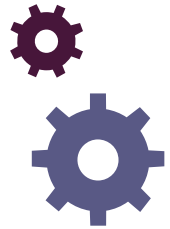
Data Analysis of Bangkok Housing

DSI03 - Project 2

Table of contents



01.Problem statement

Problem	House Buyers want to estimate the affordability of a property and make informed choices about where to live.
Background	House buyers have a budget in mind when searching for a new home so they want to narrow down their options and identify properties that fall within their budgetary constraints.
Objectives	Understand the expected costs of house owner and whether they are making a sound financial investment.

02. Data Import & Cleaning

Step 1 : Import data by using Pandas

Step 2 : Check for missing values , datatype and any obvious issues

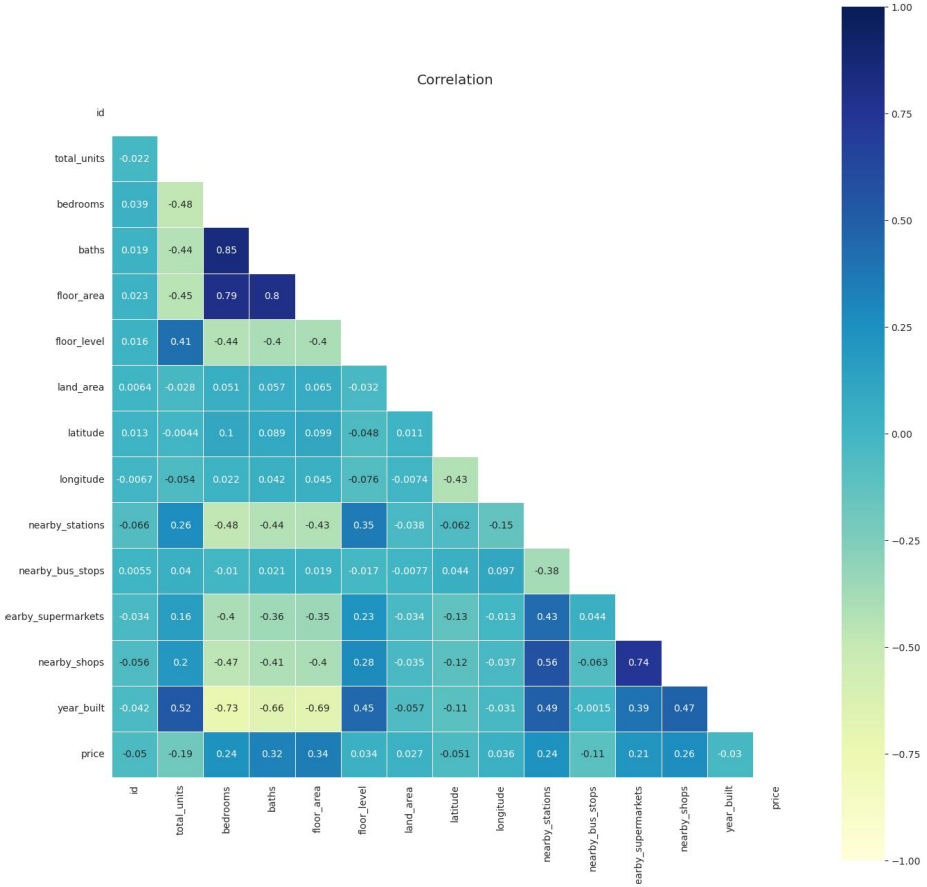
Step 3 : Fixing and cleaning data

- replace missing value in subdistrict columns using supplemental data
- replace null value with 0
- Get dummy for categories columns

Null Value

Others null values appear in features that are referring to the property surrounding so will conclude that the null values represent unavailability instead.

