

## SIG720 Task 5D

Execute your code into a jupyter notebook (.ipynb file) and keep the output, write a report (.pdf file) to answer the following questions, and submit your code and report.

In this task, the goal is to develop and evaluate various regression models to predict housing prices in Melbourne by:

- Collecting and preprocessing real estate data.
  - Building and comparing multiple regression models.
  - Evaluating model performance using appropriate metrics.
  - Discussing the applicability of the models in real-world scenarios.
1. Melbourne Housing Data acquisition: choose the top three suburbs that you want to live in Melbourne, leverage [realestate.com.au](https://realestate.com.au) to automatically or manually collect at least 150 housing data points (at least 50 for each suburb). For each housing data point, prepare as many features as you can, example features are property type, location, number of bedrooms/bathrooms, land size, sale date, etc. The target variable is sold price.
  2. Data preprocessing and exploratory data analysis:
    - a. Convert categorical variables (e.g. unit/house/apartment) using one-hot encoding.
    - b. Create new features (e.g., number of schools nearby).
    - c. Normalize or standardize numerical features.
    - d. Visualize housing price distributions, feature to target correlations, and outliers.
    - e. Identify price trends over time across suburbs.
  3. Model development: develop at least three different regression models, use MAE, RMSE and R-squared as the evaluation metrics, use k-fold cross validation to evaluate the model performance.
  4. Feature importance: identify which features more influence housing price, use model-specific methods (e.g., feature importance in tree-based methods) or other statistical methods (e.g., SHAP values).
  5. Model deployment: Develop a simple web demo application using appropriate packages (e.g., gradio, Flask or Streamlit). Allow users to input property features and receive price predictions.