Introduction to Data Science

Course Project Details

Onur Varol

@onurvarol
www.onurvarol.com
www.varollab.com

Sabancı Üniversitesi

VRLLab

Why individualized project?

- You can develop a portfolio of data science projects
    - Much more interesting in job / internship / graduate school applications
        - Time spent on the project will be useful to you
- You can learn about yourself and gain insights from your own data
- It’s much more fun for everyone to see these projects

Projects' "must-have" components

- Project will be done using **YOUR OWN DATA**
- Diverse set of techniques: EDA, visualization, machine learning models
- Github page for the project material
    - You don't need to share raw data, but analysis scripts needs to be there (Hint: search for .gitignore)
    - Website / short video / README file / report
- Participation to peer-evaluation

What to present

- Motivation: Why you are working on this project?
- Data source: Where did you get this data? How did you collect it?
- Data analysis: Techniques used, different stages of the analysis
- Findings: What you learned about yourself?
- Limitations and future work: What could be done better? Do you have any future plans about your project?

Evaluation

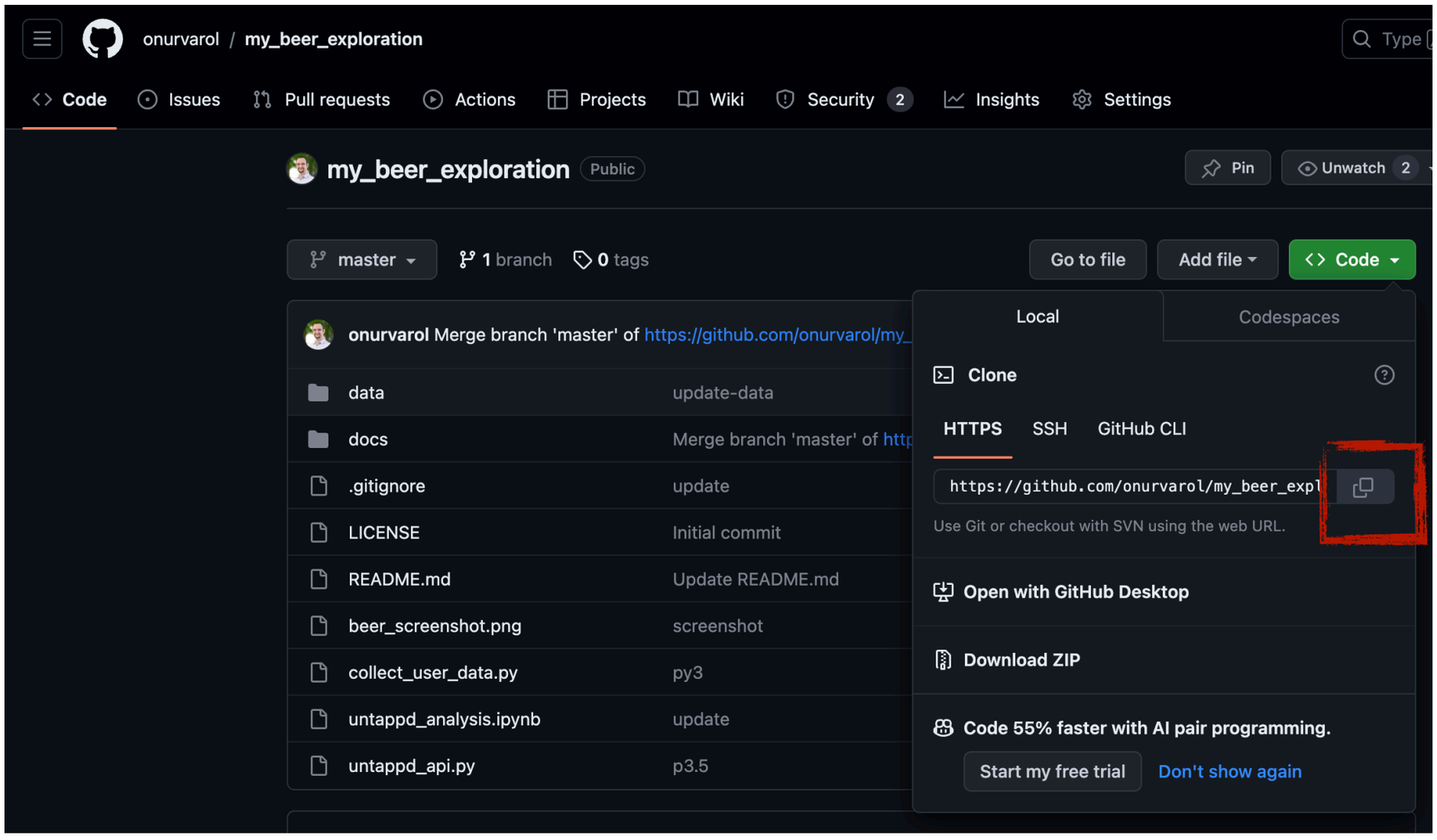
- Participation to project proposal (10%)
    - We will collect Github Repo URLs with an assignment
- Code and material review (20%)
- Quality of the project presentation (30%)
- Peer-evaluation
    - Grading other projects (10%)
    - Grades received from peers (30%)

