## Machine Learning Project Proposal

Jaylon Kiper, Diviya Manirajan

CS-320-01

Dr. Nathan Johnson

10/21/21

For this project, we will use a publicly available dataset to construct a linear regression model that predicts the value of one of the features (variables) in the dataset.

Overview:

For this proposal, we wanted to use data based on real estate price prediction for selling houses. The variables included from this data include the location of the houses with longitude and longitude, the number of convenience stores around the area, the age of the house, the distance to the nearest MRT(mass rapid transit) station, and the price of the house based on unit area. To construct a linear regression model, we will take the variables for house age and distance to the nearest MRT station with the independent variable of house price of unit area to compare how each predictor is affected. An AI program would use the model by running through each age of the house and calculating the distances to an MRT station then comparing the predictors to the house price of unit area to determine which houses are cheaper or more expensive.

**Data Sources:**

**The database name**: Real estate price prediction

**The source:** <https://www.kaggle.com/quantbruce/real-estate-price-prediction>

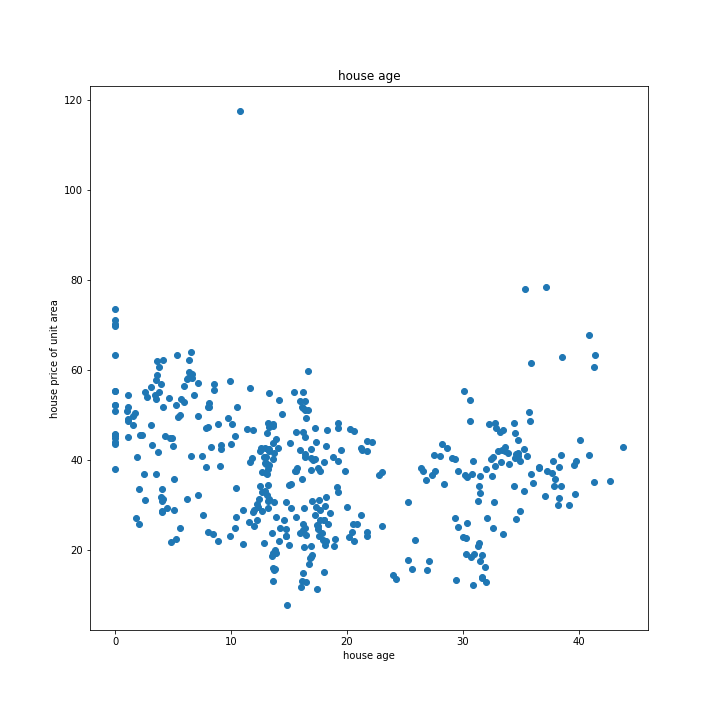
**Output Variable:** House price of Unit Area

**Predictors:** House Age and Distance to the nearest MRT station

| **No** | **Transaction date** | **House age** | **Distance to the nearest MRT station** | **Number of convenience stores** | **Latitude** | **Longitude** | **House price of unit area** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2012.917 | 32 | 84.87882 | 10 | 24.98298 | 121.54024 | 37.9 |
| 2 | 2012.917 | 19.5 | 306.5947 | 9 | 24.98034 | 121.53951 | 42.2 |
| 3 | 2013.583 | 13.3 | 561.9845 | 5 | 24.98746 | 121.54391 | 47.3 |
| 4 | 2013.5 | 13.3 | 561.9845 | 5 | 24.98746 | 121.54391 | 54.8 |
| 5 | 2012.833 | 5 | 390.5684 | 5 | 24.97937 | 121.54245 | 43.1 |
| 6 | 2012.667 | 7.1 | 2175.03 | 3 | 24.96305 | 121.51254 | 32.1 |
| 7 | 2012.667 | 34.5 | 623.4731 | 7 | 24.97933 | 121.53642 | 40.3 |
| 8 | 2013.417 | 20.3 | 287.6025 | 6 | 24.98042 | 121.54228 | 46.7 |
| 9 | 2013.5 | 31.7 | 5512.038 | 1 | 24.95095 | 121.48458 | 18.8 |
| 10 | 2013.417 | 17.9 | 1783.18 | 3 | 24.96731 | 121.51486 | 22.1 |

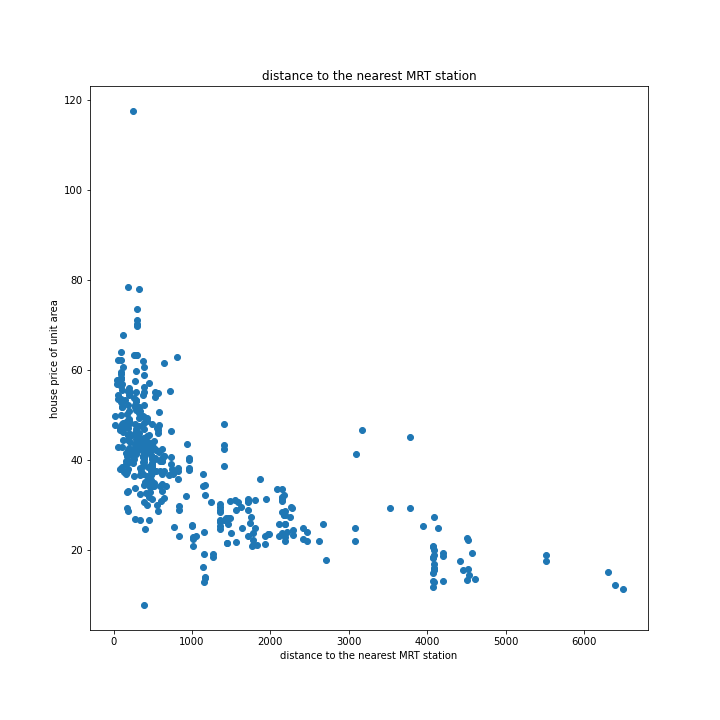
**Model Details:**

House Age vs. House Price of Unit Area



Using this scatterplot, we can see that there is a linear relationship between the age of a house and the house price of the unit area. The two variables seem to have a negative correlation wherein as the age of a house increases, the house price decreases.

House Age vs. Distance to Nearest MRT Station



Using this scatterplot, we can see that there is a linear relationship between the distance to the nearest MRT station and the house price of the unit area. The two variables seem to have a negative correlation wherein as the distance to the nearest MRT station increases, the house price of unit area decreases.