# Capstone Project

Due date: 2025-1-5

Your task is to demonstrate the data science skills you learned in the class using a data set of your choice. The evaluation of the capstone project makes up 50% of the final score. You will work in a team made up of 3 people, present in the class, and submit a written report (along with data and code) on https://learn.tsinghua.edu.cn. The presentation and the written report count toward the points for the capstone project (20 pt and 30 pt, respectively). There are no mandatory subjects for the analysis. However, please note the following:

- Data collection and cleaning are important topics in data science. When choosing the data to analyze, collecting data from multiple sources is highly encouraged, especially via web scraping. Dirty and unstructured data with missing values are welcome. Well-formatted data in a table that others have prepared for you is not encouraged.
- Exploratory data analysis is another important topic in data science. Data visualization can be helpful here. The goal of the exploratory analysis is to understand the overall distribution and relation of the variables, to provide initial insights and hypotheses, and to uncover or design important features for prediction.
- Data visualization can be an important part of the project, but data visualization is not recommended to exceed 50% of the work.
- Similarly, it is not recommended that machine learning makes up more than 50% of the work.
- The style of the report should resemble a short paper, with the following sections: introduction (background, significance, and literature review), methods (for data collection, cleaning, and analyses), results, discussion, and conclusion.
- The project will be evaluated based on the data science skills that you demonstrate in the project, the amount of effort you put into it, innovativeness, the logic of the analyses, clarity and detailedness of the writing, etc. Summarize the contribution of each author at the end of the report.
- If your dataset is too large, upload it to https://cloud.tsinghua.edu.cn and provide a download link in your report.
- There is no extension for the deadline. Late submissions are not accepted and will not have a score.