What you need to do by Nov 30th

- Prepare a **public** Github repository for your project.
    - **DO NOT** change the name of the project, because we will grade what we will find in it at the end of the semester.
    - Create a README.md file and describe your dataset, project idea and plan. This will be the material we look for this assignment.
    - You can turn your repo private between 05/12/24 to 10/01/25, but DO NOT forget to make it again public.
    - In the meantime, make sure to **update repo regularly**, so that we can monitor your activity.
- Submit the URL of the repo to SuCourse Assignment

What is the submission format

- We will **only** ask for a **public** Github repository URL.



- We will automate cloning Github repositories with a script. <u>Mistakes in the URL format will cause you points!</u>

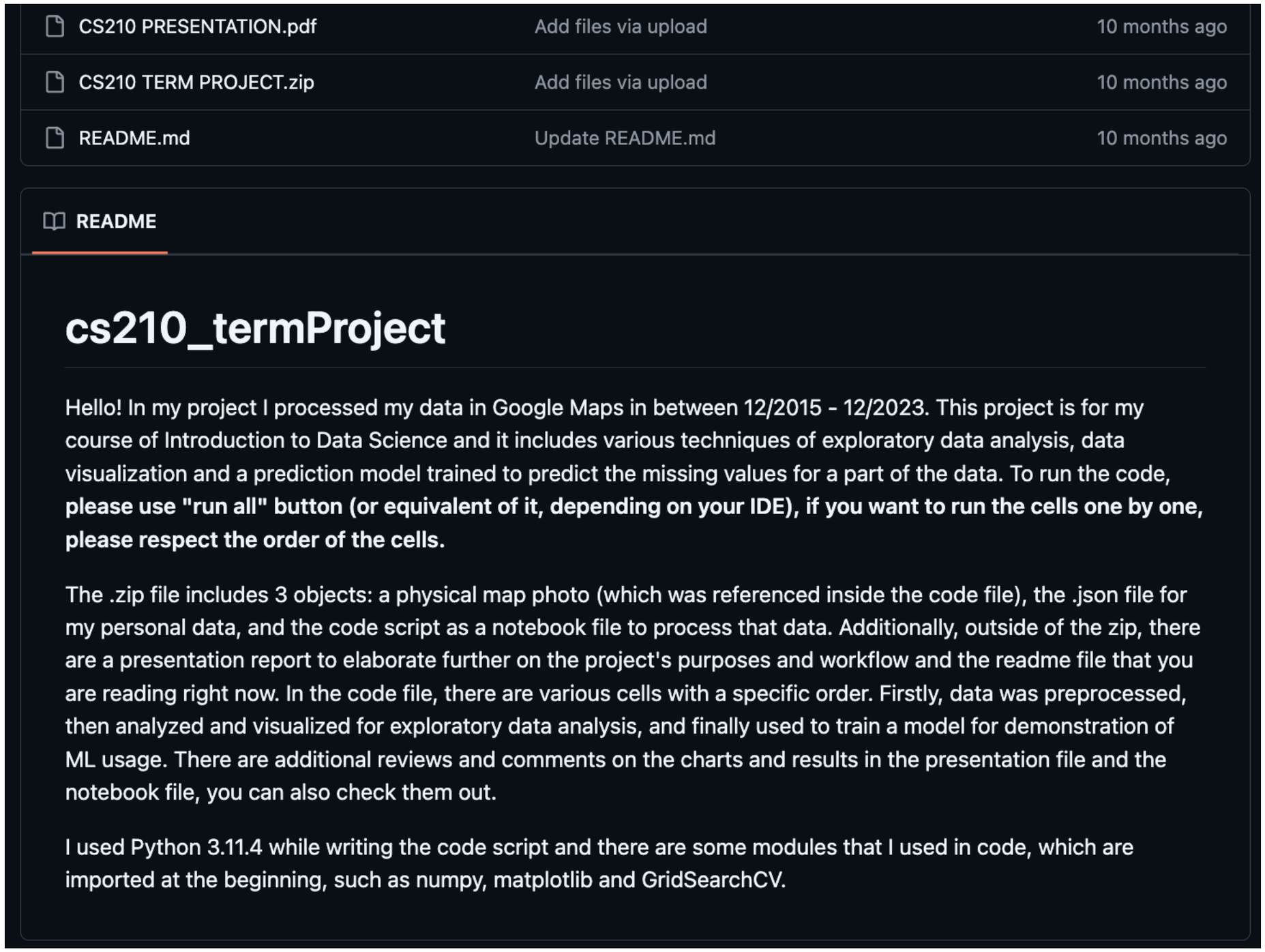
Sample projects

- Reddit usage analysis: https://github.com/firatbatar/reddit-usage-analysis

- Spotify and health dashboard: https://github.com/egeyardimci/CS-210Proj

- Lichess game analysis: https://github.com/cerendurucinar/CS210-Fall-2023-Project

Don’t do this!


CS210 PRESENTATION.pdf
Add files via upload
10 months ago
CS210 TERM PROJECT.zip
Add files via upload
10 months ago
README.md
Update README.md
10 months ago
README
cs210_termProject
