

MATH3836 2024 Project Specification

March 2, 2024

1 the administration part

- **released date:** 02/03/2024
- **submission date:** 20/04/2023
- This is a group project of two students
- this is also an open-ended project, and markings are primarily focused on the effort and innovation of the project.
- this project is worth 35% of the entire subject mark

2 Introduction

2.1 Rationale

In MATH3836, you have studied a variety of quintessential techniques for data mining. Since data mining is one of the most practical skills in academic and business, therefore, this project will give you the opportunity to translate what you have learned into practical applications.

2.2 Setting

Imagine that you are an entrepreneur trying to develop an innovative data mining tool on which to base your startup business. To limit the scope, the datasets used in this project will be sourced from <https://data.gov.hk/en/>

This project mimics part of this process. Your tasks include, but not limited to:

1. Design a data product for specific application/discipline, such as transport or retail.
Here, your design must address some fundamental questions, including, who the users of this data product are, is your goal realistic, what makes this data product unique, and what impact this product may have on the society.
2. Download one (or a few) “open source” datasets from <https://data.gov.hk/en/> that can be used as the basis for your product
3. Apply the set of techniques you acquire in this course to the datasets you downloaded, in order to get closer to your product goals.
(Optionally) you can use a programming language or application programming interface (API) to help illustrate your point. Any programming language is permissible.
4. You also need to design any future steps that may be taken to bring the data product to life.

You need to submit all relevant materials relating to this project, including but limited to:

1. Report

A report should be detailing all the necessary steps you have taken to achieve the goal(s). Note that it is not necessary to include a lot of background knowledge to make up the length of the report, unless this consists mainly of your own contributions. Note that you can **not** re-submit the report after the due date. Reports have no word limit, but should document everything you have done. You may include any coding (if any) as attachment. Keep in mind that I'm looking for your efforts in this project, so you can also include some failed attempts in the report if it helps to further illustrate your efforts.

The submission will be on Moodle.

2. Presentation PowerPoint

You will need to submit a PowerPoint your presentation before the deadline along with the report.

Although changing the PPT after the deadline is not ideal. However, if you feel it is necessary, you can still change the content of PowerPoint after the deadline, and your new presentation should clearly show the changes you made.

3. Presentations time

Presentations are best kept under 20 minutes. Students should keep background information to a minimum and just focus on what you have done. You should allow 10-15 minutes for Q&A. The assessors will include the Professor and the teaching assistant.

The presentation times will be after the project submission date. The latest dates for presentations will be announced shortly. (Students are always welcome to specify a presentation time earlier than the last possible date)

3 Grading

Broadly speaking, your assignment will be graded in three aspects:

1. understanding of subject knowledge:

Students need to apply a broad range of skills acquired in the subject, i.e., apply a variety of data mining tasks to datasets and products. You have to use a wide range of techniques, not just limited to one or two.

2. uniqueness and innovations

Students are required to demonstrate deep consideration and innovation in the design and implementation of your data products. Why is this data project unique and innovative?

3. presentation

This project will involve a presentation component. Not only do students need to have a very detailed report, but actual presentation is also crucial for a high score in this subject.

4. how far away from the example scenario given

5. Although this is a group project, each student is individually graded for the quality of their responses to the questions. Note that answering *"I don't know the answer because other student has done that part"* is not acceptable.