Peer-graded Assignment: Data Scientist Role Play: Profiling and Analyzing the Yelp Dataset

Submit by Sep 14, 3:59 AM -03

Submit your assignment soon
Even though your assignment is due on Sep 14, 3:59 AM -03, try to submit it 1 or 2 days early if you can. Submitting early gives you a better chance of getting the peer reviews you need in time.

Instructions   My submission   Discussions

Welcome to the final assignment of our course! This assignment is designed to test your knowledge of a wide range of concepts and SQL design techniques discussed throughout the course. For this assignment, specifically, you are going to be **playing the role of a real-world data scientist** using SQL to both answer specific questions for an organization and make inferences based on your discoveries.

We will be using a dataset from a US-based organization called Yelp, which provides a platform for users to provide reviews and rate their interactions with a variety of organizations – businesses, restaurants, health clubs, hospitals, local governmental offices, charitable organizations, etc. Yelp has made a portion of this data available for personal, educational, and academic purposes.

You will be asked a series of questions regarding the data to help you profile and better understand the data in the first part of the assignment. Once you have answered each question, you will come up with your own question for analysis and will prepare a dataset for the analysis you choose to do in the second part of the assignment.

Naturally, the questions will require you to write a variety of SQL statements. The reading in this lesson entitled, **"Yelp Data Set SQL Lookup,"** will allow you to run all of the queries you need to against the Yelp Data Set below. Also, be sure to check out the **Yelp Dataset ER Diagram** and instructions below for more detailed information on the dataset and how to complete the assignment. The Grading Criteria Overview section includes **the Worksheet** to be filled out and turned in.

Review criteria   **less** ⌃

This is a 2-part assignment. In the first part, you are asked a series of questions that will help you profile and understand the data just like a data scientist would. For this first part of the assignment, you will be assessed both on the correctness of your findings, as well as the code you used to arrive at your answer. You will be graded on how easy your code is to read, so remember to use proper formatting and comments where necessary.

In the second part of the assignment, you are asked to come up with your own inferences and analysis of the data for a particular question you want to answer. You will be required to prepare the dataset for the analysis you choose to do. As with the first part, you will be graded, in part, on how easy your code is to read, so use proper formatting and comments to illustrate your intent as required.

For both parts of this assignment, use the following "worksheet." It provides all the questions you are being asked, and your job will be to transfer your answers and SQL coding into this worksheet where indicated so that your peers can review your work. You should be able to use any Text Editor (Windows Notepad, Apple TextEdit, Notepad ++, Sublime Text, etc.) to copy and paste your answers.

> YelpDataCourseraPR2.txt

Once you have completed all your queries and finished the worksheet, upload your completed copy using the appropriate area below. You will then be asked to review several of your peers' worksheets.

Yelp Dataset ER Diagram   **less** ⌃

The entity relationship (ER) diagram below, should help familiarize you with the design of the Yelp Dataset provided for this peer review activity.