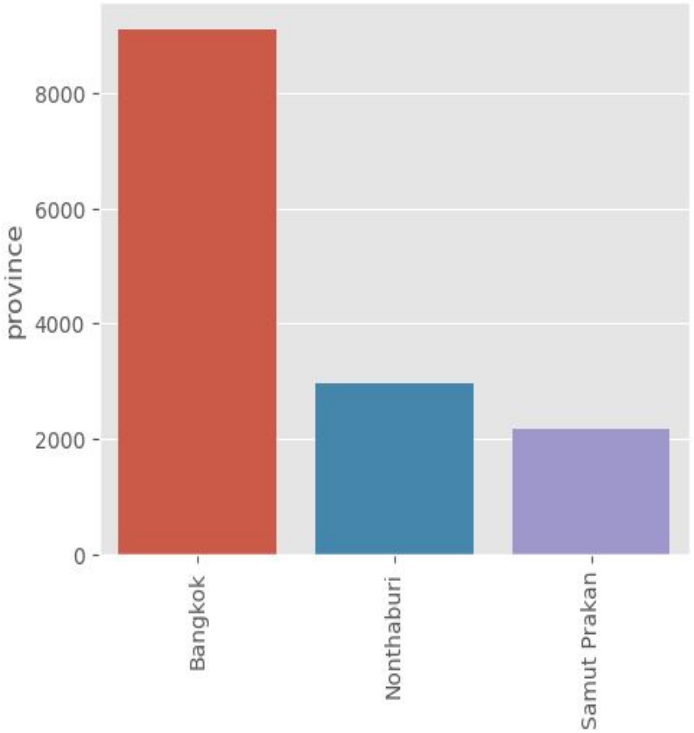
index	0
id	0
province	0
district	0
subdistrict	0
address	0
property_type	0
total_units	3762
bedrooms	43
baths	35
floor_area	0
floor_level	6178
land_area	9353
latitude	0
longitude	0
nearby_stations	0
nearby_station_distance	7043
nearby_bus_stops	8262
nearby_supermarkets	386
nearby_shops	0
year_built	0
month_built	5874
facilities	0
price	0

