

MSCS3045 — Fall 2022

Final Project

As it says in the syllabus, 40% from your grade is assigned to projects (Labs). From this percentage, 10 % will be assigned to the final project. For your final project, you will work on something related to data analysis in Python. You can select any area (like sports, medicine, weather, games...) and try to solve some of its related problems by using Python. Your objective is to do data analysis activities which you already learned before from all labs. Use this project as a chance to gain more knowledge about tools, functions, and/or techniques we covered in this class so far, or to learn some new tools or techniques that we did not get to discuss.

Some possible project ideas include:

1. Select any real dataset and do Linear Regression activities on this file to extract an interesting information.
2. Do Cleaning and Visualization activities on your data file, your objective is to show an important information which are hiding in these data

You are not allowed to use any private data and you must cite your references.

Resources (very incomplete)

1. Any related articles
2. Chapters in the textbook that we did not cover.

Some places to find data

1. 10 Open Data Sources You Need to Know: <https://dzone.com/articles/10-open-data-sources-you-wish-you-knew>
2. Kaggle: <https://www.kaggle.com/datasets>
3. GroupLens Research: <http://grouplens.org/datasets/movielens/>
4. U.S. Census Bureau: <http://2010.census.gov/2010census/data/>
5. Centers for Disease Control and Prevention: <http://www.cdc.gov/datastatistics/>
6. National Climate Data Center: <http://www.ncdc.noaa.gov/>
7. Massachusetts Open Data Initiative: <http://www.mass.gov/portal/governmenttaxes/open/opendata.html>
8. Worcester State University Institutional Research: <http://worcester.edu/IR/default.aspx>
9. Public Data Sets on Amazon Web Services: <https://aws.amazon.com/publicdatasets/>
10. New York City Open Data: <https://nycopendata.socrata.com/>

Final Project Requirement

1. Project Proposal

You must submit a project proposal by:

1. Proposal: Nov. 15, midnight.

Your project proposal must give a brief description of what you plan to do, including possible tools and data sets (if applicable), the size of the group, and your justification for it being a group project.

11. Project Reports

You need to must submit 1 report about what you are doing in your projects before the presentation time. Each student must submit an individual report (in case you are working in a group). The report length is one page minimum. In the report, the student must explain what she/he did in her/his project each week.

-Report: **Due to Nov. 21, midnight.**

3. Final Project Deliverables

- Your project description, as a pdf file, the content of this file might include the answers of the following questions:

-The Content of your project pdf file

- Why you chose this project?
 - What does your project do/how does it work?
 - What tools did you use?
 - What did you have to learn about them?
 - If you are doing analysis of a dataset:
 - How did you select the data?
 - Where did the data come from?
 - What is the format of the data?
 - Show an example of your data (just a few lines) and explain what the fields represent.
 - What are your results?
 - What is the most important thing you learned?
 - What would you do differently?
 - What other things would you have liked to have done with this project?
-
- Your code.
 - Use either MLA or APA formatting styles in your summary. Formatting styles provide a concrete structure and format for written material that is beneficial for both readers and writers and allows people to quickly locate the sources referenced in your paper. The two most popular academic writing styles that you may run across are MLA and APA.
 - Final submission: **December 12th, 2022 - 11:59pm at the latest.**

4. Final project grading policy

- Project-Proposal: 20%
- Project -Report: 25%
- — Project-result, tool, final-submission: 55%



Copyright © 2022 Nada Alsallami. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.