

Guidelines for Project 1

This document contains guidelines, requirements, and suggestions for Project 1.

Team Effort

Before anything, remember that Projects are a **group effort**: Working closely with your teammates is a requirement. This both helps teach real-world collaborative workflows, and enables you to tackle more difficult problems than you'd be able to working alone.

In other words, working in groups allows you to **work smart** and **dream big**. Take advantage of it!

Project Proposal

Before you start writing any code, your group should outline the scope and purpose of your project. This helps provide direction and prevent [scope creep](https://en.wikipedia.org/wiki/Scope_creep).

Write this as a brief summary of your interests and intent, including:

- * The kind of data you'd like to work with/field you're interested in (e.g., geodata, weather data, etc.)
- * The kinds of questions you'll be asking of that data
- * Possible source for such data

In other words, write down what kind of data you plan to work with, and what kinds of questions you'd like to ask of it. This constitutes your Project Proposal/Outline, and should look something like this:

> Our project is to uncover patterns in criminal activity around Los Angeles. We'll examine relationships between types of crime and location; crime rates and times of day; trends in crime rates over the course of the year; and related questions, as the data admits.

Finding Data

Once your group has written an outline, it's time to start hunting for data. You are free to use data from any source, but we recommend the following curated sources of high-quality data:

- * [data.world](<https://data.world/>)
- * [Kaggle](<https://www.kaggle.com/>)
- * [Data.gov](<https://www.data.gov>)
- * [Public APIs](<https://github.com/abhishekbanthia/Public-APIs>)
- * [Awesome-APIs List](<https://github.com/Kikobeats/awesome-api>)
- * [Medium APIs List](https://medium.com/@benjamin_libor/a-curated-collection-of-over-150-apis-to-build-great-products-fdcfa0f361bc)

Chances are you'll have to update your Project Outline as you explore the available data. **This is fine**—adjustments like this are part of the process! Just make sure everyone in the group is up-to-speed on the goals of the project as you make changes.

Make sure that your data is not too large for local analysis. **Big Data** datasets are difficult to manage locally, so consider a subset of that data or a different dataset altogether.

Data Cleanup & Analysis

With data in hand, it's time to tackle development and analysis. This is where the fun starts!

Inevitably, the analysis process can be broken into two broad phases: **Exploration & Cleanup** and **Analysis** proper.

As you've learned, you'll need to explore, clean, and reformat your data before you can begin to answer your research questions. We recommend keeping track of these exploration and cleanup steps in a dedicated Jupyter Notebook, both for organization's sake and to make it easier to present your work later.

Similarly, after you've massaged your data and are ready to start crunching numbers, you should keep track of your work in a Jupyter Notebook dedicated specifically to analysis.

During both phases, **don't forget to include plots**! Don't make the mistake of waiting to build figures until you're preparing your presentation. Creating them along the way can reveal insights and interesting trends in the data that you might not notice otherwise.

We recommend focusing your analysis on techniques such as aggregation, correlation, comparison, summary statistics, sentiment analysis, and time series analysis.

Finally, be sure that your projects meet the [technical requirements](TechnicalRequirements.md).

Presentation

After you've analyzed your data to your satisfaction, you'll put together a presentation to show off your work, explain your process, and discuss your conclusions.

This presentation will be delivered as a slideshow, and should give your classmates and instructional staff an overview of your work. PowerPoint, Keynote, and Google Slides are all acceptable for building slides.

As long as your slides meet the [presentation requirements](PresentationRequirements.md), you are free to structure the presentation however you wish, but students are often successful with the format laid out in the [presentation guidelines](PresentationGuidelines.md).

Presentation Requirements

The presentation requirements for Project 1 are as follows.

Your presentation must:

- * ☐ Be at least 8-10 min. long
- * ☐ Describe the core message or hypothesis for your project.
- * ☐ Describe the questions you and your group found interesting, and what motivated you to answer them
- * ☐ Summarize where and how you found the data you used to answer these questions
- * ☐ Describe the data exploration and cleanup process (accompanied by your Jupyter Notebook)
- * ☐ Describe the analysis process (accompanied by your Jupyter Notebook)
- * ☐ Summarize your conclusions. This should include a numerical summary (i.e., what data did your analysis yield), as well as visualizations of that summary (plots of the final analysis data)
- * ☐ Discuss the implications of your findings. This is where you get to have an open-ended discussion about what your findings "mean".
- * ☐ Tell a good story! Storytelling through data analysis is no different than in literature. Find your narrative and use your analysis and visualization skills to highlight conflict and resolution in your data.

- - -

Presentation Guidelines

You are free to structure your presentations to your liking, but students tend to have success with the following format.

- * Title Slide

- * Include the name of the Project and Group Members

- * Motivation & Summary Slide

- * Define the core message or hypothesis of your project.
 - * Describe the questions you asked, and _why_ you asked them
 - * Describe whether you were able to answer these questions to your satisfaction, and briefly summarize your findings

- * Questions & Data

- * Elaborate on the questions you asked, describing what kinds of data you needed to answer them, and where you found it

- * Data Cleanup & Exploration

- * Describe the exploration and cleanup process
 - * Discuss insights you had while exploring the data that you didn't anticipate
 - * Discuss any problems that arose after exploring the data, and how you resolved them
 - * Present and discuss interesting figures developed during exploration, ideally with the help of Jupyter Notebook

- * Data Analysis

- * Discuss the steps you took to analyze the data and answer each question you asked in your proposal
 - * Present and discuss interesting figures developed during analysis, ideally with the help of Jupyter Notebook

- * Discussion

- * Discuss your findings. Did you find what you expected to find? If not, why not? What inferences or general conclusions can you draw from your analysis?

- * Post Mortem

- * Discuss any difficulties that arose, and how you dealt with them
 - * Discuss any additional questions that came up, but which you didn't have time to answer: What would you research next, if you had two more weeks?

- * Questions

- * Open-floor Q&A with the audience

- - -

Technical Requirements

The technical requirements for Project 1 are as follows.

- * [] Use Pandas to clean and format your data set(s)
- * [] Create a Jupyter Notebook describing the ****data exploration and cleanup**** process
- * [] Create a Jupyter Notebook illustrating the ****final data analysis****
- * [] Use Matplotlib to create a total of 6-8 visualizations of your data (ideally, at least 2 per "question" you ask of your data)
- * [] Save PNG images of your visualizations to distribute to the class and instructional team, and for inclusion in your presentation
- * [] Optionally, use at least one API, if you can find an API with data pertinent to your primary research questions
- * [] Create a write-up summarizing your major findings. This should include a heading for each "question" you asked of your data, and under each heading, a short description of what you found and any relevant plots.

- - -