



Group Project Presentation
on
**Predicting The Reality Of House
Prices In New Taipei City**
CHIA-HAO HSU

Introduction & Background

Introduction(5W1H)

What issue?

How much budget we should prepare to buy a house.

Where?

New Taipei City

When

The transaction in 2019

Who(target audience)

People in general and Employees

Why

Less knowledge or idea about how to buy a ideal house.

How

The analysis of linear regression to predict house price.



Whole Progresses



- 1 Introduction to the dataset
- 2 Data Exploring
- 3 Feature engineering
- 4 Regression model
- 5 Conclusion



Introduction to the Dataset

01

Data Sources

URL:<https://data.gov.tw/dataset/125903>

The screenshot shows a detailed view of a dataset page from the Government Information Open License Terms-First Edition. At the top, there's a navigation bar with links for ALL DATA SETS, SITE GUIDE, INTERACTIVE ZONE, INFORMATION STORY BUILDING (NEW), LATEST NEWS, and ADVISORY GROUP. Below the navigation is a breadcrumb trail: Home > Data set > ABOUT THE PLATFORM > Information-sales cases. The main title is "Real estate real price registration information-sales cases". A "Related Information Set" button is visible. The "Data set rating" is shown as 5 stars with the note "No votes yet". The "Data set description" includes two points: 1. Real-value registration information for real estate sales cases, including information on the target location (de-identification), area, and total price. 2. This data set is updated every 10 days. The "Description of main fields" lists various fields such as rps01 through rps27, detailing district, land section, building section, house number, total land transfer area, urban/non-urban zoning, transaction date, number of floors, building type, main use, building material, completion month, total area of building transfer, construction current structure pattern-room, construction current structure pattern-hall, construction status-wei, construction status-compartment, management organization, total price, unit price per square meter, parking category, vehicle displacement to total area, total parking price, remarks, and number. Below this, there are download links for CSV and View data, and a provision authority listed as New Taipei City Government Lands Bureau. The page also includes a contact name (Miss Liu) and provides details about update frequency, authorization, billing method, date added, data set type, interpretation, and keywords.

Related Information Set

Data set rating: ★★★★★ No votes yet

Data set description: 1. Real-value registration information for real estate sales cases, including information on the target location (de-identification), area, and total price. 2. This data set is updated every 10 days.

Description of main fields: district (village township), rps01 (subject to transaction), rps02 (land section location building section house number), rps03 (total land transfer area square meters), rps04 (urban land use zoning), rps05 (non-urban land Use zoning), rps06 (non-urban land use planning), rps07 (transaction date), rps08 (number of transaction pen s), rps09 (transfer level), rps10 (total number of floors), rps11 (building type) , Rps12 (main use), rps13 (main building material), rps14 (building completion month), rps15 (total area of building transfer square meters), rp s16 (construction current structure pattern-room), rps17 (construction current structure pattern-hall) , Rps18 (construction status-wei), rps19 (construction status-compartment), rps20 (with or without management orga nization), rps21 (total price), rps22 (unit price per square meter), rps23 (parking category), rps24 (Vehicle disp lacement to total area square meters), rps25 (total parking price), rps26 (remarks), rps27 (number)

Information download: CSV | View data | Real estate real price registration information-sales cases

Provision authority: New Taipei City Government Lands Bureau

Provide the name of Miss Liu ((02) 2960-3456 ext. 3353)

