

SDSC5001 Project

1. **Team Formation Submission (1 point):** Each team should have four or five students. Please form your team and submit your team name and team members by 5 Oct (Sun) @ 11:59 PM. The submission link will be announced later. Your submitted file should be named as “Team-Formation-XX”, where “XX” is your team name.
2. In the project, you should identify and model a valuable problem which could be solved by statistical/machine learning methods. You need to collect data, propose good models, run learning algorithms, validate/assess your models, interpret the results and draw conclusions in the format of a written report.

You are also encouraged to participate the real project in any applied fields; but **it is not permitted to directly use your ongoing project in other courses or your thesis work**. Your project must be a new and independent one *exclusively* for this course.
3. There is no restriction or preference on the topic, the methods, the algorithms, the programming language (SAS, R, Python, etc.) used in your project. The scientific value of your project is the main point for evaluation.
4. **Project Proposal Submission (4 points):** Your team is required to propose a topic of interest and submit a proposal (at least half a page in PDF file with team name) to ([Canvas->Assignment: Project Proposal](#)) by 2 Nov (Sun) @ 11:59 PM. The file name should be “Team-Proposal-XX”. Each team only needs to submit one copy of the project proposal.
5. Written report requirements: the submission consists of three parts:
 - **Poster:** a single powerpoint slide in A0 size (1189 × 841 mm)
 - **Main report:** a well-organized PDF document within **6 pages** following smooth logic such as Cover page (course name, title of project, student names and UIDs, date), Contributions of team members, Background, Data sources, Methods and algorithms, Results, Interpretation, Conclusion and discussion
 - **Appendix:** source codes in PDF file
6. **Report Submission (15 points):** You should submit your poster with name “Team-Poster-XX” to [Canvas->Assignment: Poster](#), submit your main report with name “Team-Main Report-XX” to [Canvas->Assignment: Main Report](#), and submit your appendix with name “Team-Appendix-XX” to [Canvas->Assignment: Appendix](#). Each team only needs to submit one copy of their work. The submission deadline is 30 Nov (Sun) @ 11:59 PM.
7. **Grading Policy:** The total points for project is 20. Team formation counts 1 point, proposal counts 4 points, and written report counts 15 points. No source codes results in zero point of the report. Plagiarism is strictly prohibited.

Appendix:

The following is a collection of available datasets you may use in your project. However, you are free to use any other dataset that is not listed here.

<https://www.kaggle.com/datasets>

<https://paperswithcode.com/datasets>

<https://archive.ics.uci.edu/datasets>

<https://data.world/>

<https://data.gov.hk/>

<https://www.kdnuggets.com/datasets/index.html>

https://en.wikipedia.org/wiki/List_of_datasets_for_machine_learning_research

You can easily find more from popular data science forums or blogs such as <https://medium.com/topic/data-science>, <https://www.jiqizhixin.com/> (Chinese), <https://towardsdatascience.com/>, <https://www.quora.com/Where-can-I-find-data-sets-for-regression>, <https://blog.bigml.com/list-of-public-data-sources-fit-for-machine-learning/>.

***** End *****