

Introduction to Computer Science I

Project: Fatal Police Shootings

Overview

The goals of this project are to:

- Gain experience using Python to analyze real-world data.
- Practice using Python dictionaries.

In this project, you will use Python dictionaries to analyze real-world data on fatal police shootings. You will use dictionaries to investigate whether black people are disproportionately the subject of fatal police shootings, when compared with their representation in the general United States population. Task 1 does not involve any Python programming. Instead, you'll gain familiarity with the data before you start working with it. For Task 2, you will read and add comments to code that has been provided to you, reading data and organizing it into a dictionary. In Task 3, you will use the dictionary that we have constructed, and write some additional code to answer questions about the data. Finally, in Task 4, you will write a short reflection on your findings. As always, be sure to use meaningful variable names and add helpful comments in your program.

Task 1: Understanding the data

For this project, you will use data on fatal police shootings that has been compiled by the Washington Post, and is publicly available on GitHub at <https://github.com/washingtonpost/data-police-shootings>. We have posted the data file, `fatal-police-shootings-data.csv`, on the course webpage, to make it a bit easier for you to access.

Before you start working with this data using Python, it's important to understand the data that we're working with. To do this, you will spend some time reading about the data, and looking at the `.csv` file in a spreadsheet program such as Excel or Google Sheets. We have provided a template file, `template.txt`, for you to record your investigations. In this file, you will write your answers to the following questions:

Question 1. Follow the link <https://github.com/washingtonpost/data-police-shootings> to read about the data. Write a three sentence summary of what is included in this data, and where it comes from.

Question 2. Open the file `fatal-police-shootings-data.csv` in a spreadsheet program, such as Excel or Google Sheets. For each column in the data file, answer the following questions.

- What is the name of the column header? For example, the first one is “id”.
- Write one sentence explanation of what information is in this column.

- Give a few examples of the values in the column, and what they mean. For example, some examples of values in the first column are 22, 325, and 140.¹

Question 3. In order to compare the representation of black people among the subjects of fatal police shootings to their representation in the general population, we need to know the percent of the United States population which is black. You can find this by googling racial demographics of the United States. Write the answer in the template file, and include a citation.

Task 2: Reading the data into a dictionary

The next step for this project is to read the data from the file `fatal-police-shootings-data.csv` using Python, and organize the data into a Python dictionary. This step has been done for you, and is posted on the course webpage as the file `police_shootings.py`. Download this file, and the `.csv` file containing the data, and save them in the same folder on your computer.

For this task, you don't need to write any code yet. Instead, you will add comments explaining the code that has been provided for you, and answer questions about the structure of the dictionary stored under the variable name `database`. In addition to reading the code, it can be helpful to run the module, and then use `print` statements to investigate the structure of the dictionary.

Add comments to the code in the file `police_shootings.py`, and answer the following questions in `project8_template.txt`:

Question 1. What information from the file `fatal-police-shootings-data.csv` is stored in `database`? That is, which columns from the `.csv` file are used? Refer to the columns by name, not by number.

Question 2. What information from the file `fatal-police-shootings-data.csv` do we use as the keys in `database`?

Question 3. What is the type of the values in `database`? Hint: you can find this using the line `type(database[key])`, where you replace `key` with an example of a key in `database`.

Question 4. Looking at one of the values in `database`, what information does it store? Give a specific example.

Task 3: Using the database

In this task, you will use Python and the dictionary `database` which has been created to ask and answer questions about the data. To do this, you will write additional code in the file `police_shootings.py`, continuing where the provided code ended, writing code to complete the following tasks:

¹Note that this dataset allows only a small number of categories for race, and the dataset fails to recognize that some people are multiracial. This is a weakness of the dataset.

- Print the name of the subject of the fatal police shooting with ID number 1694. If you've followed current events in the past few years, this should be a familiar name.
- Print the name of all subjects of fatal police shootings in Minnesota in the dataset. This will require iteration using a `for` loop.
- Create a new dictionary, called `race_counts`, which is initialized as an empty dictionary, and built by iterating through `database`, and following an accumulator pattern. The purpose of this dictionary will be to count the number of occurrences of each race among subjects of fatal police shootings. The keys in this dictionary should be races, and the corresponding values should be the number of subjects of that race.
- Print the fraction of fatal police shootings with a black subject. This should be a number between 0 and 1, and can be computed by dividing the number of fatal police shootings with a black subject (you can get this from the dictionary `race_counts`), by the total number of fatal police shootings (how can you get this without hard coding?).