Update frequency: Every 10 days

Authorization: Government Information Open License Terms-First Edition

Billing method: free

Date Added: 2015/02/09

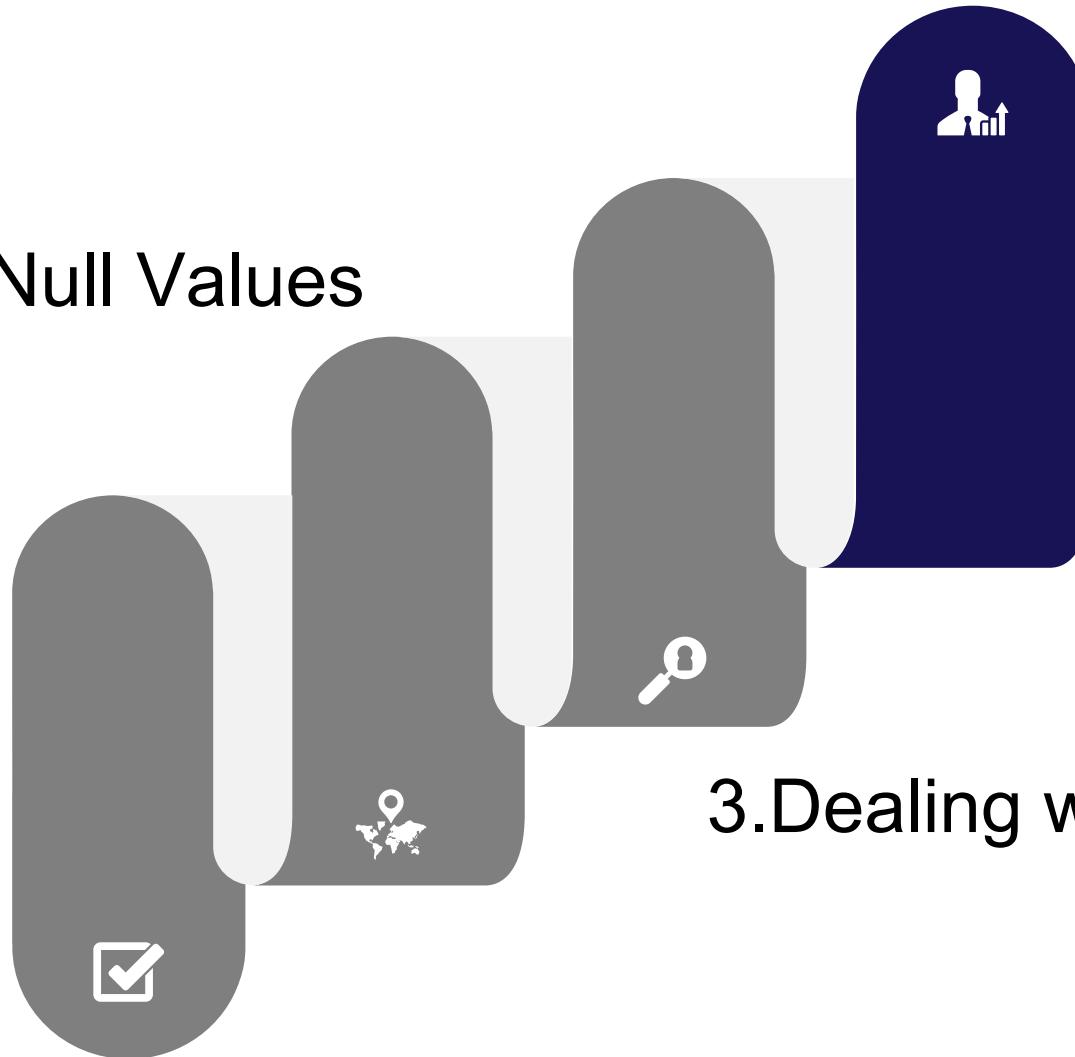
Data set type: System interface program

Interpretation: 2020/05/28 16:06

Keywords: Land price

Introduction to the Dataset

1. Initial Language
that was translated



3. Dealing with Null Values

4.
Rows 3637
Columns 26

Introduction of dataset(1)



Predictor



Response

Variable	Type	Unit	Description
district	object		The place where houses located
subject of transaction	object		sell house, land, or house+land
total land transfer area square meters	float64	square meters	The area of the land
urban land use zoning	object		Living,Business,Factory,Agriculture
non-urban land use zoning	object		hillside conservation, rural,general, industrial, forest, national park
non-urban land use planning	object		general, rural, hillside, forest, transportation,water
transaction year and month	int64	datetime	year month date

Introduction of dataset(2)



Predictor



Response

Variable	Type	Unit	Description
number of transaction pens	object		landX,houseY,parkZ X,Y,Z are the numeric value
transfer level	object		1,2,3 means it sell three levels together
total floors	float64	level	The total levels of building
building type	object		apartment, residential building,suite,office,factory
main use material	object		steel, brick, reinforced concrete construction
with or without management organization	object		Yes,No

Introduction of dataset(3)



Predictor



Response

Variable	Type	Unit	Description
construction room	int64	room	the total number of rooms(halls+bathrooms)
construction hall	int64	room	the total number of halls
construction bathroom	int64	room	the total number of bathrooms
construction status-compartment	object		Yes,No
year and month of completion of construction	float64	datetime	year/ month /date
total area of building transfer square meters	float64	square meter	
main use	object		living,business,factory,Parking

Introduction of dataset(4)



Predictor



Response

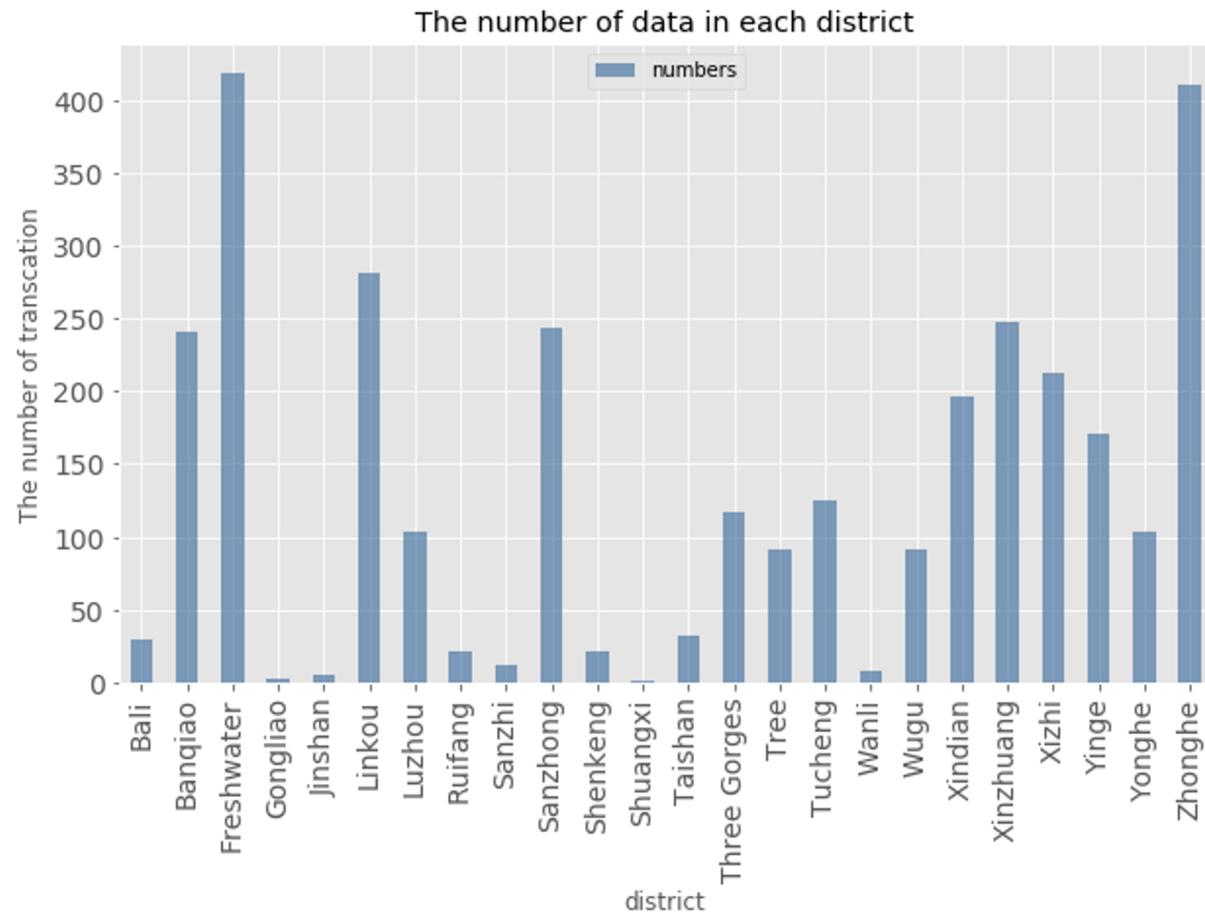
Variable	Type	Unit	Description
Total price	int64	dollar	The price sells in the transaction
unit price square meter	int64	dollar	Total price/ total land transfer area square meters
parking category	object		garage, vertical lifting parking system, parking area, other
Car displacement to total area square meter	float64	square meter	The area of parking space
total parking price	float64	dollar	The price sells in the parking space
number	object		The unicode of the transaction



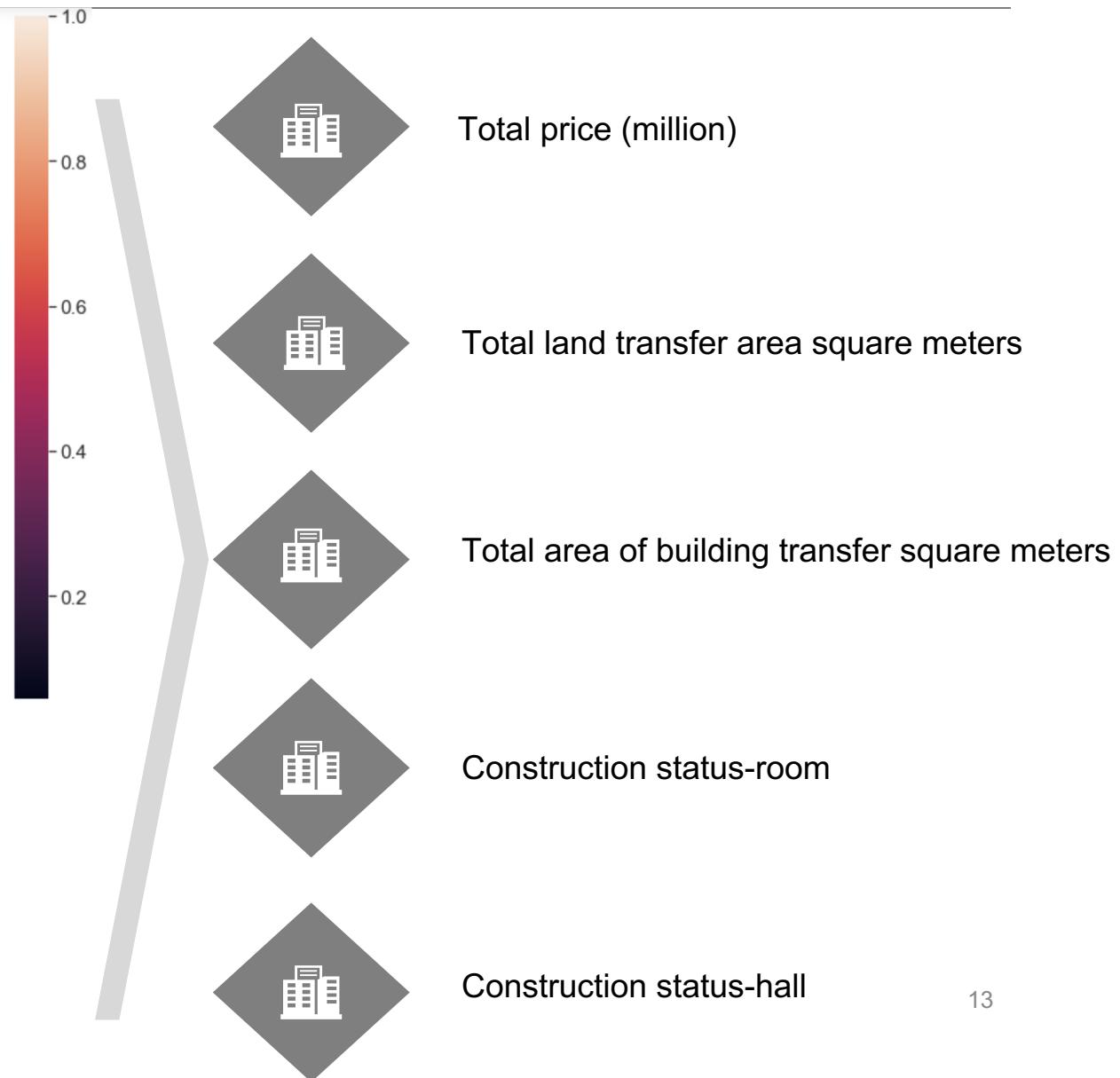
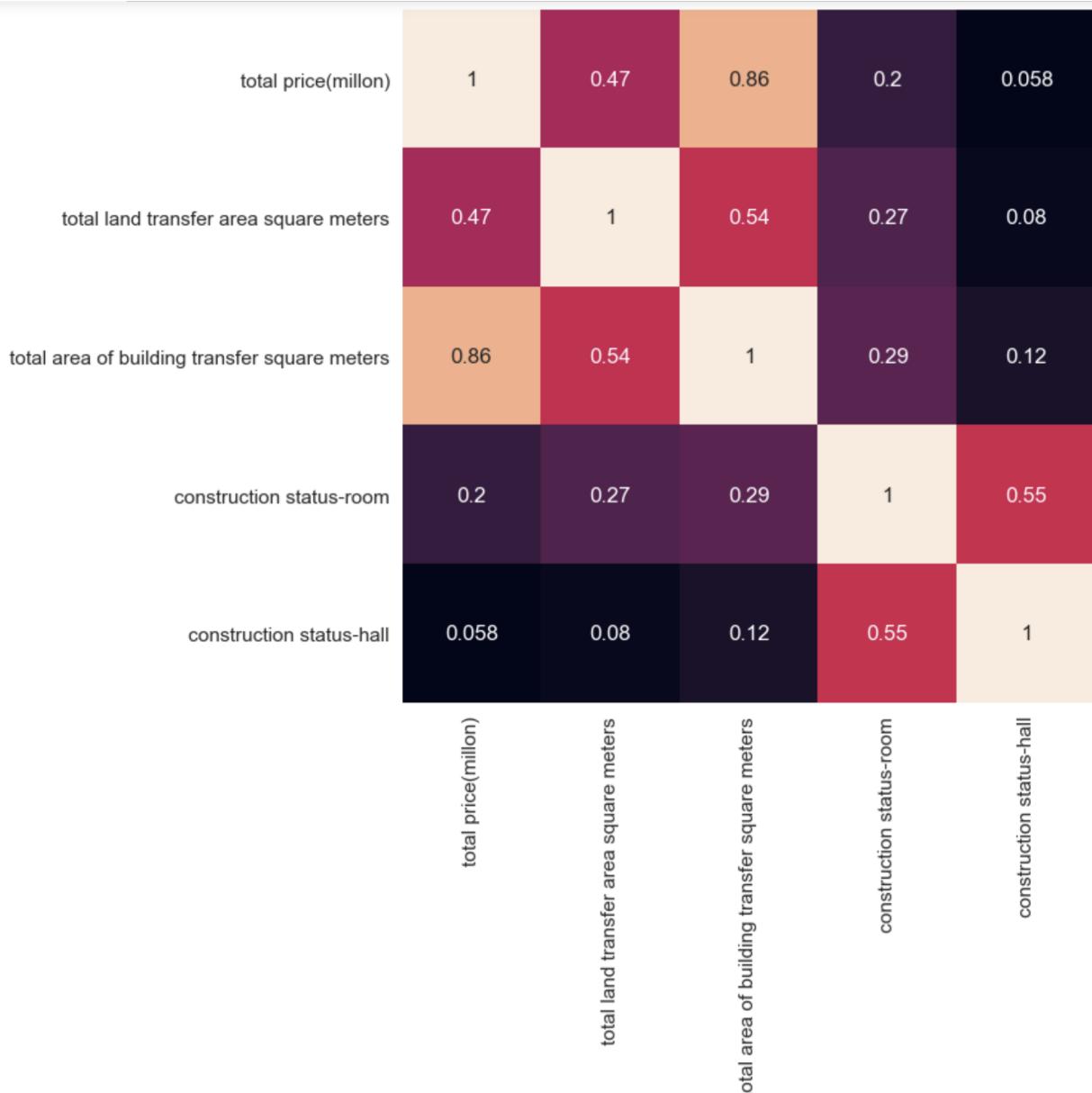
Data Exploring

02

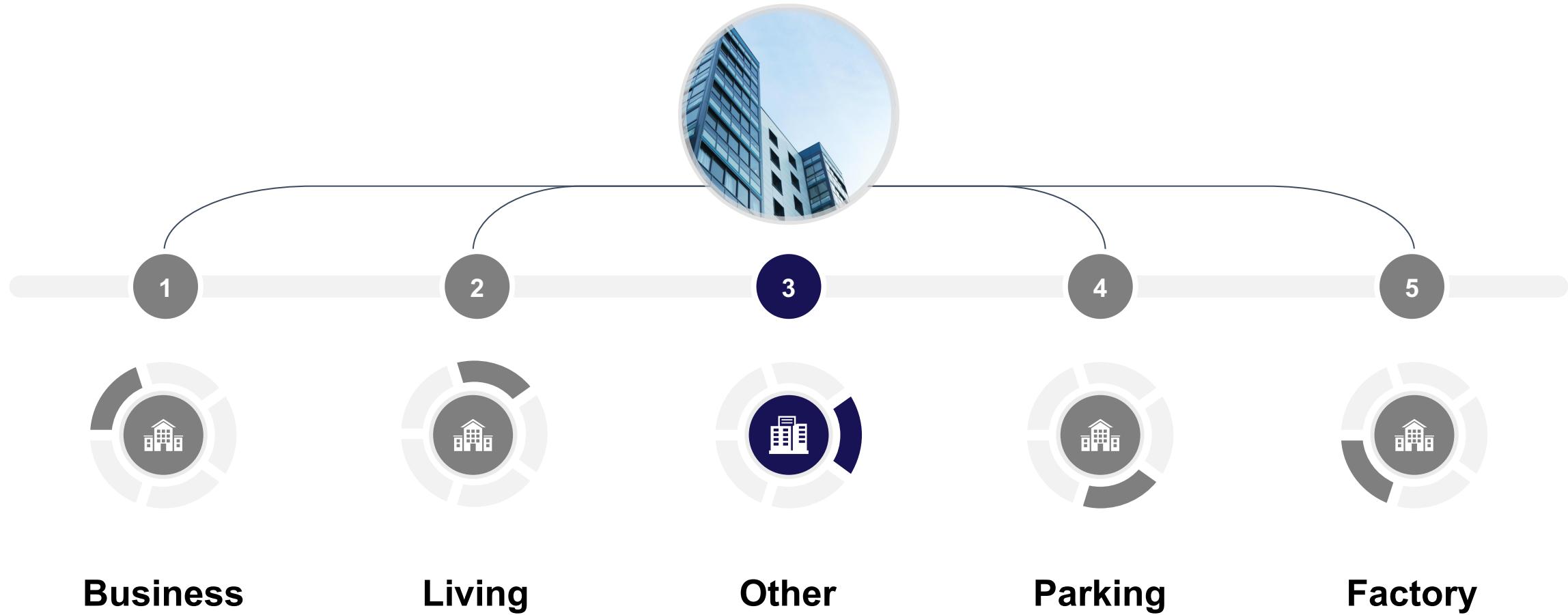
The Number Of Data In Each District



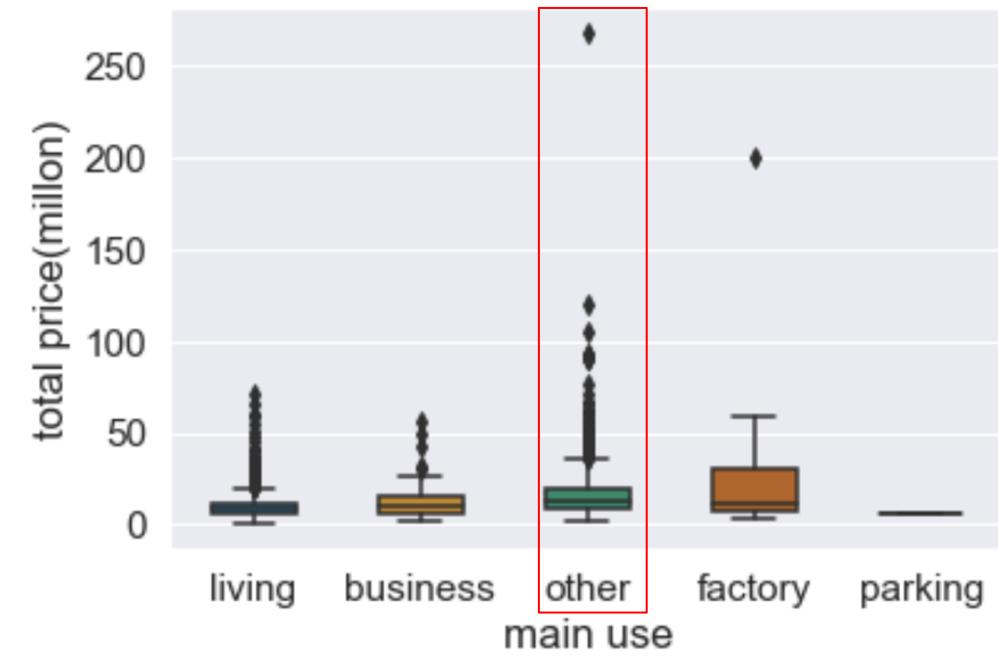
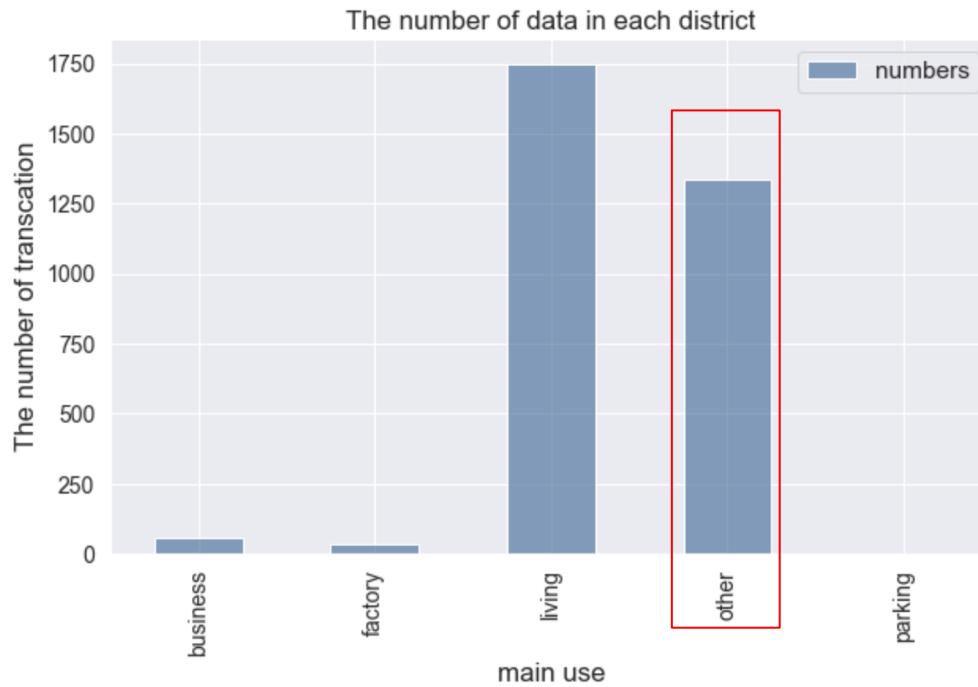
Heat Map Of Five Main Variables



Main Use For Building

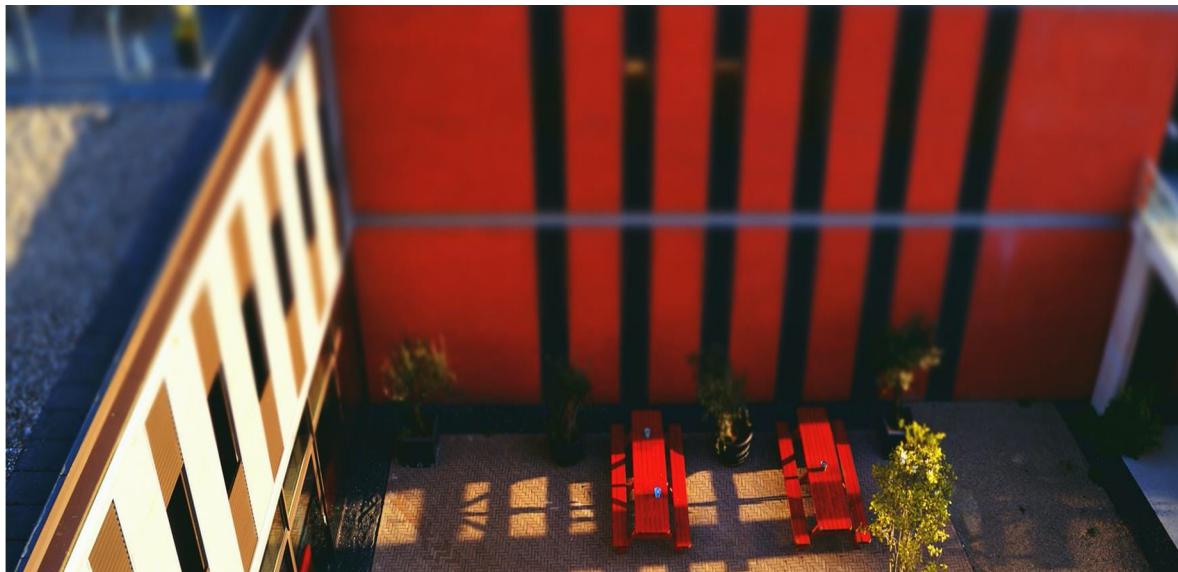
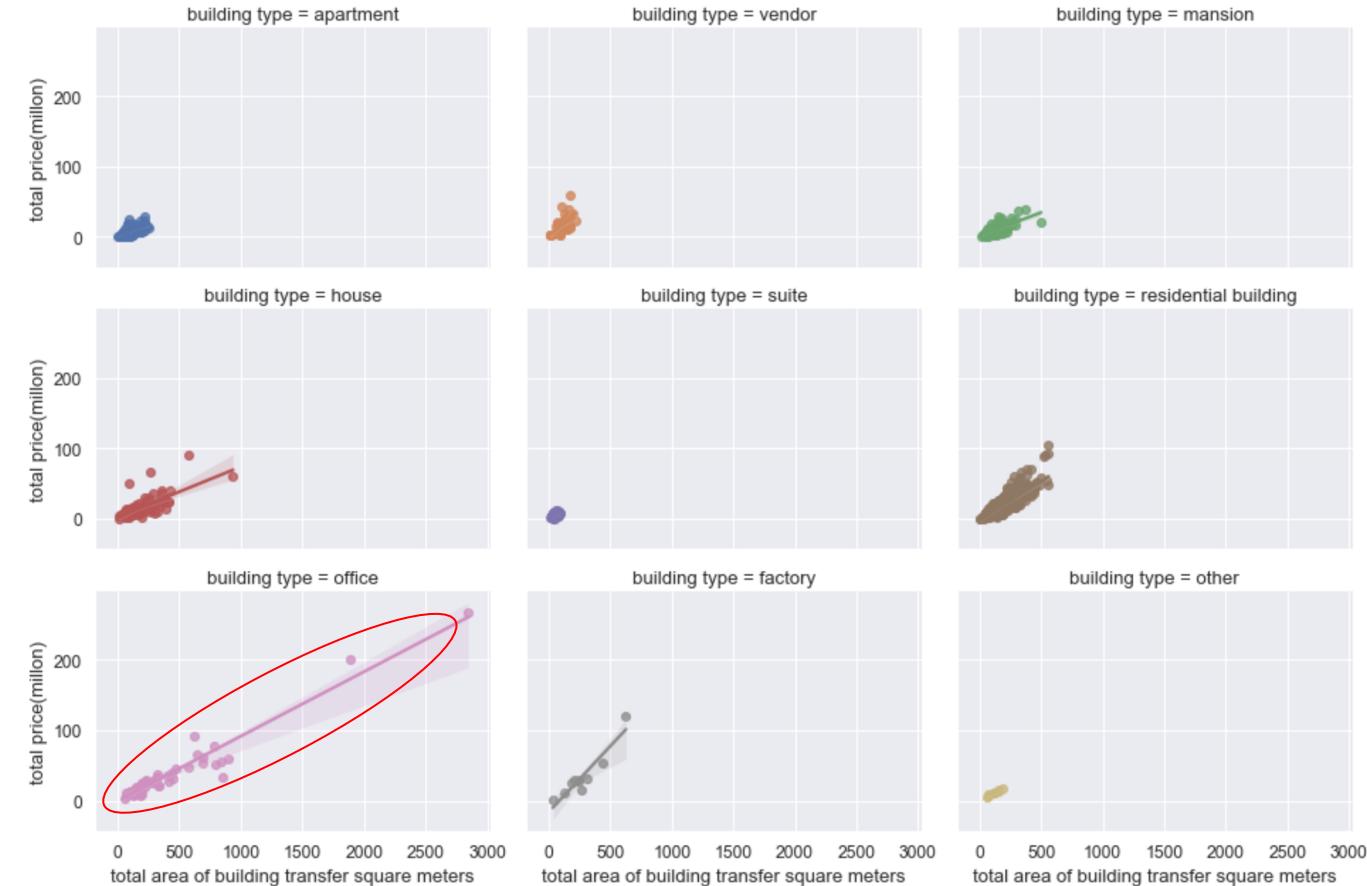


Main Use For Building



- Main use of 'other' has almost 1250 numbers of transaction, but 'other' has lots of outliers in total price.
- Learning the special trend for these cases.

— Building Type

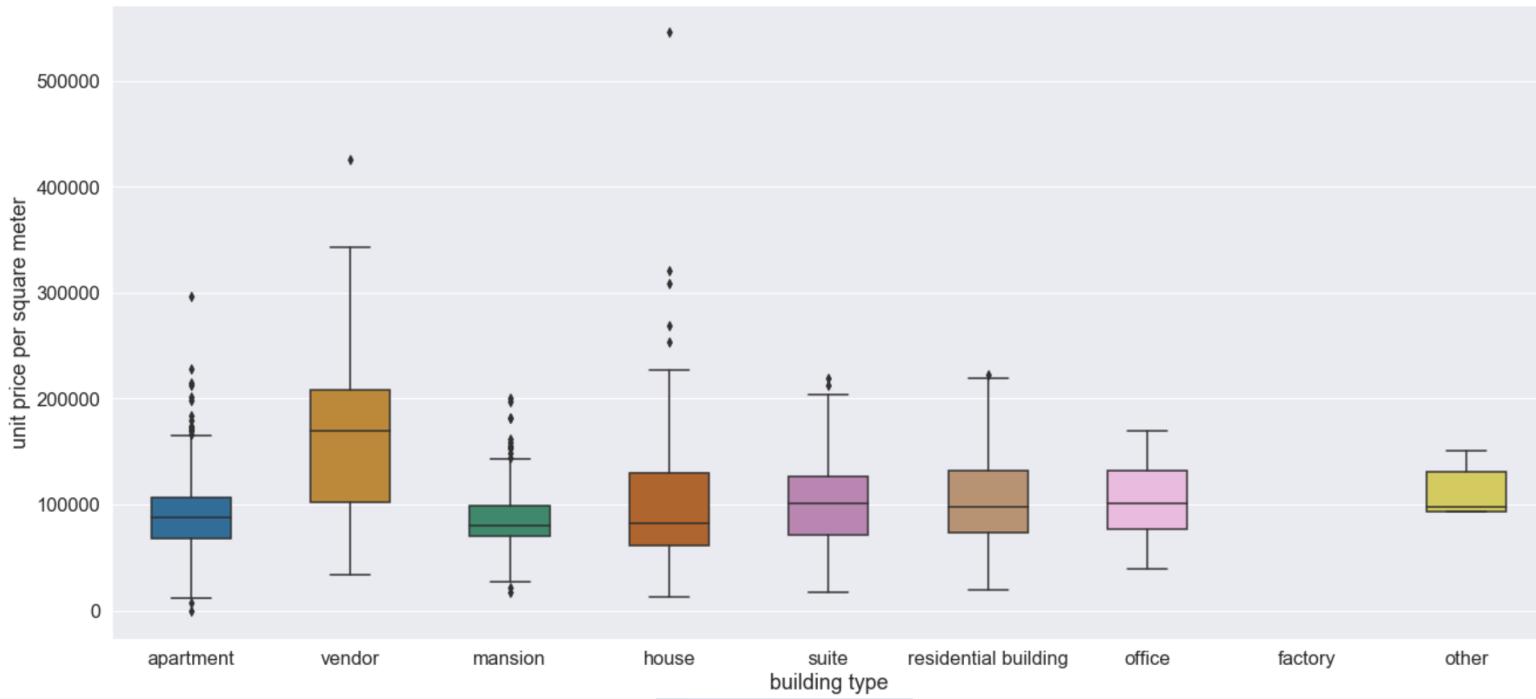


square meters price

In each building type, the correlation shows significant, but some of data in 'office' have too high prices.



Unit Price Per Square Meter By Building Type

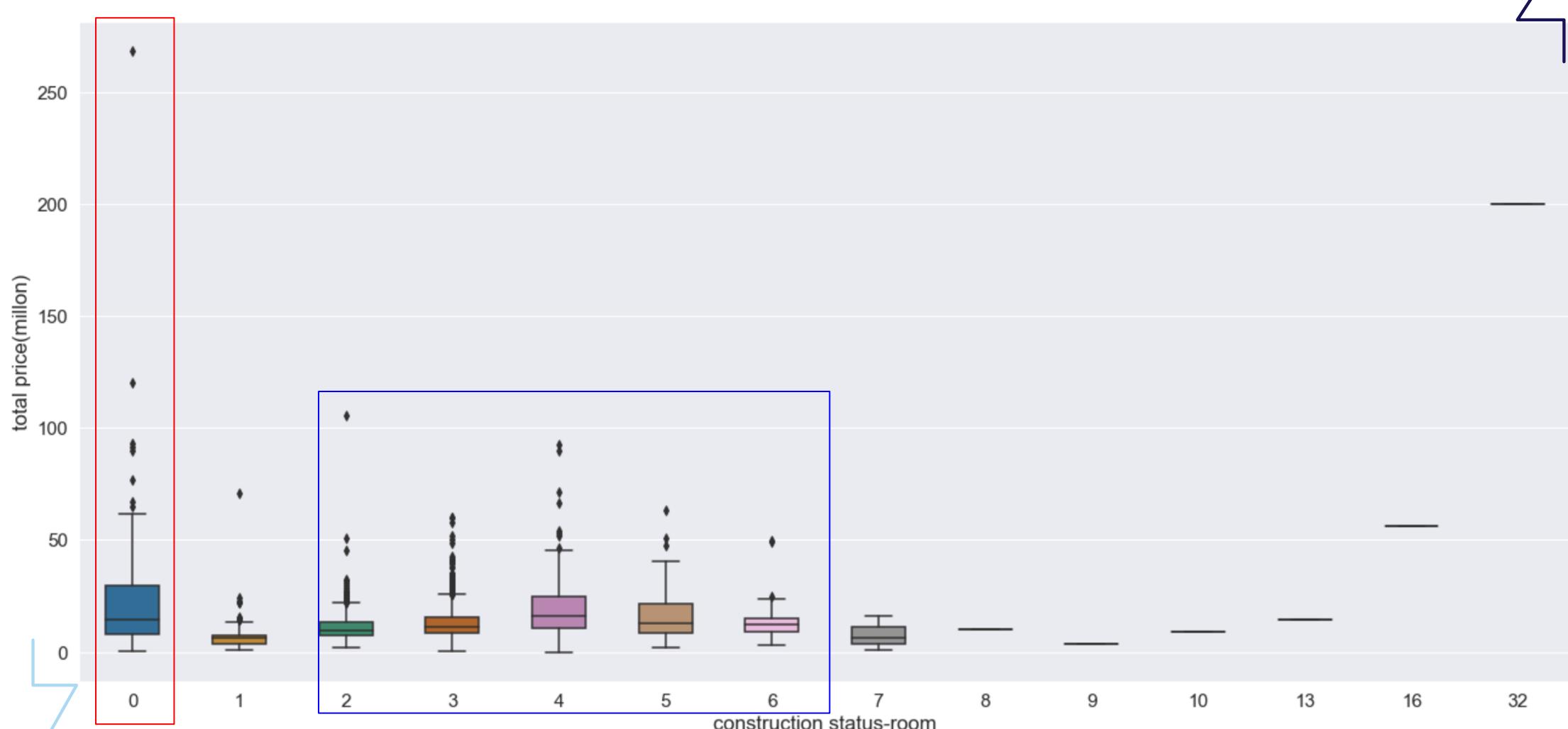


Much Outliers
in Apartment and House



Vendor worth
much price

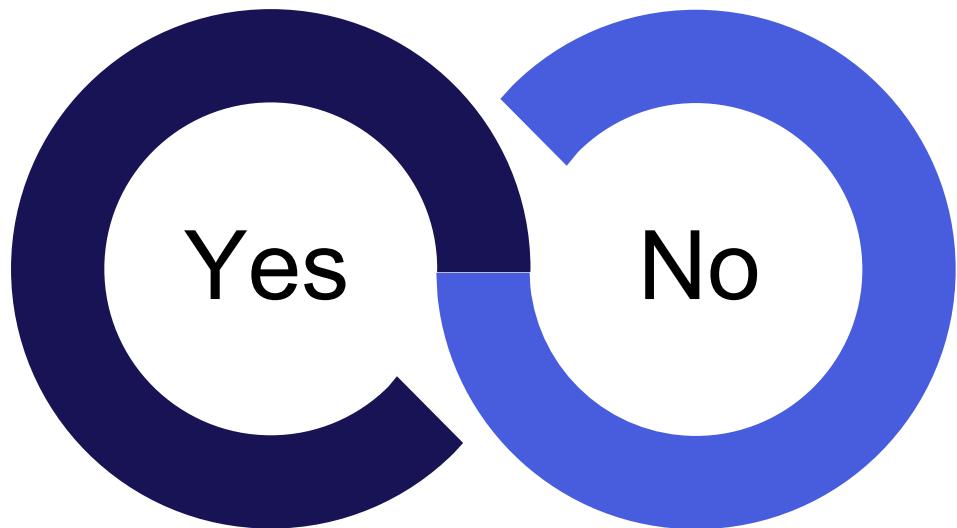
Number Of Construction Status-Room



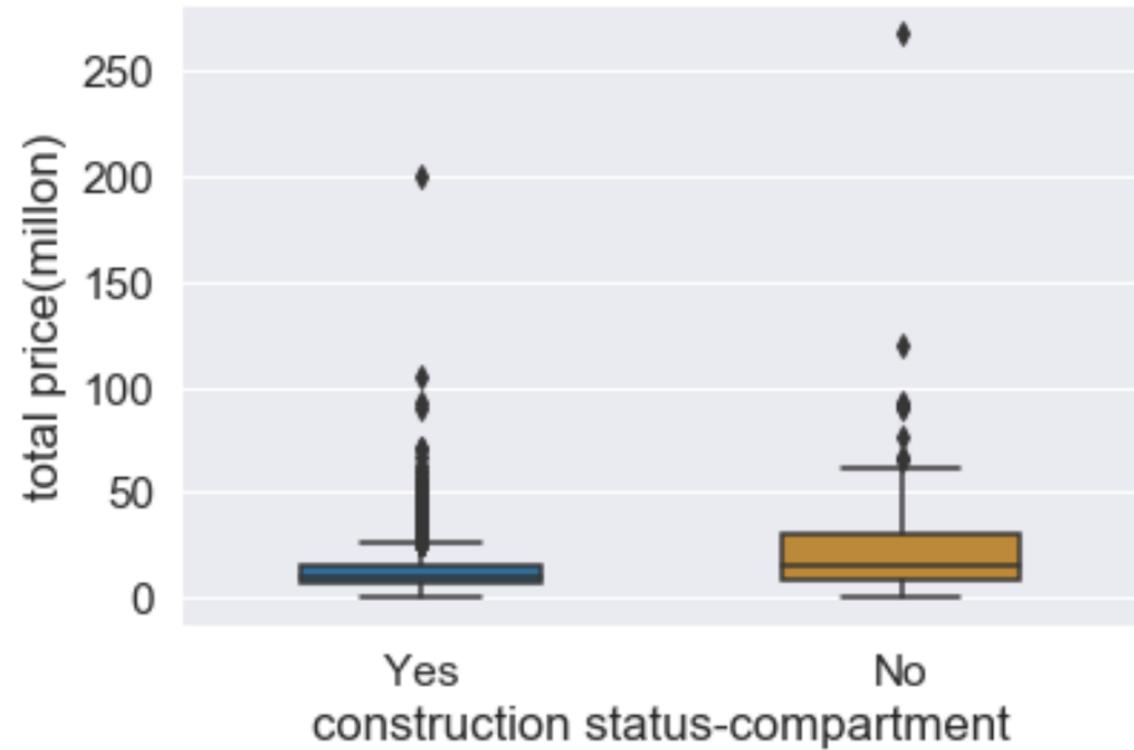
- If $2 < \text{the number of status-rooms} < 6$, the total price is high,

