HDB Resale Price Prediction Model Improvement Report

The initial neural network architecture was built using TensorFlow's Keras API and consists of 3 layers. This includes a one input layer, two hidden layers and a single neuron output layer. An input layer is determined by X_train.shpae[1] and accepts 51 features. Two hidden layers with 32 neurons and 16 neurons respectively both using the ReLU activation function. A single neuron output layer with a linear activation function which is suitable for regression tasks. The total number of trainable parameters in the initial model is 2,209. The model was trained for 10 epochs with a batch size of 32. The performance metrics for this initial model on both the training and test datasets are as follows:

```
Model: "sequential_1"
```

Layer (type)	Output Shape	Param #
dense_3 (Dense)	(None, 32)	1,664
dense_4 (Dense)	(None, 16)	528
dense_5 (Dense)	(None, 1)	17

```
Total params: 6,629 (25.90 KB)
Trainable params: 2,209 (8.63 KB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 4,420 (17.27 KB)
```

```
df_results_initial = df_results.copy() # Make a copy to preserve initial results
print("Initial Model Results:")
print(df_results_initial)
```

```
Initial Model Results:
                                Train      Test
Root Mean Squared Error       86321.39  86585.84
Mean Aboslute Error           65007.56  65114.84
Mean Aboslute Percentage Error    10.84     10.84
R2 score                           0.82      0.82
```

In the phase of new model improvement, the model has 5 layers that includes 4 hidden layers and 1 output layer. Four hidden layers with 64, 32, 16 and 8 neurons respectively both using the ReLU activation function. The total number of trainable parameters in the initial model is 6,081. The model was trained for 10 epochs with a batch size of 32. The total training time for this new model is 166.96 seconds. The new model has significantly improved performance as indicated by the much lower RMSE score and higher R2 score 0.91 on both the training and testing datasets. The performance metrics for this initial model on both the training and test datasets are as follows:

```
--- Training New Model ---
New model training time: 166.96 seconds
```

Model: "sequential_4"

Layer (type)	Output Shape	Param #
dense_16 (Dense)	(None, 64)	3,328
dense_17 (Dense)	(None, 32)	2,080
dense_18 (Dense)	(None, 16)	528
dense_19 (Dense)	(None, 8)	136
dense_20 (Dense)	(None, 1)	9

Total params: 18,245 (71.27 KB)
Trainable params: 6,081 (23.75 KB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 12,164 (47.52 KB)

```
3843/3843 ━━━━━━━━━━━━━━━━━━━━ 5s 1ms/step
1647/1647 ━━━━━━━━━━━━━━━━━━━━ 2s 1ms/step
New Model Results:
                                  Train      Test
Root Mean Squared Error         61569.5  61820.89
Mean Aboslute Error             44709.16 44759.74
Mean Aboslute Percentage Error      7.28     7.29
R2 score                            0.91     0.91
```

Comparison for both model

The new model significantly improved the predictions performance compared to the initial model. The Root Mean Squared Error (RMSE) decreased from approximately 86,585 to 61,820 on the test datasets showing a much lower average prediction error. Similarly, the R2 score increased from 0.82 to 0.91 suggesting that the new model explain a much larger proportion ot the variance in HDB resale prices. This model performance improvement can be assigned to the increased capacity of the deeper neural network. By adding more layers and parameters, the model was able to learn more complex features from the HDB transaction data. This allowed it to capture more complex non-linear relationships between the input features such as floor areas, flat types, etc and the adjusted resale price that leading to better prediction on unseen data. While the new model needed a longer training time 166,96 seconds compared to the initial models, the new model gains a better predictive performance. Therefore, a new neural network proved to be a more effective modelling approach for this HDB resale price prediction problem.