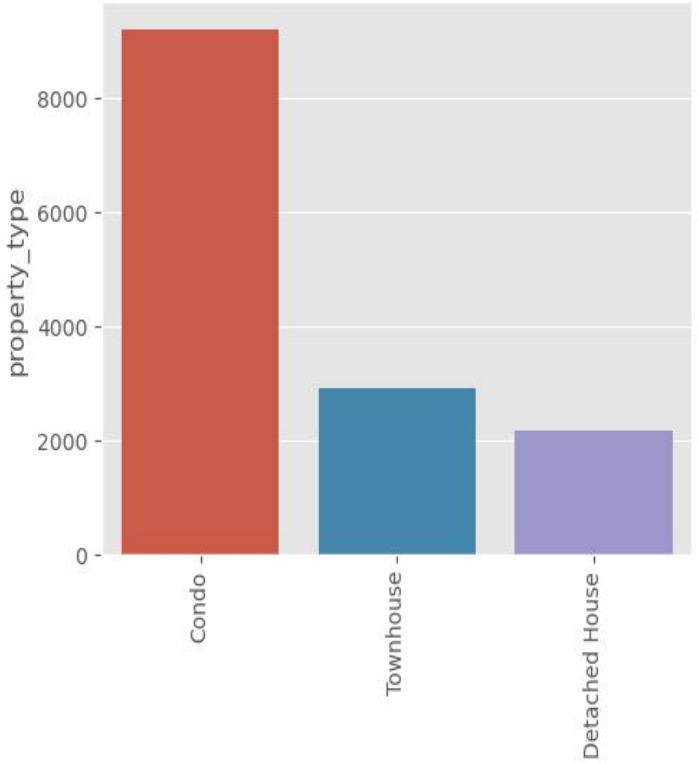


0.3 Exploratory Data Analysis & Visualisation

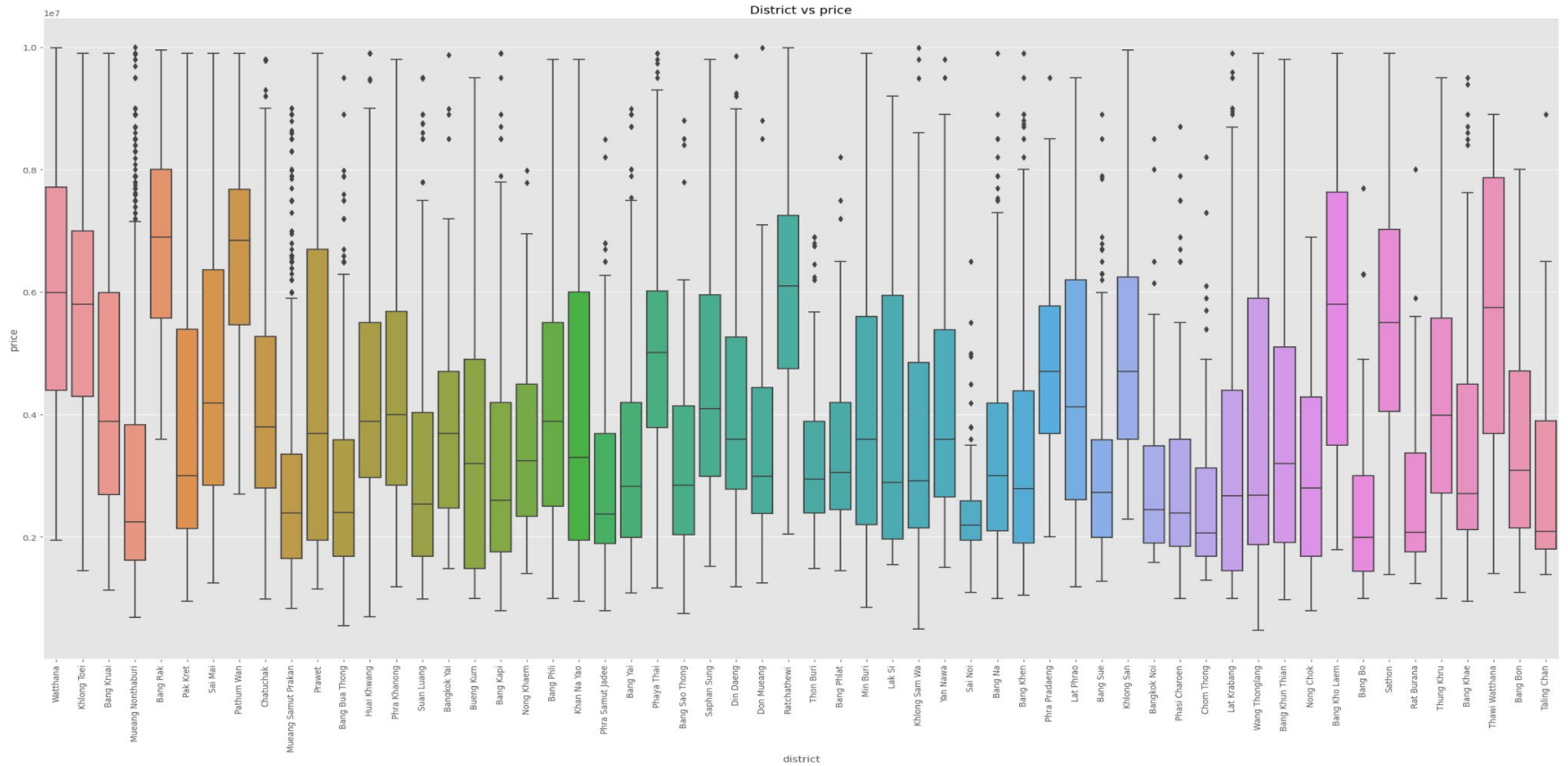
Refer to heatmap : Price has strong correlation with bedrooms, baths , floor area , nearby stations , nearby shops and nearby supermarkets

