

MAS DSE 230 - Project Description

Overview: For the project, you will apply what you have learned from this course to analyze real-world data. The project consists of three main components: project proposal presentation, final project presentation, and peer review. Requirements for each project component are described below.

Teams: This project is to be completed in teams of 2 people. Please sign up for teams on the sign-up sheet that we will provide for you.

Important Note: Be sure that your team submit your own work on this project. If we find that you have simply copied an analytics solution from another team or from a website, you will receive a 0 for the project and will be reported to the Academic Integrity office. And we will check!

Grade Breakdown

The grading breakdown is as follows:

- Project: 35 points total
 - Project proposal presentation: 10 points
 - Final project presentation: 20 points
 - Peer review: 5 points
 - Evaluate project presentations of two other teams: 4 points
 - Evaluate team members: 1 point

Data

We will provide some datasets that you can use for your project. You can also find a dataset on your own to use.

Project Requirements

General Requirements

- Use PySpark or Dask
- Include one classification **or** regression **or** cluster analysis task
- Describe problem
 - To include: Explain why problem is interesting, what real-life application is being addressed
- Describe analysis task
 - To include: type of task (e.g., classification), how does task related to business problem
- Describe data
 - To include: data quality issues, characteristics of the dataset (summary statistics, correlation, outliers, etc.), plots
- Describe data preparation process
 - To include: data cleaning steps, features used, train/validation/test datasets
- Describe analysis approaches
 - To include: input, setup, and output of model(s)
- Describe challenges and solutions
 - To include: challenges encountered, solutions to address challenges

- Describe analysis results and insights gained
 - To include: discussion of results, insights gained from analysis
- Describe future work
 - To include: lessons learned, next steps, what you would have done differently

Project Proposal Presentation - 10 points

- In Session 3, you will present the problem you want to address, the dataset you want to use, analysis task you want to perform, insights you hope to gain, and potential challenges with the data and/or the task.

Final Project Presentations - 20 points

- In Finals Week, you will present the problem description, analysis task, data preparation process, analysis approaches, challenges and solutions, analysis results and insights gained, and future work.

Peer Review - 4 points

- Each team will evaluate the final presentation of two other teams. This involves asking questions and providing feedback during Q&A. We will assign teams for you to review.

Team Evaluation - 1 point

- This is required of each person on the team. Evaluate your partner based on their contribution, quality of work, teamsmanship, and promptness. We will provide a template.