In the template file `template.txt`, answer the following question:

Question 1. How does the proportion of black subjects in fatal police shootings compare to the proportion of black people in the United States population?

Next, you'll restrict your focus to the fatal police shootings where the subject of the shooting is unarmed, and find the proportion of these shootings with a black subject. To do this, write code in the file `police_shootings.py` to complete the following tasks:

- Create a new dictionary called `unarmed_selection`, which is initialized as an empty dictionary, and built by iterating through `database`, and following an accumulator pattern. This dictionary should have the same structure as `database`, except it will only contain entries for fatal police shootings where the subject was unarmed.
- Create a new dictionary, called `unarmed_race_counts`, which is initialized as an empty dictionary, and built by iterating through `unarmed_selection`, and following an accumulator pattern. The purpose of this dictionary will be to count the number of occurrences of each race among subjects of fatal police shootings, including only those where the subject is unarmed. The keys in this dictionary should be races, and the corresponding values should be the number of subjects of that race.
- Print the fraction of unarmed fatal police shootings with a black subject. This should be a number between 0 and 1, and can be computed by dividing the number of unarmed fatal police shootings with a black subject (you can get this from the dictionary `unarmed_race_counts`), by the total number of unarmed fatal police shootings (how can you get this without hard coding?).

In the template file `template.txt`, answer the following question:

Question 2. How does the proportion of black subjects in fatal police shootings where the subject is unarmed compare to the proportion of black people in the United States population? How does it compare to the proportion of black subjects in all police shootings?

Task 4: Reflection

Question 1. For your final task of this project, write a reflection (at least five sentences) on what you learned from this project, in the file `template.txt`. This can include your reaction to the results of the project, as well as the process of working with the data. This reflection will be graded for completion, not on the content of what you write.

Submitting your work

For this project, you will need to submit the following file:

- `police_shootings.py`, which is a Python file containing the program you wrote.
- `template.txt`, which is a text file containing your answers to all of the questions for this project.

Grading

You will earn one point for each of the following accomplishments:

- You have given a correct answer to Question 1 of Task 1, in the file `template.txt`, summarizing the data and its source.
- (2 points) You have given a correct answers to Question 2 of Task 1, in the file `template.txt`, describing the columns in the dataset.
- You have given a correct answer to Question 3 of Task 1, in the file `template.txt`, finding the percent of people in the United States who are black.
- As part of Task 2, you have added helpful comments to `police_shootings.py`, making it easier to understand the code that creates the dictionary database.
- You have given a correct answer to Question 1 of Task 2, in the file `template.txt`, correctly identifying the information from the dataset that is stored in our database.
- You have given a correct answer to Question 2 of Task 2, in the file `template.txt`, correctly identifying the keys in the dictionary.
- You have given a correct answer to Question 3 of Task 2, in the file `template.txt`, correctly identifying the type of the values in the dictionary.
- You have given a correct answer to Question 4 of Task 2, in the file `template.txt`, correctly describing the structure of the values of the dictionary.
- In `police_shootings.py`, you have correctly printed the name of the subject of the fatal police shooting with ID number 1694.
- In `police_shootings.py`, you have correctly printed the names of the subjects of all fatal police shootings in Minnesota from the dataset.

- In `police_shootings.py`, you have correctly constructed the dictionary `race_counts`, as described.
- In `police_shootings.py`, you correctly print the proportion of fatal police shootings with a black subject.
- You have given a correct answer to Question 1 of Task 3, in the file `template.txt`, explaining how the proportion of fatal police shootings with a black subject compares to the representation of black people in the United States population.
- In `police_shootings.py`, you have correctly constructed the dictionary `unarmed_selection`, as described.
- In `police_shootings.py`, you have correctly constructed the dictionary `unarmed_race_counts`, as described.
- In `police_shootings.py`, you correctly print the proportion of unarmed fatal police shootings with a black subject.
- You have given a correct answer to Question 2 of Task 3, in the file `template.txt`, explaining how the proportion of unarmed fatal police shootings with a black subject compares to the representation of black people in the United States population, and with the proportion of fatal police shootings with a black subject.
- (2 points) You have written a reflection on this project in the file `template.txt`.