

Project Description

Due on July 13th, 2023

Introduction to Programming with Python

Columbia University Pre-College Program—Summer Session I

Instructor: Daniel Kadyrov

In this project, you will apply your Python skills to a real-world data analysis problem.

First, you will find a publicly available dataset that interests you. You can find datasets on websites such as Kaggle, Google Dataset Search, or Data.gov.

Using the dataset, you will perform the following tasks:

- Clean up the data and prepare it for analysis.
- Explore the data and generate descriptive statistics.
- Analyze the data and test hypotheses using appropriate methods.
- Visualize the data and communicate your findings using at least one graph.
- Narrate your work and explain your reasoning and conclusions.

You will use pandas, a popular Python library for data manipulation and analysis, to perform most of the tasks. You will also use other libraries, such as matplotlib, seaborn, or plotly, to generate your graphs. You will document your work using LaTeX or markdown, two common formats for online publication. You will also create a slideshow to present your project to the class.

You will upload your project deliverable (code, writeup, and slideshow) to Github, a platform for hosting and sharing code. You will also submit a link to your Github repository and your online writeup (using a service such as Overleaf or Github Pages) to the instructor.

This project will help you develop your Python skills, as well as your data analysis and communication skills. It will also showcase your ability to work with real data and produce professional results.