map و یا از scatter plot (همانند مثال اول این لینک) را پیشنهاد می‌کنید؟ هر کدام برای چه مسائلی مناسب هستند؟

نکته مهم در پیاده‌سازی این بخش از تمرین این است که:

- به منظور انجام قسمت visualization باید از پایگاه داده طراحی شده استفاده کنید و Query مربوطه را نوشته و با استفاده از خروجی Query نمودار خواسته شده را رسم کنید.



- در صورت لزوم تمیزسازی داده انجام شود و دلایل آن توضیح داده شود. برای مثال اگر بخش‌هایی از داده وجود ندارد، دلیل آن بیان شود و در صورت نیاز با روش مناسبی این مشکل برطرف گردد.
- برای انتخاب انواع مختلف نمودار و سایر فرض‌های انجام شده توضیحاتی ارائه نمایید.
- برای رسیدن به نتایج قابل اتکا لازم است feature های مناسبی انتخاب شوند تا اثر confounding factor ها به حداقل برسد. برای این مورد نیز در صورت نیاز دلایل خود را بیان کنید.
- در رسم نمودارها به رنگ‌ها، فونت‌ها، سایزها، تمیزی و... توجه شود. در صورتی که از زبان فارسی و فونت خاصی برای متن‌های به کار رفته در نمودارها استفاده کنید، نمره مثبت محسوب می‌شود.
- انتخاب نوع static یا interactive بودن نمودارها و انتخاب کتابخانه‌ها به عهده خودتان است.
- نمودارهای بدون تحلیل هیچ ارزشی ندارند و در هر بخش باید تحلیل خود را ارائه کنید.

### بخش سوم – کار با QGIS (امتیازی\*)

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

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 هست 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 که این فایل‌ها را به صورتی که پایگاه داده سوال اول را ساختن یک پایگاه داده جدید نیز برای این سوال بسازید و سپس این دستورات را برای وارد کردن جدول‌های مربوط به مختصات‌های جغرافیایی استان‌ها، شهرستان‌ها و کل کشور ایران وارد کنید تا جداول مربوطه ساخته شوند. و سپس اکنون در نرم‌افزار QGIS از طریق ماژول PostGIS اقدام به ایجاد یک ارتباط جدید با پایگاه داده ساخته شده در postgresql کنید و به این طریق دیتاها را داخل نرم‌افزار QGIS لود کنید. پس از انجام این موارد اکنون می‌توانید که با استفاده از دیتاهای لود شده لایه‌های جدید ایجاد کنید و این موارد را بر روی نقشه مشاهده کنید.



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

توجه: بعد از ساختن پایگاه داده باید حتماً با استفاده از دستور `CREATE EXTENSION postgis` اقدام به نصب ماژول مربوطه در pgAdmin کنید تا این که 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 (51.3867718 35.7242345) کدام استان و شهرستان را مشخص می‌کند.

b. اکنون به همان شیوه قبل فایل `Iran_inland_waters.sql` را نیز وارد پایگاه داده خود کنید و داده‌های مربوط به آب‌های درون سرزمینی ایران را می‌توانید که مشاهده کنید. این پایگاه داده را نیز در QGIS مشابه قبل لود کنید و مشاهده کنید.

c. اکنون بیان کنید که دو مختصات (53.6166 31.0624) و (61.2768 31.2736) مشخص کننده چه دریاچه‌هایی هستند. و همچنین در کدام استان و شهرستان قرار دارند.



## نکات تحویل

- مهلت ارسال این تمرین تا پایان روز سه شنبه ۳۰ آذر ماه خواهد بود.
- انجام این تمرین به صورت یک نفره می باشد.
- لطفا هر گونه فرض در حل سوالات را در گزارش خود ذکر کنید.
- لطفا گزارش ، فایل کدها و سایر ضمائم مورد نیاز را با فرمت زیر در سامانه مدیریت دروس بارگذاری نمائید.

HW3\_4\_[Lastname]\_[StudentNumber].zip

- در صورت وجود سوال و یا ابهام می توانید از طریق رایانامه های زیر با دستیاران آموزشی در ارتباط باشید:
- در ایمیل خود هر دو دستیار را به عنوان گیرنده قرار دهید تا به سوال مربوطه شما پاسخ دهند.

[p.zilouchian@ut.ac.ir](mailto:p.zilouchian@ut.ac.ir)

[m.vatandoust@ut.ac.ir](mailto:m.vatandoust@ut.ac.ir)