Project 2

Bangkok Housing Price

At first with individual perspective I try to calculate `price_per_sqm` to leverage unit since `land_area` and `floor_area` is different weight in remote province

But found out that in this data set `land_area` didn't correlate that much may be location range isn't different too much

| id - | 1 | -0.019 | 0.037 | 0.018 | 0.023 | -0.034 | 0.0033 | 0.013 | -0.0067 | -0.066 | -0.039 | -0.032 | -0.056 | -0.042 | -0.05 |
|-----------------------|---------|---------------|------------|---------|--------------|---------------|-------------|------------|-------------|-------------------|--------------------|-----------------------|----------------|--------------|---------|
| total_units - | -0.019 | 1 | -0.24 | -0.23 | -0.25 | 0.23 | 0.055 | 0.068 | -0.075 | 0.03 | 0.15 | -0.051 | -0.035 | 0.15 | -0.28 |
| bedrooms - | 0.037 | -0.24 | 1 | 0.85 | 0.8 | -0.25 | -0.0056 | 0.1 | 0.023 | -0.48 | -0.42 | | -0.47 | -0.74 | 0.24 |
| baths - | 0.018 | -0.23 | 0.85 | 1 | 0.8 | -0.23 | 0.016 | 0.089 | 0.042 | -0.44 | | | -0.41 | -0.66 | 0.33 |
| floor_area - | 0.023 | -0.25 | 0.8 | 0.8 | 1 | -0.25 | 0.027 | 0.099 | 0.045 | -0.43 | | | | -0.69 | 0.34 |
| floor_level - | -0.034 | 0.23 | -0.25 | -0.23 | -0.25 | 1 | 0.024 | -0.018 | -0.15 | 0.3 | 0.21 | 0.1 | 0.18 | 0.3 | 0.2 |
| land_area - | 0.0033 | 0.055 | -0.0056 | 0.016 | 0.027 | 0.024 | 1 | 0.0053 | -0.012 | -0.0095 | -0.013 | -0.013 | -0.002 | -0.0067 | 0.038 |
| latitude - | 0.013 | 0.068 | 0.1 | 0.089 | 0.099 | -0.018 | 0.0053 | 1 | -0.43 | -0.062 | 0.055 | -0.13 | -0.12 | -0.11 | -0.051 |
| longitude - | -0.0067 | -0.075 | 0.023 | 0.042 | 0.045 | -0.15 | -0.012 | -0.43 | 1 | -0.15 | 0.011 | -0.0071 | -0.037 | -0.031 | 0.036 |
| nearby_stations - | -0.066 | 0.03 | -0.48 | | -0.43 | 0.3 | -0.0095 | -0.062 | -0.15 | 1 | 0.33 | 0.41 | 0.56 | 0.49 | 0.24 |
| nearby_bus_stops - | -0.039 | 0.15 | -0.42 | | | 0.21 | -0.013 | 0.055 | 0.011 | 0.33 | 1 | 0.42 | 0.42 | 0.39 | -0.057 |
| nearby_supermarkets - | -0.032 | -0.051 | | | | 0.1 | -0.013 | -0.13 | -0.0071 | 0.41 | 0.42 | 1 | 0.73 | 0.38 | 0.23 |
| nearby_shops - | -0.056 | -0.035 | -0.47 | -0.41 | | 0.18 | -0.002 | -0.12 | -0.037 | 0.56 | 0.42 | 0.73 | 1 | 0.47 | 0.26 |
| year_built - | -0.042 | 0.15 | -0.74 | -0.66 | -0.69 | 0.3 | -0.0067 | -0.11 | -0.031 | 0.49 | 0.39 | 0.38 | 0.47 | 1 | -0.03 |
| price - | -0.05 | -0.28 | 0.24 | 0.33 | 0.34 | 0.2 | 0.038 | -0.051 | 0.036 | 0.24 | -0.057 | 0.23 | 0.26 | -0.03 | 1 |
| | - pi | total_units - | bedrooms - | baths - | floor_area - | floor_level - | land_area - | latitude - | longitude - | nearby_stations - | nearby_bus_stops - | nearby_supermarkets - | nearby_shops - | year_built - | price - |

- 0.8

- 0.6

- 0.4

- 0.2

- 0.0

- -0.2

- -0.4

- -0.6

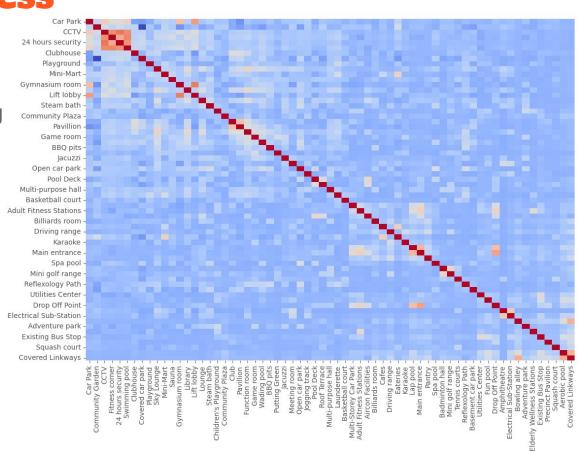
So I try with EDA process

Scanning for `nil` value and clean its up with `mean` value (after use individual expectation to only 1 floor for Detach House etc..)

Classified `property_type` and `province` also `district` with `get_dummies`

Extract `facilities`

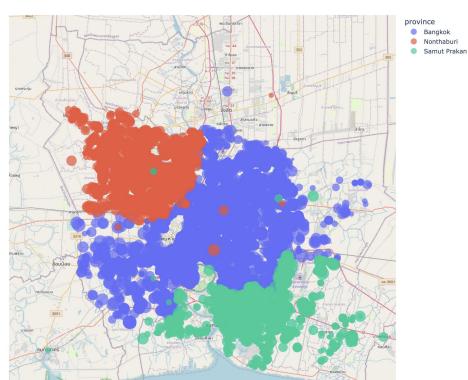
into column and modeling



After all of normal process of EDA Kaggle score is better than worst benchmark

model

After seeing the map



I try experiment (with some guidance from Nozomi)

As hypothesis that `Siam` is the most expensive housing price in Bangkok

So .. as Spider Man far From Home I code

Experiment How Far From Siam

create function to get distant from Siam (hypothesis that Siam is Expensive price for Accomodation)

And got a luck that R2 score is better

4 fake-or-dead



1,183,779

5

9s



Your Best Entry!

Your most recent submission scored 1,183,779, which is the same as your previous score. Keep trying!

82%

Better that Dummy Regression

-14.41%

From the Best Model

Thank you