Hello! In my project I processed my data in Google Maps in between 12/2015 - 12/2023. This project is for my course of Introduction to Data Science and it includes various techniques of exploratory data analysis, data visualization and a prediction model trained to predict the missing values for a part of the data. To run the code, please use "run all" button (or equivalent of it, depending on your IDE), if you want to run the cells one by one, please respect the order of the cells.
The .zip file includes 3 objects: a physical map photo (which was referenced inside the code file), the .json file for my personal data, and the code script as a notebook file to process that data. Additionally, outside of the zip, there are a presentation report to elaborate further on the project's purposes and workflow and the readme file that you are reading right now. In the code file, there are various cells with a specific order. Firstly, data was preprocessed, then analyzed and visualized for exploratory data analysis, and finally used to train a model for demonstration of ML usage. There are additional reviews and comments on the charts and results in the presentation file and the notebook file, you can also check them out.
I used Python 3.11.4 while writing the code script and there are some modules that I used in code, which are imported at the beginning, such as numpy, matplotlib and GridSearchCV.

Deadlines

- Since there are several hard deadline for grade submission, there will not be any late submission. **PLEASE PLAN ACCORDINGLY!**
- **Deadline for submitting the URL with README.md file: Nov 30th**
- **Deadline for submitting the project material: Jan 10th**
- **Deadline for peer-evaluations: Jan 12th**
- **Last day for us to submit grades: Jan 16th**