

Project 2

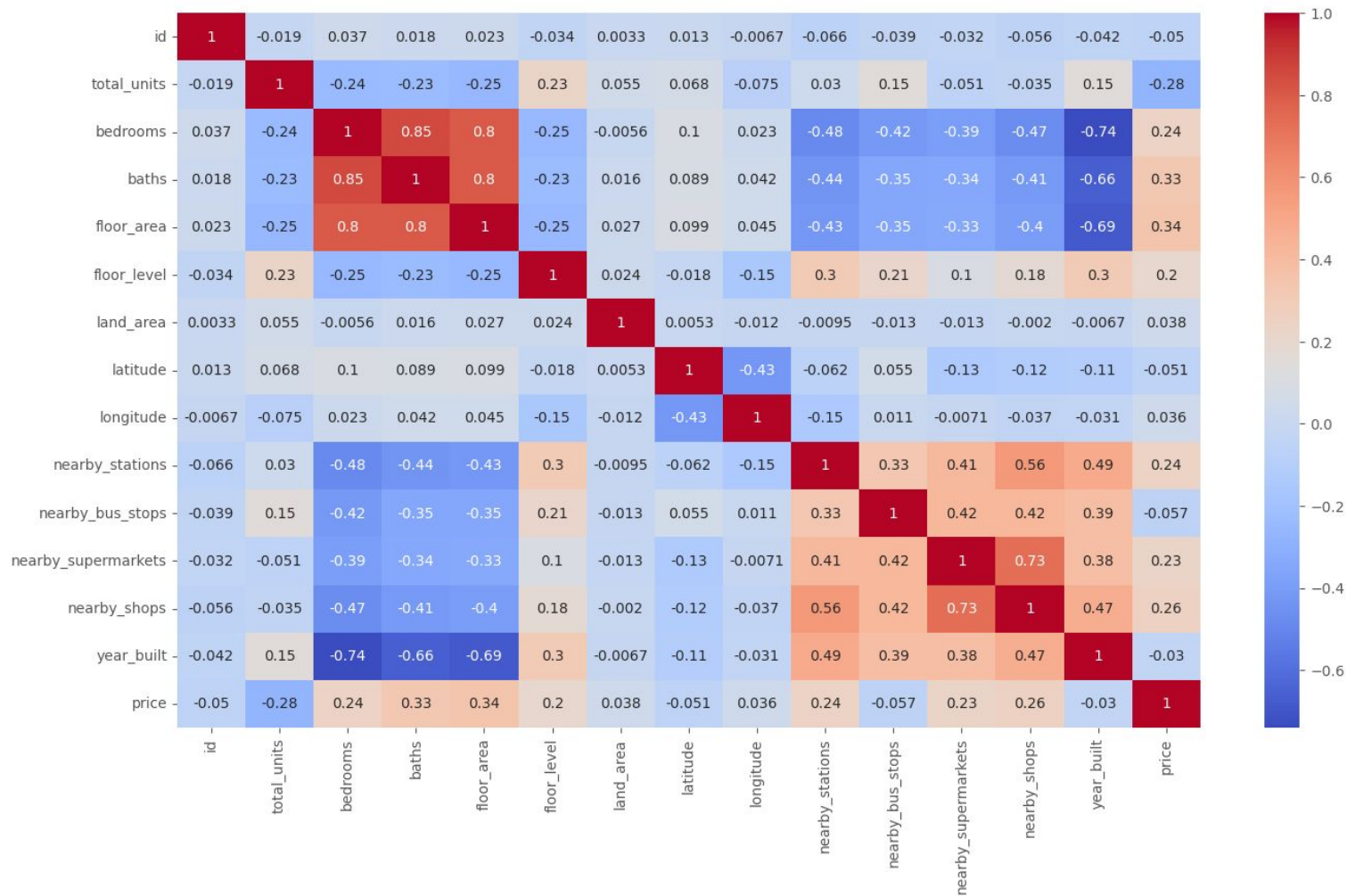


Bangkok Housing Price

Thought Process

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



Thought Process

So I try with EDA process

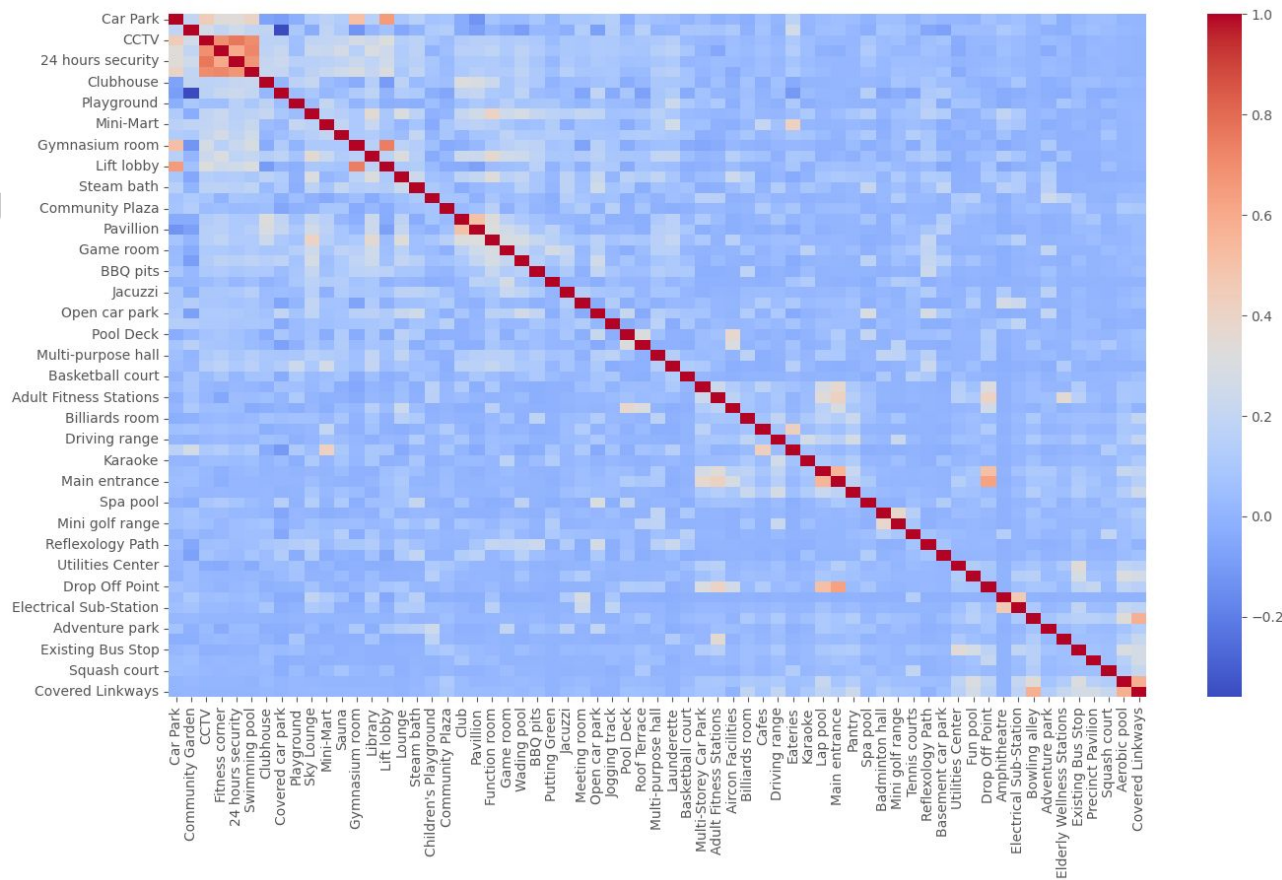
Scanning for `nil` value and clean its up with `mean` value (after use individual expectation to only 1 floor for Detach House etc..)

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

Thought Process

Extract `facilities`

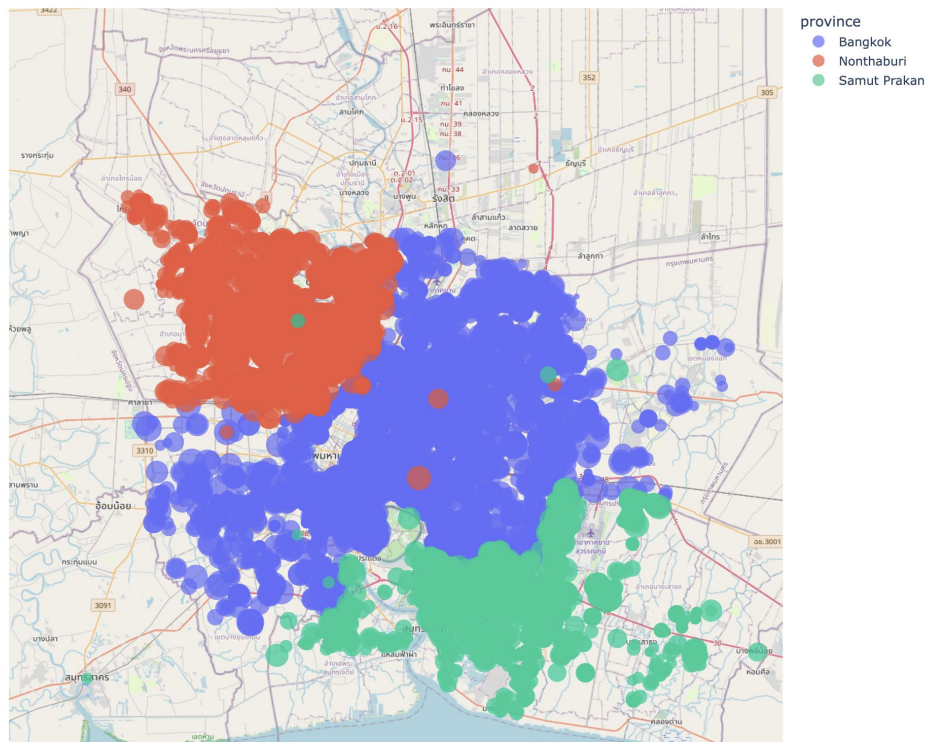
into column and modeling



Thought Process

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

After seeing the map



Thought Process

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)

```
def distant_to_siam(lat, long):  
    # https://maps.app.goo.gl/E954o1rAdTKxDHyG6  
    siam_lat = 13.7465  
    siam_long = 100.532  
    return np.sqrt((lat - siam_lat)**2 + (long - siam_long)**2)
```

16]

✓ 0.0s

Python

Thought Process

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 than Dummy Regression

-14.41%

From the Best Model

Thank you