**Real Estate Price Prediction**

**Problem Statement:**

Consider a real estate company that has a dataset containing the house price of unit area in the state of Maryland, USA. The company wishes to use the data to optimize the prices based on important factors such housing age, distance to the nearest MRT station, number of convenience stores, latitude, and longitude

**Essentially, the company wants —**

* To identify the variables affecting house prices per unit area,
* To create a linear model that quantitatively relates house prices with variables such as house age, distance to the nearest MRT station etc
* To know the accuracy of the model, i.e. how well these variables can predict house prices.

**Business Goal**

I am required to model the house price of unit area with the available independent variables. It will be used by the management to understand how exactly the house price of unit area vary with the independent variables. They can accordingly manipulate the design of the houses, the business strategy etc. to meet certain price levels. Further, the model will be a good way for management to understand the pricing dynamics of a new market.

Dataset

I got my dataset from Kaggle. The dataset has been uploaded to GitHub. It has 441 rows and 8 columns

**Important features**

House age, distance to the nearest MRT station, number of convenient stores, Latitude and longitude

**Target**

House price of unit area

**Data splitting into training and testing data**

**80% of the data is set aside for training our model while 20% is for testing**