6.1 Data Analysis Lab pt. 1

Instructor Guide

Overview

In this inquiry-based lab, students will formulate questions about one or more databases, and write the code to answer these questions. This lab involves a brief introduction to dictionaries, although its focus is not dictionaries but rather the process of asking questions about datasets and translating those questions to code. It begins with a heavy guided practice component, where the instructor walks students through the process of understanding databases and answering questions about them. Students will then work on individual practice, where they are writing code to answer questions provided by the instructor. In the second-day followup to this lab, students will ask their own questions and work with mentors to formulate code that answers those questions.

Learning Goals

- The high-level concept of a dictionary
- How to access data in a dictionary
- How to iterate over dictionaries
- How to use accumulators to store values across multiple iterations of a loop
- How to convert high-level questions to systematic steps in computer code
- Basic algorithmic thinking, such as finding the max within a data set

Personal Growth Goals

- Problem-Solving: Students will get walked through and practice a problem solving approach that is particularly useful at analyzing datasets: 1) understanding the dataset, 2) posing questions about the dataset, and 3) determining how to systematically analyze the data set to answer the questions

- Computer Programming is a Tool: In prior labs, students use computer programming to make end-to-end systems (i.e. a make-your-own-adventure game), or to create graphics. This is the first lab where they are using computer programming not to create ends, but as a means to seek out an end -- a tool to answer a larger question.

Skills Required

- Complete understanding of loops, conditionals, and variables
- Familiarity with using lists -- especially indexing into lists
- A curiosity to ask/answer questions about data

Resources Required

- Computers for either every student or every pair of students
- Python 3 and a text editor needs to be installed on all the computers
- One mentor per 2-3 students
- A projector to project the central instructor's computer

Instructor Preparation

1. Make sure all the computers students will use have Python and a text editor (right now, we use Pyzo) installed (check to see that students have a way to save/access files)
2. Load the following programming files onto each computer.
    a. 06_01_IMDB.py
    b. 06_02_IMDB.py
    c. 06_02_US_Presidents.py
    d. 06_02_NFL.py
    e. IMDBDatabase.py
    f. USPresidentsDatabase.py
    g. NFLDatabase.py

In Depth Description of Lab Activities

Phase 1: Setup

1. Before the students arrive, open the following files in a text editor on each computer:
    a. 06_01_IMDB.py
    b. 06_02_IMDB.py
    c. 06_02_US_Presidents.py
    d. 06_02_NFL.py
    e. IMDBDatabase.py
    f. USPresidentsDatabase.py

g. NFLDatabase.py

Phase 2: Introduction | Review

1. Review the following topics verbally with students as they enter the class.
    a. Lists: high-level concept, indexing
    b. Loops -- how to loop over a list, how to initialize a variable outside a loop and update it inside

Phase 3: General Lecture & Guided Practice

1. This lecture will be taught by a central instructor, who will write up code while his/her computer is projecting for the whole class to see (or write Python code on a board). Students will follow along on their computers. The central instructor will go over Challenges 1.1 - 1.5 in front of the whole class. Content to cover before and while walking students through Challenges 1.1 - 1.5:
    a. Dictionaries
        i. What is a dictionary?
            1. Similar to a list, except instead of accessing information using its location (index), you access it using some other values, such as a string.
            2. Example: an actual dictionary
        ii. Syntax for accessing elements from a dictionary
            1. Same as indexing into a list, except you use the key instead of the index
        iii. Nested Dictionaries
            1. Sometimes, it is not enough to just store data with one level of association. For example, you might want to store the names of all players on a hockey team, and also store each player's statistics. For this you use a nested dictionary -- a dictionary of dictionaries
            2. Example: the glossary of a book. You look up a word, it gives you page numbers, and you have to go to the page numbers to find the instances of the word. This is two levels of association.
            3. Accessing nested dictionaries: hockeyTeams["Penguins"]["Sidney Crosby"]["Age"]
        iv. Looping over dictionaries:
            1. You can loop over every key in a dictionary using "for...in..."
    b. How to reason about simple algorithms that loop over a database
        i. What are you looping over?
        ii. What do you want to keep track of while looping?
        iii. How should you initialize those variables?
        iv. In what cases do you modify the variables?
        v. Once the loop is done, is there any final modification you have to do to the variables?

Phase 4: Independent Practice

1. Students should have completed 1.1-1.5 while following along with the teacher, asking questions along the way. They will now complete challenges 1.6-1.7, working with their peers/mentors as necessary to determine how to write the code.

Phase 5: Start Self-Guided Inquiry

1. If students have extra time and have finished the bonus activity, they can start the self-guided inquiry portion of the lab, which is intended for the next session.
2. To do so, students need to pick a database of interest: IMDB, US Presidents, or NFL. They will load the corresponding 06_02 file for that database, begin by understanding the database and completing the warmup activity, and then begin asking and answering their own questions about the database (with the help of peers/mentors as necessary).

Phase 6: Pack up | Review

1. Then, with the last five minutes that they have, have the students discuss among themselves the following questions (in no particular order):
    a. What are other questions they might want to ask about these databases?
    b. What other databases might they be interested in having access to, and what would they want to answer about that?
    c. Now that they have used data analysis to reveal trends in datasets, how can they use knowledge of those trends? What are practical applications of data analysis?
2. These questions may be useful to use this as a form of review, and can also be used to increase interest in the subject.

Lesson Plan

(:10) means that this part should be done by the tenth minute of the lesson

1. Setup (:0)
2. Introduction | Review (:10)
3. General Lecture & Guided Practice (:30)
4. Independent Practice (:55)
5. Pack up | Review (:End)

Take Away

After completing this lab, students should recognize the high-level idea of dictionaries and, more generally, data storage. They should have an idea of how to begin understanding unfamiliar databases, how to ask questions about data sets, and how to translate those questions into code.