MMDT091_May Mon
Mentor - Ma Nu Wai Thet

This time, we are going to test with multiple variables. The value we want to predict is the price of housing based on the bedrooms and sqrt_living.

First, we defined the X and Y values in the dataset, the values we want to use.

Then we used the scaler for the X values. This is because our values have different scales.

There are three main variables in the datasets which are sqrt_living, bedrooms, and price.

When we analyze the value we get from the dataset, the results we got from the data are,
[-53567.44561984 288667.55112691] 540166.7341666666

The bedroom data is minus and I think that is something we need to consider.

As far as I understand, this means if we increase the number of rooms by 1 , we can get a decrease in price -53567.44561984.

However, regarding the sqrt_living, if we increase 1, we can get a higher price by 288667.55112691.

If we don't increase the price, we can get the average price of 540166.7341666666.

Also if we compare, actual value and predicted value by X_test = [[2, 860]]
We can see the difference.

The actual price from dataset - [235932.46722026]
The scaled data from the model - 303760.2105263158

I also test with 3-bedroom, 1400 sqft house
And then the results are

Predicted price: [347895.94679533]
Average actual price: 356309.53684210527

It is not really accurate, it can mostly predict the estimated values.

My final reflection from this exercise is that although we need to consider other factors for our pricing for the house, with these two data, we can explain and understand how real world estate pricing works and how we can predict price based on two models.

Regarding the results, people can look for the size of the house rather than how many bedrooms in each house. That is also reasonable.