Faculty of
Computer and Informatics
Department of
Artificial Intelligence and Data Engineering

YZV211 Introduction to Data Science and Engineering
Fall 2023
Project

You will propose and design a data science and engineering project. Your project evaluation will be performed in two stages. First you will be required to submit a proposal/design document.

The project topics are limited to the themes available in Turkish Statistical Institute which provides statistical data about various themes shown below.

You are expected to choose a project topic where you use the main data the above themes from the Turkish Statistical Institute (https://data.tuik.gov.tr). An example project topic may be "*The Effect of Depreciation of Turkish Lira on House Sales Both to Locals and Foreigners*". The project must benefit from at least two different sources of data, one of them must be from Turkish Statistical Institute. For example, if you choose to work on house sales statistics you may need to get additional data about average house prices[1].

You will be working as a team of 3 people at most.

[1] For example, a report on this may be accessed from https://image5.sahibinden.com/staticContent/cf40508c/1dff/4ed6/be39/ea3279454c44/1692261349.pdf, Please consider that there may be other data sources also.

Proposal document

(Should be submitted by 19th December 2023), at least 2 pages A4.

Project Team
Team members should be listed.

Project Title

Summary
One or two lines explaining the project.

Problem definition
One or two lines explaining the problem formally.

Dataset(s) Sources
Explaining the datasets of the project. The project must use at least two different sources of data. Please talk to us if you have a special reason to use a single data source.

Possible Research Questions
Describe what research questions can be addressed by this study in the capacity of the datasets.

Design and (partial) implementation document (Should be submitted by 3th January 2024)

Submit a pptx and ipynb file to present your work. Alternatively, you can use extensive markup and other plugins to create your presentation in the ipynb file. In that case the ipynb file must be formatted completely. Visual quality must be high. You can use plugins to assist in preparation of the notebook-presentation. However, we must be able to see and evaluate your presentation without any installations (in case submit a pdf version as well).

The presentation sections:
Project title, team
Problem definition
Same as in the proposal. But you may have changed it slightly after studying the data.
Datasets and sources
Explain datasets and sources, how they merged.
EDA
Data example, some visualizations from your data.
Research Questions and Proposed tests
Explain some simple research questions you have asked to the data.

1. Get at least one research question tested. Write a hypothesis. Answers including the statistical tests, tables, figures.
2. Write at least three more research questions. Write your hypothesis. we do not expect you to answer them all. Some of the questions may require machine learning methods. Explain them.

Potential products or how to convert this project to a service or product
Explain how this project can be converted to a product or a service.

Any issues related to data engineering.
Explain which data engineering issues needs to be addressed for a product/service. E.g. for continuous functioning.

Any issues related ethics.
Explain the ethical issues of the project or future studies. Do you need to consider getting permissions, privacy, ethics etc.

Conclusion
Few bullet points, what is accomplished in this project.

Presentation (Submitted by 3th January 2024)

5-minute oral presentation of your work. Presentation does not need to show all the analysis in the project.
Teams members must share the presentation speech equally.
This presentation must be uploaded to a streaming site such as Youtube. Please do not forget to submit this link and make sure that the link is accessible for at least one month.

Grading Scheme

	Weight
Proposal Document	3
Design and Implementation Doc Overall Quality	3
Presentation	2
Datasets	3
EDA	3
Research Questions-1 Implemented	3
Research Questions-2-3 Designed. (Listing test and ML methods)	2
Products or Service Idea Design	2
Data Eng. Issues	2
Ethics Issues	2
Total	**26**