

# Data Mining Principles Class Project

---

The class project gives you hands on practice on data mining. The project will require you to develop and document a project from (any) data (preferably a dataset from an organization), and finally implement the project using Python.

## **The project will be completed in two stages:**

1. Project design. The team will collect and analyze a related implementation of an organization of your choice (make sure you have sufficient domain knowledge).
2. Data mining. The team will implement the project in Python and will conduct regular data mining tasks learned in class.

## **Team Policy**

- This is a team project.
- A team should not have more than **5** students.
- All team members must contribute to the success of the project.
- If you are not able to find a teammate, please email me and I will find a teammate or coordinate a team for you.
- **All team members should be responsible for the project progress.**

## **Schedule**

Date	Milestone	Points
Week-1	Project announcement is out.	
Week-5	M1: Team is formed, and a brief description of the project is due.	20 pts
Week-10	M2: Project Presentation and Final report	80 pts

## **Team Submission Policy:**

- Reports are submitted on Canvas under Class Project category.
- Each team **only** needs to submit one file for each submission. One team member will submit the report on behalf of the team.

## **Grades:**

- Total point of the class project is 100. That is a substantial part of your final grade.

### Grading

The evaluation is based on the following aspects.

- **Clarity.** Are the organization/business data strategy clearly explained?
- **Novelty.** How novel is your business/organization choice?
- **Completeness/Substance.** Does the model have enough substance, such as unstructured data, data mining algorithms, visualization
- **Correctness.** Is the project design sound and well-chosen?

Example projects:

Kaggle.com is a good start. To just give you an idea, some of the impressive ones I have seen so far after all machine learning, data mining, and deep learning training:

- Divvy Bikes Graph Analysis
- AirBnb Price Prediction
- Wholefoods Nutrition Prediction
- Salary Prediction by using Indeed and Glassdoor
- H-1 Visa Application Prediction
- Starbucks Coffee Classification
- Fashion Brand Detection
- Investment Decision Forecasting
- Covid-19 Disease Spread Analysis

You can narrow down the scope to the techniques that you learn in this class and provide a robust implementation.

M1 requirement: Team list, role of each member, case description (min 1 page single space excluding references)

M2 requirement: Final white paper report (Min 10 pages single space excluding references) and a jupyter notebook with predictions.csv (if applicable). Max 10 slides ppt presentation, professional outfit and ready to answer all related questions about the project.

PS: I will dedicate sufficient time for the projects require extensive research and implementation.

Good luck!