Condominium Price Prediction

Bangkok

Self Introduction

Chalermchon Wongsopa (Im)

Career History:

- Data Analyst / Business Intelligence Analyst
  KASIKORNBANK, *Bangkok*

- Data Analyst
  The BRS (Market research company), *Bangkok*

Education:

- Master in Agricultural and Resource Economics
  Kasetsart University, *Bangkok*

- Bachelor in Economics
  Chiang Mai University, *Chiang Mai*

Problem Statement

Problem Statement

Covid-19

- Prices decreased due to covid situation.

- Many campaigns and discounts are being used to reinforce market demand.

- Condominium prices are likely to increase in 2023 due to higher construction costs and also residential market recovery.

Problem Statement

Price & Demand Indices by Property Type

| Condominium | Single-detached house | Townhouse |

Price Index   Demand Index
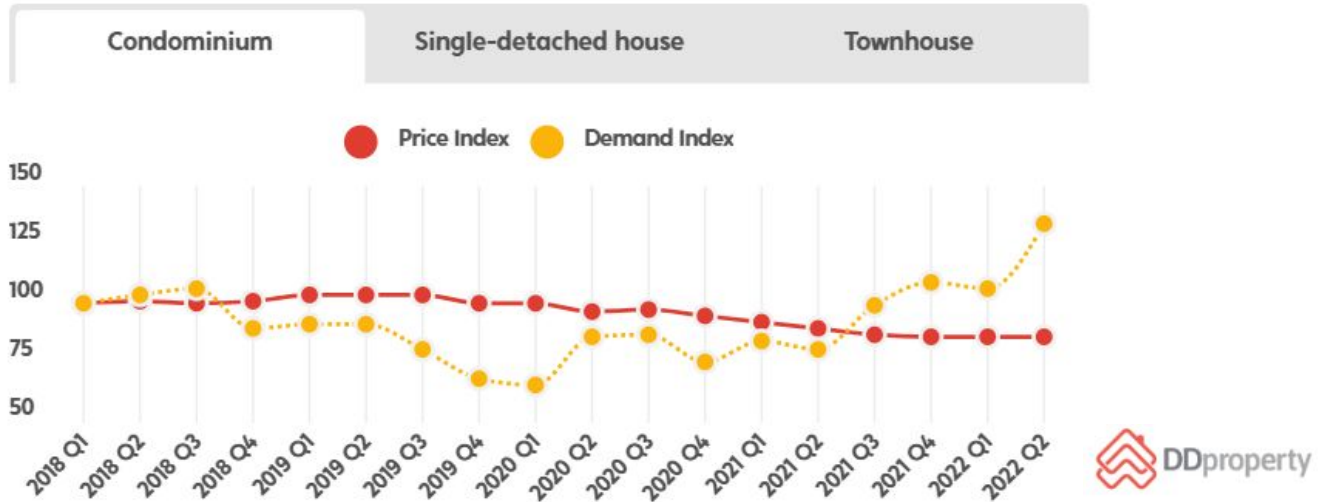
Objectives

- To conduct a model that could help investors to predict condominium prices

- To be one of the tools that helps buyers find real condominium values

Agenda

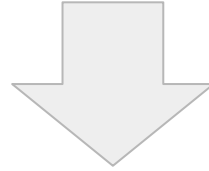**1** **Data Collection**

2 **Exploratory Data Analysis**

3 **Model Development**

4 **Conclusion & Future Works**

Data Collection

Data Collection

Unable to find raw Bangkok condominium data

Data Scraping

Data Scraping

Potential Features:

- District
- Developer
- Year build
- Towers
- Floors
- Sale price
- Rental price
- Bedrooms
- Bathrooms
- Internal areas

Data Scraping

SCRAPED

Amenities

✅ Elevator	✅ Parking	✅ 24-hours Security
✅ CCTV	✅ Swimming Pool	— Sauna
✅ Gym	— Garden / BBQ	— Playground / Kid's Area
— Shop on Premise	— Restaurant on Premise	✅ Wi-Fi

Location and Neighborhood

SCRAPED

Transportation

🚆	**Phra Khanong** 880 m / 4 minutes by car	
🚆	**On Nut** 1.2 km / 12 minutes by car	
🛣️	**Phra Khanong Toll Plaza** 1.7 km / 7 minutes by car	
🚆	**Ekkamai** 1.7 km / 8 minutes by car	
🛣️	**At Narong 2 Toll Plaza** 2.2 km / 15 minutes by car	
🛣️	**At Narong 1 Toll Plaza** 2.3 km / 16 minutes by car	

Neighborhood   Street View
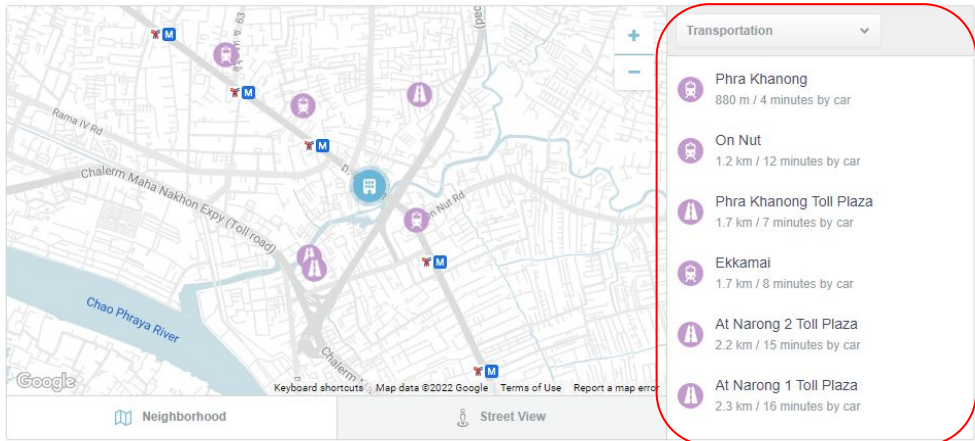
Market Stats for Aspire Sukhumvit 48

SCRAPED

Current asking price per sqm.	Asking price change from last quarter	Asking price change from last year	Achievable gross rental yield	Rent price change from last year
฿97,051	• 0 %	↓ -7.54 %	5.48 %	↑ 8.58 %

Potential Features:

- Elevator
- Parking
- 24-hour security
- CCTV
- Swimming pool
- Sauna
- Gym
- Garden/BBQ
- Playground/Kid's area
- Shop on premise
- Restaurant on premise
- Wi-Fi

Challenges of Data Quality

1   **Web users' typos**

- Double typing numbers of bedroom / bathroom

- Input whole condominium area instead of individual room area

2   **Old format page**

- Some old formats of web pages still exist on the website

3   **Different format of missing values**

- Each feature has it own way presenting absence of value

Exploratory Data Analysis

Data Preparation

Captured all 50 districts of Bangkok

Removed different rooms but same element

Rows
22,915

DATA LOSS 30%

DATA LOSS 13%

32,686
Rows

19,986
Rows
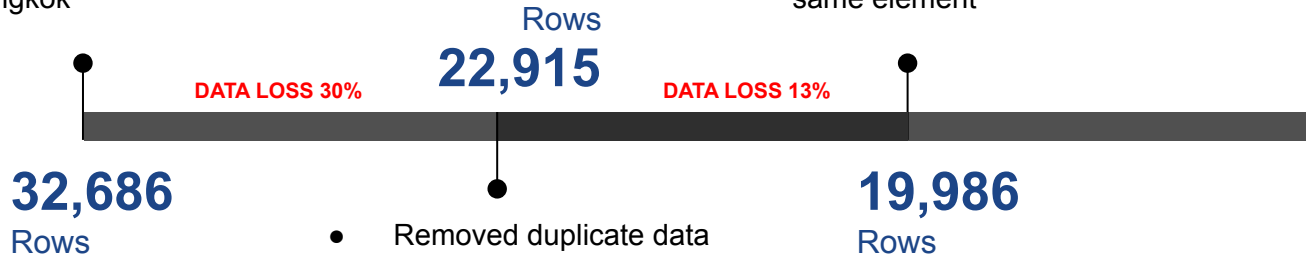
- Removed duplicate data from scrapping process

- Removed data from old web page format

Outlier Detection

Sale Price Histogram

Skewness: 9.66
Max: 580 mb
Min: 0.3 mb
Mean: 10.3 mb
Median: 4.5 mb
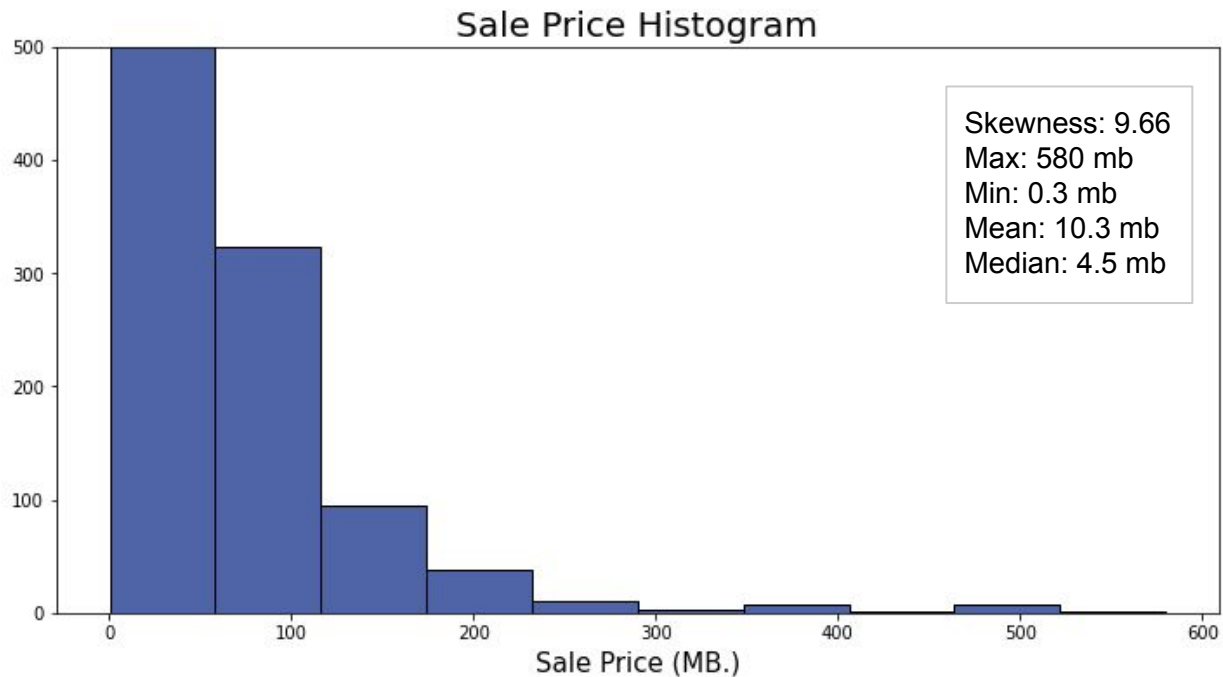
- Nature of condominium sale price is usually **extremely vast from different values for each district**

- To make our model perform better, at first stage, **we reduced our data scope to condominiums that price lower than 10 MB.**

- As per data reduction, **current dataset has 15,391 rows**

Sale Price Distribution of New Dataset

Sale Price Histogram

Skewness: 0.766
Max: 9.99 mb
Min: 0.3mb
Mean: 4.03 mb
Median: 3.45 mb

Model Development

Ways of Model Improvement

Missing Values

- Dataset dropping null rows

- Dataset with median imputation

- Dataset replacing null with zero

Feature Engineer

- Normal sale price

- Log transformation sale price

Models

- Linear Regression

- Random Forest

- XGBoost

- Neural Networks

24 Combinations

Missing Values Handling

Split

Training Set

Test Set
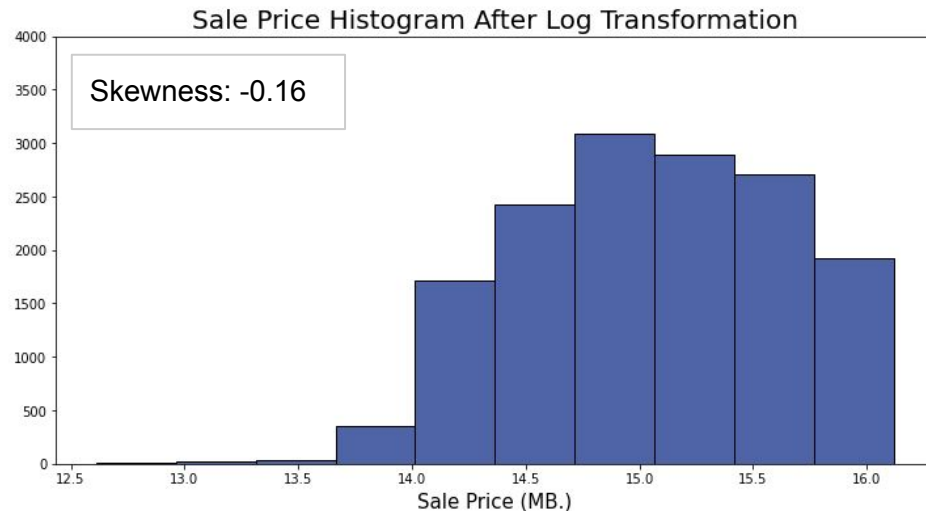
- Drop some data that contain missing values

- Impute missing values with column median

- Fill in zeros for some missing values

- Drop data that contain missing values

Feature Engineer

Sale Price Histogram — Skewness: 0.766

Sale Price Histogram After Log Transformation — Skewness: -0.16
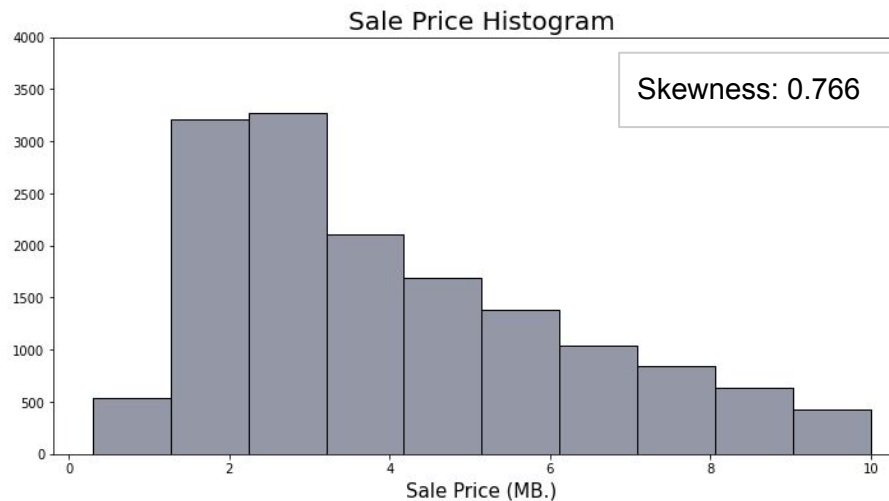
- Log transformation can reduce original `Sale Price` skewness.

Model Results

	Model	Fill N/A	Target (Sale Price)	Metrics	
				R-squared	**RMSE**
BASE LINE	Linear Regression	Drop	Normal	0.78	1,065,223
	Random Forest	Drop	Normal	0.90	692,664
	Random Forest	Median	Normal	0.91	678,173
	XGBoost	Drop	Log transform	0.93	654,897

Comparing Residual Plots

Random Forest
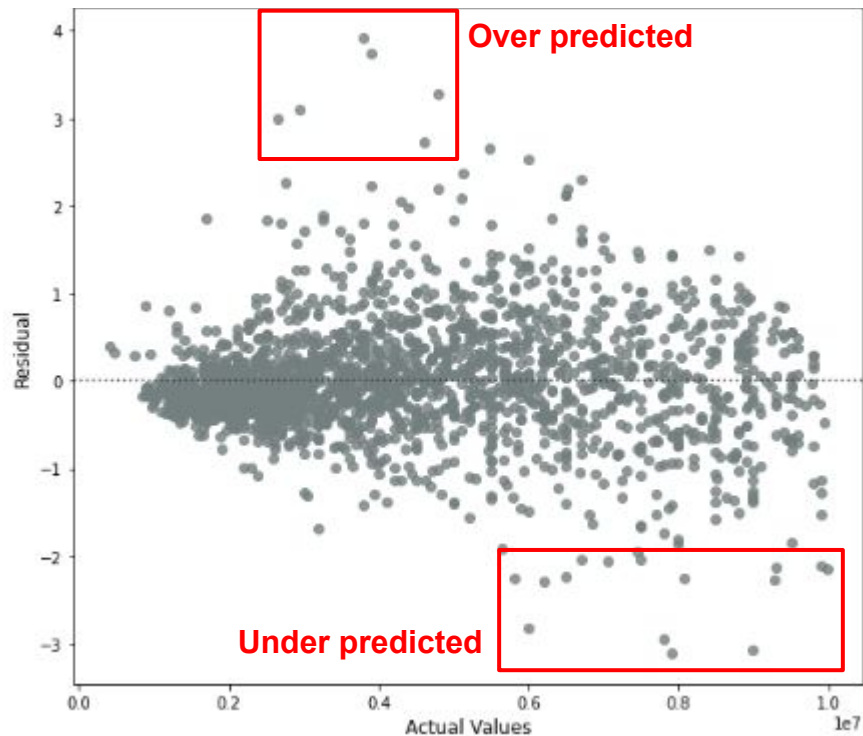
XGBoost with Log Transformation
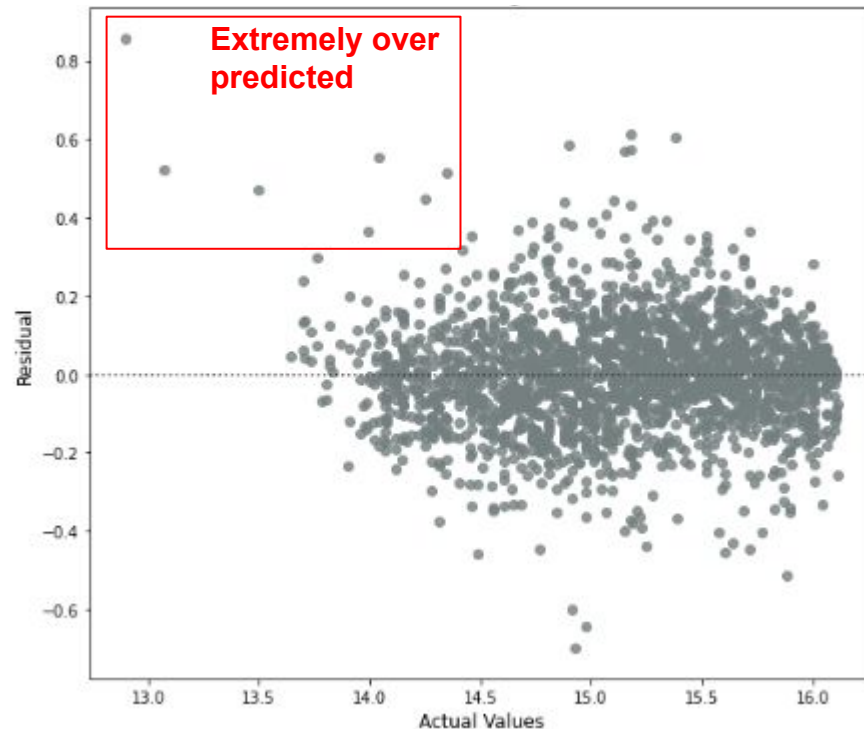
- XGBoost with log transformation does not perform well at lower price condominiums

Final Model's Performance

Test Case	Metrics	
	R-squared	**RMSE**
Split test dataset by **sale price**		
<= 5 MB	0.75	536,723
> 5 MB	0.57	887,431
Split test dataset by **number of bedroom**		
1 Bedroom	0.91	598,009
2 Bedrooms	0.81	891,786
More than 2 Bedrooms	0.43	1,078,638

- The final model tends to perform well with lower than 5 MB condominiums, which usually are one bedroom.

Test with the Other Website

DDproperty by PropertyGuru

฿**7,900,000** Baht Mortgage: ฿33,944 /month

2 🛏 2 🛁 57.6 sqm ฿137,153 / sqm

คอนโด the coast Bangkok (เดอะ โคสต์ แบงค็อก), Bangkok
Bang Na, Bang Na, Bangkok

🚶 Nearby Stations

🟦 11 mins (850 m) to BL30 Sam Yot MRT

🟦 13 mins (970 m) to BL31 Sanarn Chai MRT

PSM	Floor Level
฿137,153 / sqm	21
Built Year	Listing ID
N/A	8906630

Prediction

```
make_prediction(final_model, test_data)
◀
Expected value is: 6978676.703333332 THB
```

1 million baht off

Test with the Other Website

DDproperty
by PropertyGuru

฿6,100,000 [Baht]   Mortgage: ฿26,210 /month

1 🛏 1 🛁 55 sqm ฿110,909 / sqm

59 Heritage Sukhumvit 59, Bangkok
18 Soi Sukhumvit 59, Sukhumvit Road, Khlong Tan Nua, Watthana, Bangkok

🚶 **Nearby Stations**

🟩 6 mins (430 m) to E6 Thong Lo BTS

🟩 9 mins (650 m) to E7 Ekkamai BTS

Furnishing	Floor Level
Fully Furnished	19
Tenure	Built Year
Freehold	March, 2009

Key Features	Facilities
🏠 Covered car park	🎥 CCTV
🏋 Fitness corner	💂 24 hours security
🏊 Swimming pool	

Prediction

```
make_prediction(final_model, test_data)
```

Expected value is: 6091393.7514285715 THB

10 thousand baht off

Test with the Other Website

DDproperty
by PropertyGuru

฿2,590,000   Baht   Mortgage: ฿11,129 /month

Studio 🛏 1 🚽 27 sqm ฿95,926 / sqm

IDEO O2 บางนา, Bangkok
195 Sanphawut Road, Bang Na, Bang Na, Bangkok

🚶 Nearby Stations

🟩 8 mins (570 m) to E13 Bang Na BTS

🟩 11 mins (850 m) to E12 Udom Suk BTS

Furnishing	Floor Level
Fully Furnished	11
Tenure	Built Year
Freehold	January, 2017

Key Features | Facilities

🛒 Covered car park   📹 CCTV

🏋 Fitness corner   🏋 Gymnasium room

💂 24 hours security   🏊 Swimming pool

Prediction

```
make_prediction(final_model, test_data)
Expected value is: 2586482.848666667 THB
```

4 thousand baht off

Conclusion & Future Works

Conclusion

1 The final model is Random Forest with missing values replacement by the median.

2 Model performs better on lower price (< 5 MB) condominiums.

3 Currently, model performance based on data.

- If we have a large and up-to-date dataset, model tends to perform well.

4 Outliers are one of the important areas of concern when working with price prediction data.

Future Works

1 Spend more time with feature selection and feature engineering processes.

2 Explore more hyperparameter tuning.

3 Add emotional features such as interior design.

Thank you for your attention