IS5126
Final Group Project

- **All project deliverables (including presentation slides) are due on 7th November 2025 (Friday), 6:30pm**
- **Group in-person presentation on 7th or 14th November 2025, 6:30pm to 9:45pm.**

Your group will identify business problems, collect, and prepare datasets, perform analyses, address the plausible questions, produce professional report and present your work on the dates specified above. The project should highlight some novel insights and demonstrate a mature and sophisticated understanding and interpretation of the data. The group should use the proper technique with the correct interpretation. Students must demonstrate the ability to extract value out of the data given its constraints such as sample bias, missing data, methodological constraints, industry-specific knowledge, and practical constraints.

In addition, groups are encouraged to explore how Large Language Models (LLMs) and Generative AI workflows can be integrated into their projects. This may include, but is not limited to:

- using LLMs to assist in data cleaning, exploration, or feature generation,
- designing workflows that combine traditional analysis with LLM-based insights,
- applying generative AI APIs to create unique outputs (e.g., summaries, recommendations, scenario generation) that improve the approach or user experience.

Clarification on model performance

- This project will not be judged on raw model performance. Marks will not depend on metrics such as accuracy, AUC, RMSE, MAE. A high metric alone will not improve your grade, and a modest metric will not reduce it.

- Instead, your evaluation focuses on the clarity of the problem formulation, quality of the data work, appropriateness of methods, soundness of reasoning, and the value and credibility of insights and recommendations. Negative or null results are acceptable if they are well justified and thoughtfully analysed.

The project is designed to offer the opportunity for creativity by exploring unconventional ideas. The scope of the topic does not have to be extensive to be valuable; a well-defined "small" question with clear identification and relevant data often yields better insights. You have the freedom to choose your project's topic and can go beyond the course's scope. However, it is advisable to avoid topics with challenging hardware requirements, or those that are overly common, like stock, real estate, crypto, gold and property price prediction.

The details of the final group project deliverables are as follows:

- Present your work in-person on the dates specified above**.**
    - Each group is allowed 12 minutes for presentation and 5 minutes for comments and questions.
    - Every group member must speak during the presentation. A timer will ensure the presentation stay within the allotted time.

    - The final version of the presentation slides must be submitted with the other deliverables by the dates specified above.

- A single zip-file with filename 'GroupXX_FinalProject.zip', where `XX' is your two-digit group number, is to be submitted to Canvas assignment folder and contains the following:
  - *Project Report (PDF)*
    There is no page limit for the Project Report which should be kept succinct and concise.
      - As a guideline, your group may aim for 15 pages (double-space, font size 12). The report should include the details of the methods and results. Conclude from the project and discuss the value of the results, including actionable business recommendations and managerial insights.
  - *All dataset(s) in serialized format (e.g., CSV)*
  - *URLs / Addresses of API Endpoints and API keys*
  - *All source codes* (.ipynb) such that your results are reproducible from the raw dataset.
  - *Presentation Slides* (.pptx), the final version.

Rubrics

Your project will be assessed based on (as a guideline, try answering the questions):

1. **The novelty and value of your application *(20%)***
    - How is the focal question or problem rooted from current practice or belief?
    - Why is your solution superior to existing approaches?
    - What is the novelty of the data collected or the way you combine or segment the datasets?
    - What is the novelty of the way you applied the methods on the dataset?
    - Was any Generative AI / LLM-based workflow integrated to create unique outputs or insights that significantly improved the analysis or user experience?
2. **The attention to details on the data collection and exploration *(10%)***
    - How was the data collected? How was it sampled?
    - What method or technique was used for collecting, preparing, exploring, and visualising the data?
    - How was the data handled? Any issues with missing data and combining datasets?
    - How large is the dataset?
3. **The validity of your methods *(30%)***
    - To what extent did you consider alternative factors and hypotheses in your analysis?
    - How robust is your analysis? Can it maintain validity under varying conditions?
    - Model choice should be justified by the question and constraints, not by chasing marginal metric gains.

4. **The coherence of interpretation and conclusions *(20%)***
    - Could the estimators be interpreted in a meaningful way?
    - Reason to believe the validity of analysis results? Will this work on a new data sample? What was the scope of the analysis?
    - What kind of selection bias may exist?
    - What was the managerial insights and business value of the analysis?
5. **The quality of the presentation and report. *(20%)***
    - Was the question or problem well motivated and highlighted in a clear fashion?
    - Was the presentation concise and coherent on selected methods in a logical way?
    - Did the report highlight key findings and contributions?
    - Was the information and analysis presented in a clear manner, with an engaging narrative or surprising findings?
    - Was it easy to digest the results with proper visualization, tables, and explanations?
    - How well were references organized and connected to the analysis?
