Data Analysis Lab

Instructor Guide

Overview

The lab introduces students to lists and dictionaries in Python. The central idea is that students will use lists and dictionaries to store, analyze, and do analysis on some data. At the beginning of the lab mentors should talk with their students about what topic the student might be interested in analyzing data about. Sample topics include: information about countries (demographics, GDP, etc.), characteristics of movies (actors, rating, genre, etc.), data surrounding one or more sports, etc. All the sample coding files are written using the lens of favorite movies, but they should be able to give the instructor an idea about how to convert topics the students are interested in into a lesson about lists and dictionaries.

Learning Goals

- Concept of lists (an ordered collection of objects)
- Concept of dictionaries (a collection of objects, but instead of accessing objects based on an order you access them based on other unique descriptors, such as looking up words in a dictionary)
- Creating lists/dictionaries in Python
- Looping over lists/dictionaries
- Mutability and destructive/nondestructive modifications
- Finding elements in list
- Using lists with/in functions & List properties (len, min, max, sum)
- Using dictionaries with/in functions & dictionaries properties (len, copy, popitem, etc.)
- Using dictionaries as rudimentary databases

Personal Growth Goals

- Students will learn how to think about a problem in a computational, step-by-step manner
- Students will realize that everyday pieces of technology that they use to access information, such as IMDB and/or databases like Google, are implemented in relatively simple ways that they are capable of understanding and emulating

Skills Required

- An understanding of the print function
- A basic understanding of functions and methods
- An understanding of loops and conditionals

Resources Required

- One mentor per student or per 2 students.
- One computer per student with Python 3 and an editor installed
- A projector to project the central instructor's computer

Instructor Preparation

- Load the programming files onto each computer
- Each instructor should have access to the sample programming files (labelled 1 to 4 in the GitHub folder), so they know what key information to cover (the sample programming files serve as a script of sorts).

In Depth Description of Lab Activities

This lab will be taught by a central instructor, who will write up code while his/her computer is projecting for the whole class to see. Students will follow along on their computers, and mentors will make sure they stay on track. There are checkpoints built into each activity to allow students to practice the skills they learnt and give mentors the time to make sure the students have understood what is being taught.

Before starting the lab, mentors should discuss with students to determine what topics are interesting to them. Examples include movies, sports, geography, historical figures, etc. Basically, anything that has enough "data" to be organized into lists and dictionaries, and that the mentor and/or student are familiar enough with to organize into lists and dictionaries. For example, with historical figures, lists can be of US presidents ordered by their inauguration date, and dictionaries can include other attributes of the president like their approval rating. Mentors

and students should feel free to look up some of the data, as long as it is readily accessible and will not take too much time out of the lab.

The central mentor will use the frame of favorite movies (i.e. what is provided in the sample programming files), but when the students and mentors emulate what the central instructor is doing, they should use the topic of their choice.

This lab is broken up into four activities: Basic Lists, Using Lists, Basic Dictionary, and Using Dictionaries.

Activity 1: Basic Lists

In this activity, instructors will introduce the following concepts through the frame of topic chosen above:

- Creating lists
- Printing lists
- Indexing into lists (zero based index)
- Adding elements to a list (by element and by index)
- Removing elements from a list (by element and by index)

A sample of what the instructor might type and/or say for each item noted above, through the frame of favorite movies, is given here: 1-favoriteMovies-sample.py

Activity 2: Using Lists

In this activity, instructors continue using the frame decided on above:

- Looping over lists
- Copying lists
- Sorting lists
- Using lists in functions

A sample of what the instructor might type and/or say for each item noted above, through the frame of favorite movies, is given here: 2-favoriteMoviesFunctions-sample.py

Activity 3: Basic Dictionaries

In this activity, instructors will introduce the following concepts through the frame decided on above:

- Creating dictionaries
- Printing dictionaries
- Indexing into dictionaries
- Adding elements to a dictionary
- Removing elements from a dictionary
- Iterating over a dictionary

A sample of what the instructor might type and/or say for each item noted above, through the frame of a movie collection, is given here: 3-categorizingMovies-sample.py

Activity 4: Using Dictionaries

This activity breaks from the classroom structure in the other activities, with a central instructor doing most of the teaching. Instead, most of the teaching in this activity will happen solely mentors and students.

The goal of this activity is to manipulate data in one of three provided datasets: IMDB, NFL, and US Presidents. The IMDB Dataset contains a dictionary of 15 of IMDB's 250 most popular films of all time. The NFL Dataset contains a dictionary of 10 of the best football teams in the 2016-2017 season (based on best offense, defense, or passing stats). The US Presidents Dataset contains a dictionary of the last 10 presidents, along with data and statistics about their presidency. The goal of this activity is for students to write functions that manipulate one of these datasets.

The central instructor should lead a discussion about what all students might want to know with each of these datasets. Mentors will then work with students to determine what they might want to know about the data. At the end, the class will come together and students will share interesting insights they found.

Example questions:

- IMDB:
    - Write a function that takes in the movie database and a genre, and returns a list of movies within that genre
    - Write a function that takes in the movie database and a rating, and returns a list of all movies with higher ratings
    - Write a function that flips the movie database around, so as opposed to searching for a movie name and getting a list of actors in that movie, you search for an actor name and get a list of the movies s/he was in.
    - Write a function that takes in the movie database, and returns a list of the movies in that database, sorted by year of release. [OPTIONAL]
- NFL:
    - Write a function that takes in the NFL database and outputs a list that orders the team by average points per game.
    - Write a function that takes in the NFL database and a touchdown percentage, and outputs the quarterbacks with a greater touchdown percentage.
    - Write a function that takes in the NFL database and outputs a list that orders the quarterbacks by number of yards per attempt.
- US Presidents:
    - Write a function that takes in the US Presidents database and outputs a list that orders presidents by average approval rating.

- Write a function that takes in the US Presidents database and outputs a dictionary where the keys are colleges and the values are the number of US presidents that attended there.
- Write a function that takes in the US Presidents database and outputs a list that orders presidents by unemployment rate in the last month of their presidency.
- Write a function that takes in the US Presidents database and outputs a dictionary where the keys are political parties and the values are lists of the names of presidents from that party.
- Write a function that takes in the US Presidents database and outputs a list that orders presidents by how long they stayed in office.

A sample of what the instructor might type and/or say for the IMDB database is given here: 4-imdb-sample.py

Lesson Plan

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

1. Setup (:10)
2. Activity 1: Basic Lists (:25)
3. Break (:30)
4. Activity 2: Using Lists (:55)
5. Activity 3: Basic Dictionaries (:75)
6. Break (:80)
7. Activity 4: Using Dictionaries (:120)