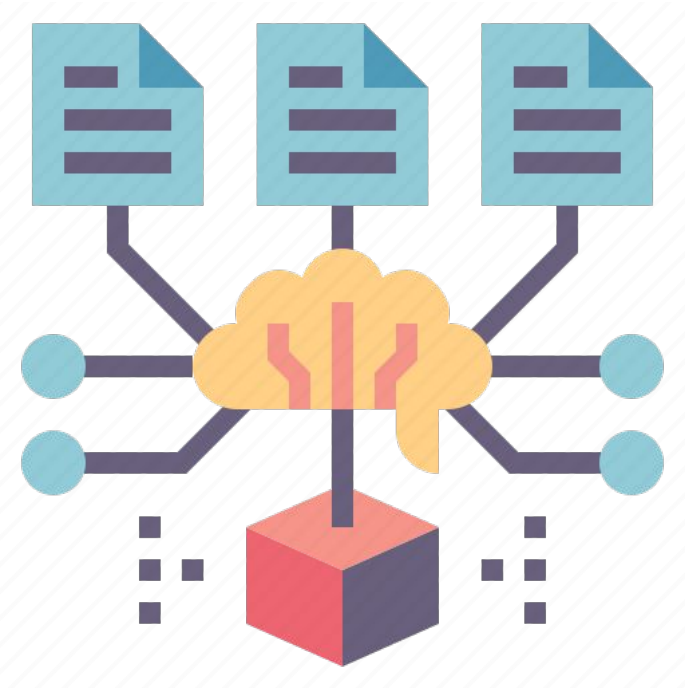
Bar Graph : Most of data province is in Bangkok and data property type is condo type





Boxplot : Data columns district has lot of outlier and different district has some effect with price





04

Modelling

Selected Features

	Features	Model type	R-square	RMSE
Model 1	bedrooms , baths , floor area ,nearby stations / shops / supermarkets	Linear Regression	0.40956	1675002.69
Model 2	Model 1 + property type + province	Linear Regression	0.46725	1591075.38
Model 3	Model 1 + property type + District	Linear Regression	0.62274	1338896.16
 **Model 4**	Model 3 + land area + longitude + floor level + year built	Linear Regression	0.63302	1320539.28

Selected Model Type : Model 4

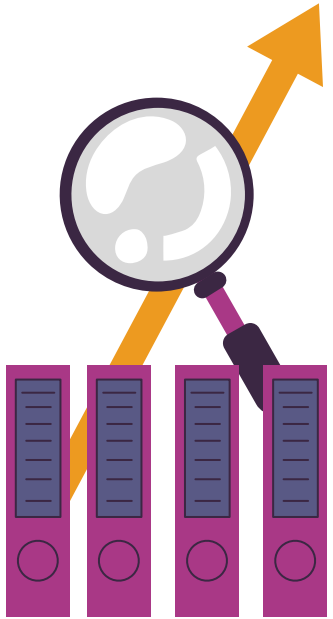
	R-square	RMSE	Cross Val Score	Kaggle Score
Linear Regression	0.63302	1320539.22	0.62874	1276333
 **RidgeCV**	0.63115	1320660.29	0.62692	1275923
LassoCV	0.43028	1645355.57	0.42998	1601265

05. Conclusions & Recommendations

Conclusions

House prices can vary significantly depending on various factors like features and surrounding of the house, including location, property size and house condition but house that more convenience, price will be more expensive like more rooms and use of space or near station , shop and downtown.

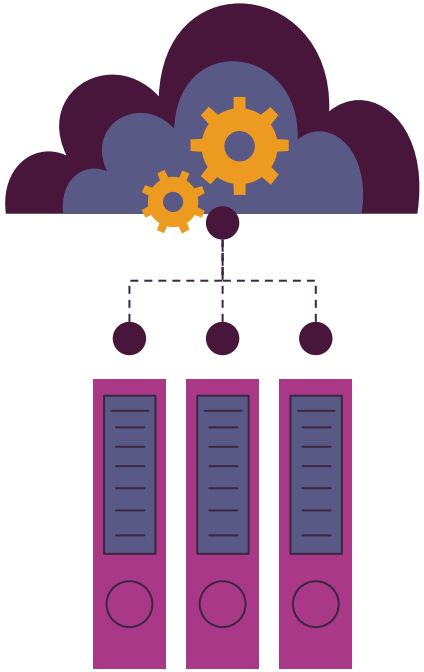
After considering all the above factors the model predicted value may not much accurate, due to incomplete Data , outlier and model complexity such as too many features or parameters.



Recommendations

For a homebuyer who have limit budget should consider create a list of features in a home that are in need to narrow down the choices and stay in the budget.

Thank you



CREDITS: This presentation template was created by Slidesgo, and includes icons by Flaticon, and infographics & images by Freepik