# CST4050 - Weekly challenges

**Challenge  A**

> Students are required to solve their challenges at home, submit their solutions to MyLearning and MSTeams and present
> their solutions in class during the last two weeks.
>
> Successfully completing the weekly challenges will enable students to build their portfolio of tasks, which is essential for earning marks in this module.
>
> For information on how marks are awarded in this module, please refer to the module specifications (*see CST4050 intro*) or consult your teacher.
>
> It's important to note that your instructor will not provide the solutions to any weekly challenges; it is solely the responsibility of students to solve them and propose their solutions in class.

# Challenge

## Description

In this challenge, you have access to a comprehensive house prices dataset (see your MyLearning folder), containing various property details, such as:

- **Rooms**: average number of rooms per dwelling.
- **Age**: proportion of owner-occupied units built prior to 1940.
- **Distance**: weighted distances to employment centres.
- **Accessibility**: index of accessibility to highways.
- **Tax**: full-value property-tax rate per $\$10,000$.
- **DisadvantagedPosition**: % lower status of the population.
- **Crime**: per capita crime rate.
- **NitricOxides**: nitric oxides concentration (parts per 10 million).
- **PupilTeacher**: pupil-teacher ratio by town.
- **Residential**: proportion of residential land zoned for lots over 25,000 sq.ft.
- **NonRetail**: proportion of non-retail business acres per town.
- **Price**: prices of the house in thousands of dollars (USD).

Your objective is to perform the following tasks:

1. Train a linear regression model designed to predict house prices (dependent variable) based on the remaining independent variables.

2. Evaluate your modle using the R squared and the MSE

3. Interpret the regression model, highlighting the significance of the $\beta$ coefficients associated with standardised independent variables. Identify which independent variable exerts the most substantial influence on house price predictions.

٤. discuss the multicollinearity in the dataset

## Submission Guidelines

- Submit your code along with comments and explanations in your code to make it easy to understand.
- Discuss and interpret any result you get.
- Fully justify your choices.
- Feel free to add any visualisations to support your claims.

**Note**. To convert your notebook into a PDF file, start by exporting it to HTML, then open the HTML file and proceed to save it as a PDF using the `Save as PDF` option when printing.