



# Final Project

Assigned: MM/DD/YYYY

Due: MM/DD/YYYY

-----

**Submission Instructions:** Submit this project to its Canvas submission slot by the due time. You will submit your code, a short document, and your presentation. These can be submitted separately or as a zip file. Your code should be submitted as a file exported from IDLE, with a .py file extension. **One person** from each group should upload the submission to Canvas; this group submission functionality is supported in Canvas. Be sure to add comments to your code to support your work.

-----

*Learning Goal: Implement a larger-scale solution to a problem which integrates your knowledge from the other units in the course. Your solution should make use of file handling, conditionals, loops, string manipulation, dictionaries, and lists. This should also serve as a fun way for you to integrate your own interests (by way of the file and data you analyze) with your new Python skills.*

## Group Details:

For your final projects, you will work in groups of 3 (approximately) to create a codebase that takes in a file of your choosing and analyzes it. Rank your preferred data to use for this project using this form by Wednesday October 26 at midnight:

<https://forms.gle/KEPDHxzSxXc13kj87>

I will automatically assign groups based on your form responses. If you have preferences about your group members, such as if there is someone you would like to work with, please email the Instructor, CCing all students involved. All students must agree to work together before the email is sent to the Instructor. There will be 13 groups of 3 and 1 group of 2.

## Project Assignment:

In this project, you will automatically process/clean the data you select / are assigned. To clean the data, you must sort it into variables and decide what to do about any missing data or outliers.

Then, with this clean data, every team member must write a function to analyze the data. Some examples of potential findings might to create a lists of all the values in a column, for which you could then take the mean or max value if it is integers (numerical data), or find the most common value if it is strings (categorical data). You could also have an interactive function which asks the user for input about a specific value, e.g. how many times does X value occur in the data, and your function could output the number associated with value X. Finally, you could make a data visualization such as a line graph which shows results over time.

In effect, there are four components to this project: taking in the file and sorting/cleaning the data, then producing 3 analyses of this data.

For example, if you are using data about temperatures in US cities from 1950-2000: first, you will need to take in the file and store it using variables. Then, using the data, you can (1)

produce a line graph with a trend line across years, (2) create an input through which the user asks for the current temperature in a specific city, month, and year, and the corresponding temperature is output, and (3) print out the average temperatures for each city.

### **Code Requirements:**

Your code will need to implement the appropriate variable types and definitions, including dictionaries and lists where appropriate. Your project must meet the following requirements:

- 1) The code must execute without any errors, and must have been tested to ensure correct output.
- 2) You must correctly take in the file and store the relevant information from the file into variables / data types e.g. lists and strings.
- 3) You must have at least 3 functions or methods.
- 4) Your code must present 3 findings / analyses of your data, such as mean, mode, probabilities, max values, data visualizations, or more. A function which takes in user input and provides a response related to your data based on that input can also count as one of your 3 findings / analyses. The code corresponding to these analyses needs to be housed inside of functions.
- 5) Somewhere in your code you must use an if statement (at least one if statement).
- 6) Somewhere in your code you must use a for loop or while loop (at least one loop).

### **Project Presentation:**

The project presentation will be 8 minutes during class for each group. You can decide how to divide the time among team members; all group members need to speak for some amount of time. During the presentation you must describe: the data you used, your general approach to cleaning and analyzing the data, your code structure, and your results/findings. You will submit your presentation (as a PDF, PPTX, or Keynote) on Canvas.

### **Project Document:**

Along with your presentation and code, you will need to submit a short (1- to 2-page document, as a PDF) which outlines the following information:

- 1) A description of your data
- 2) Your approach to programming and analysis
- 3) How you tested your code
- 4) Your code structure
- 5) A summary of your analysis and findings
- 6) Which team members completed what work
- 7) An ethics statement about your data
- 8) Your takeaways from the project and findings, and any general feelings about the project or goals for future work.

### **Ethics:**

This project involves making use of Open Source data that is available online. Part of your final project will involve reflecting on the ethics of Open Source. On November 9, we'll have a class session dedicated to thinking critically about the sorts of benefits and harms that might accrue to people in virtue of what is made visible, and invisible, by any data set. The aim of this session will be to empower you all to think critically, and carefully, about the ethical dimensions of your own project analysis, and Open data more generally. You will incorporate some of your knowledge from the November 9 session into your final document. The details of this portion will be provided on November 9.

### **Important Dates:**

- Monday October 24— Project Assigned In Class

- Wednesday October 26 by midnight—All Forms Submitted and Emails Sent to Request Specific Group Members
- Friday October 28— Groups Assigned via Email (Topics and Members)
- Monday November 7 — In-class Project Meetings to Discuss Plan and Progress
- Wednesday November 9 — In-class Ethics & Data Activity (Mandatory Attendance)
- Wednesday November 16 & Monday November 21 — In-class Project Presentations
- Monday November 21 at midnight— Project Documents, Presentations, and Code Due (Submitted on Canvas)

**Grading Rubric:**

Aspect	Portion of Grade
Approach / Concept	20%
Project Code (Implementation and Structure)	50%
Comments to Explain your Code	10%
Project Presentation	10%
Project Document	10%