

Data Science Individual Project

Objective

Provide a data-driven solution to a problem that excites you using the tools discussed (or related to) in this course.

Only visualization is not enough for this project

Datasets

The dataset can be same or different with your group project.

But you need to answer or provide additional solution to **three questions**.

Tasks

For your project, you should:

- Pick a topic/issue and three questions that excites you
- Select or create datasets
- Familiarize yourself with that data, if necessary:
 - data munging
 - feature engineering
- Choose **proper model**, **Train/Fit the model**: NumPy, Pandas, Matplotlib, SciKit, OpenCV, Keras, TensorFlow (tf), catboost, DeepLearning4j, etc.
- Analyze results

Deliverables

Deliverables for your project:

- Draft of report:
 - Required: draft of introduction, datasets, methodology.
 - Optional: draft of result, discussion

Due 12/3 (Sunday), at 11:59 pm

More details about submission will be released before the due

- Final report: introduction, datasets, methodology, results, discussion/suggestion, reference + codes

Due 12/10 (Sunday) at 11:59 pm

- Format
 - Create a repository in your Github account. If the repo is not a public one, add 'pangwit' as a collaborators by the instruction in [link \(https://docs.github.com/en/account-and-profile/setting-up-and-](https://docs.github.com/en/account-and-profile/setting-up-and-)

[managing-your-personal-account-on-github/managing-access-to-your-personal-repositories/inviting-collaborators-to-a-personal-repository](#))

- Published the report as readme in the repo
- Add all pictures in a folder graph under the content
- Add all datasets in a folder data under the content
- Add all codes in a folder codes (.ipynb files) under the content
 - make sure the codes can be run without any change (the codes about read data should under content data/...)
 - Write markdown cell and comments
- Revise the readme.md to generate the report
 - Include all required categories in Rubric
 - Include link to main .ipynb file
 - Include reference
- The link of the repo will be submitted as assignment in BrightSpace

More details about adjust readme file can be found during our lab at 12/1, or in this [reference](https://docs.github.com/en/repositories/managing-your-repositorys-settings-and-features/customizing-your-repository/about-readmes) (<https://docs.github.com/en/repositories/managing-your-repositorys-settings-and-features/customizing-your-repository/about-readmes>).

Rubric

Category	Explanation	%
Introduction	Why was the project undertaken? What was the research question, the tested hypothesis or the purpose of the research?	10
Selection of Data	What is the source of the dataset? Characteristics of data? Any munging, imputation, or feature engineering?	20
Methods	What materials/tools were used in answering the research question?	20
Results	What answer was found to the research question; what did the study find? Any visualizations?	20
Discussion	What might the answer imply and why does it matter? How does it fit in with what other researchers have found? What are the perspectives for future research?	20
Coding	ipynb file with clear comments and datafile.	10

Rubric based on the IMRAD: <https://en.wikipedia.org/wiki/IMRAD> (<https://en.wikipedia.org/wiki/IMRAD>).

Sample

https://github.com/pangwit/DS_Individual_Project_Example/
(https://github.com/pangwit/DS_Individual_Project_Example/).

In []: