

# 6.1 Data Analysis Lab pt. 2

## Instructor Guide

[Overview](#)

[Learning Goals](#)

[Personal Growth Goals](#)

[Skills Required](#)

[Resources Required](#)

[Instructor Preparation](#)

[In Depth Description of Lab Activities](#)

[Lesson Plan](#)

[Take Away](#)

## 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. At this point, students should have already completed the first part of the lab, which involved guided and individual practice in understanding the IMDB Database and analyzing it using code. In this lab, students will briefly remind themselves of what they did last time, and then jump into formulating and writing code to answer their own questions with a database of their choice.

## 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. Dictionaries: high-level concept, accessing elements
  - b. Loops: how to loop over a dictionary, how to initialize a variable outside a loop and update it inside
  - c. Data Analysis: example questions that can be asked about data sets, and how to approach writing code to answer them.

## Phase 3: Finish Independent Practice

1. If students have not yet finished challenges 1.1-1.7, they should work with peers/mentors to finish them.

## Phase 4: Warm Up

1. Individual students should decide which dataset they want to work with. There are three options: they can continue working with the IMDB Dataset, or investigate trends in a US Presidents dataset or an NFL dataset.
2. Once students have decided, have them open the appropriate 06\_02 file. Students should then complete 2.0-2.1, which are warmups intended for the students to familiarize themselves with the database, and recall what they learned last lab about writing code to analyze a database.

## Phase 5: Self-Guided Inquiry

1. Once students are done with the warmup, they should move on to challenge 2.2, which invites students to ask questions about the dataset (with sample questions provided), reason about how to write code to answer the question, and write that code.
2. Mentors should be on-hand to help students formulate questions, help them reason about how to convert it to code (especially if they choose complex questions, like those that require a dictionary as an accumulator), and help them write/debug their code.

## 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 (:25)
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 know how to begin understanding unfamiliar databases, how to ask questions about data sets, and how to translate those questions into code. They should also have an idea of why data analysis is useful, and what some social applications of data analysis are.