# CS148 Course Project Guideline

1. Objective

The objective of this course project is to get you hands-on experience in applying what you have learned in class (or beyond) to solve a real-world data science application.

1. It is a team-based project, and everyone in the same team is expected to have the same score, unless you are reported as a free-rider in the peer evaluation.

* 4-5 people for each team.

1. Topics
   * Please discuss with your TAs for the right level of data science problem. Make sure it can be solved within 3 weeks.
   * Sources: Kaggle, twitter, GitHub, other lists
     1. <https://www.kaggle.com/>
     2. <https://gist.github.com/entaroadun/1653794>
     3. Examples from past years: 1: heart disease prediction; 2: song popularity prediction; 3: card driver fatality prediction; 4: Netflix show rating prediction; 5: Hepatitis C prediction; 6: Strokes prediction; 7: movie recommendation system; 8: Apex legend: multiplayer battle royale game; 9: student performance prediction. 10: Predicting the outcome of a multi-player game based on the players’ past records 11. Classifying some objects e.g. planet, food, trash 12. Predicting prices / scores e.g. car, stock, ranking of music, etc. 13. Certain medical treatments’ success rate 14. Predict social relationship 15. Fake news prediction
2. Requirements

The general procedure follows the typical data science life cycle:

* Choose topic (decide your problem)
* Collect data
* Apply algorithms to the data (note you might need to generate sub tasks, and each sub task belongs to a different function type e.g. algorithms used in preprocessing data? ensemble models made up of different components? — not required but could be)
* Evaluate and compare algorithms
* Present your project to the whole class on the last day (visualization, presentation, communication)

1. Evaluation criterion
   * Whether the topic is a data science topic, interesting, well-motivated?
   * Whether the project covers all the 5 components?
   * Whether the project has a clear problem definition on every component (especially where you use the algorithms learned in class) involved?
   * For the specific subtask, have you chosen the proper models and tried different models?
   * Have you done model evaluation and comparison properly?
   * Have you provided visualization and interpretation to your results?
   * Have you communicated well to your audience, in terms of presentation and report?
2. Timeline (except for presentation and slide, report, code submission, other timeline is soft)
   * 11/1: **Proposal Due**
   * 12/4: Presentation
   * 12/9: Report Due
3. What to submit
   * Slides, report, code to BruinLearn (Project leader submit everything on behalf of all the team members)
   * Peer evaluation form (submit by **everyone**)
4. What to expect on the last day
   * Group Presentation + 5(QA) minutes each group
   * Peer rating, everyone will rate every group (include yourself)
   * TA and instructor rating.