but some of the data do not have room also have high price.

— Construction Status-Compartment



- The 'No' in this variable has the higher price in Q3.





Feature Engineering

03

Feature Engineering

Number of transaction pens

House
Park
Ground

Transfer level

1,2,3,4,6

Construction status-compartment

Yes, No

Building type

Apartment
Mansion
Vendor
House

Main use

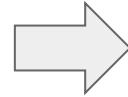
Business
Factory
Other
Living



Feature Engineering

(1) Dividing the ‘transaction pen’ into more detail columns.

Number of transaction pens
Ground3
Park2
House4

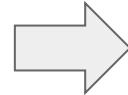


House	Ground	Park
4	3	2

Feature Engineering

(2) Count the ‘transfer level’.

transfer level
1,2,3,4,6
13,15,19,23
7,8

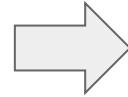


number of level
5
4
2

Feature Engineering

- (3) Change the category variable into numeric variable

construction status-compartment
Yes
No
No

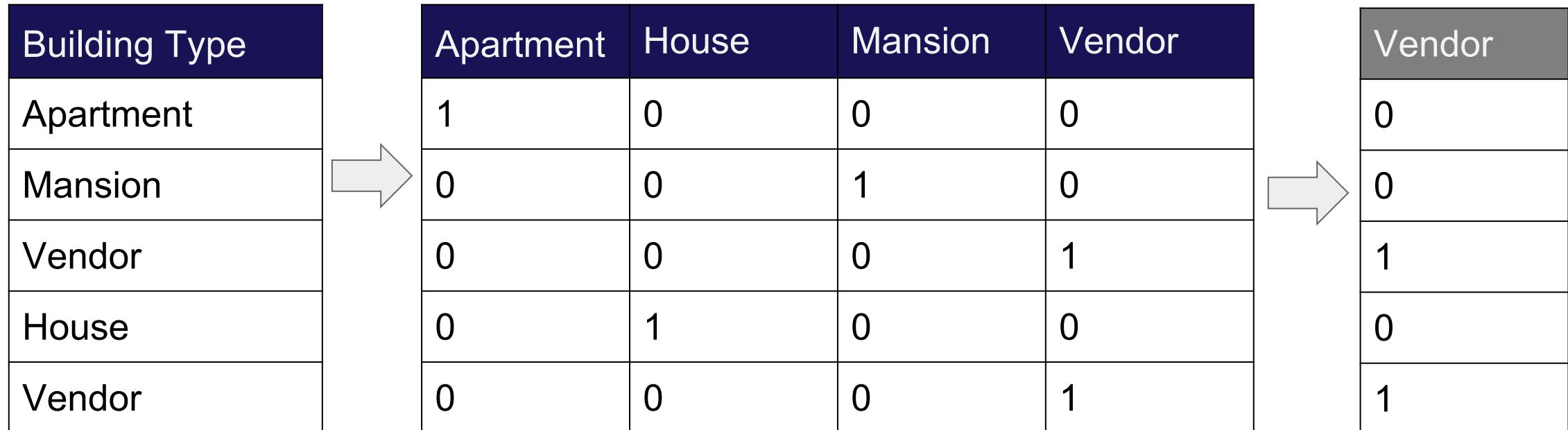


construction status-compartment
0
1
0

Feature Engineering

(4-1) Change the category variable into numeric variable

(4-2) Keep the column “vendor”



Feature Engineering

(5)-1 Change the catagory variable into numeric variable

(5)-2 Keep the column “other”

Main Use	Business	Factory	Living	Other	Other
Living	0	0	1	0	0
Factory	0	1	0	0	0
Other	0	0	0	1	1
Business	1	0	0	0	0
Other	0	0	0	1	1

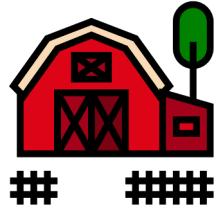


Regression Model

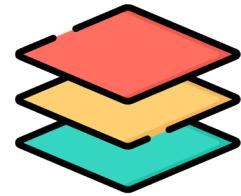
04

Regression Model

Predictors



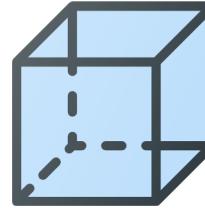
Total land transfer
area square meters



Total area of
building transfer
square meters



Ground



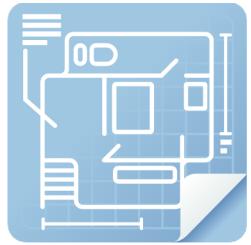
Construction status-
compartment



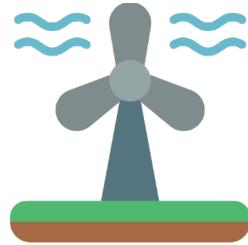
Number of levels

Regression Model

Predictors



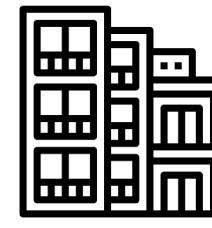
**Construction status-
room**



Main use (other)



Park



House



**Building type
(vendor)**

Regression Model

Response



01

Linear Regression

02

Random Forest Regression

03

XGBoost Regression

Regression Model

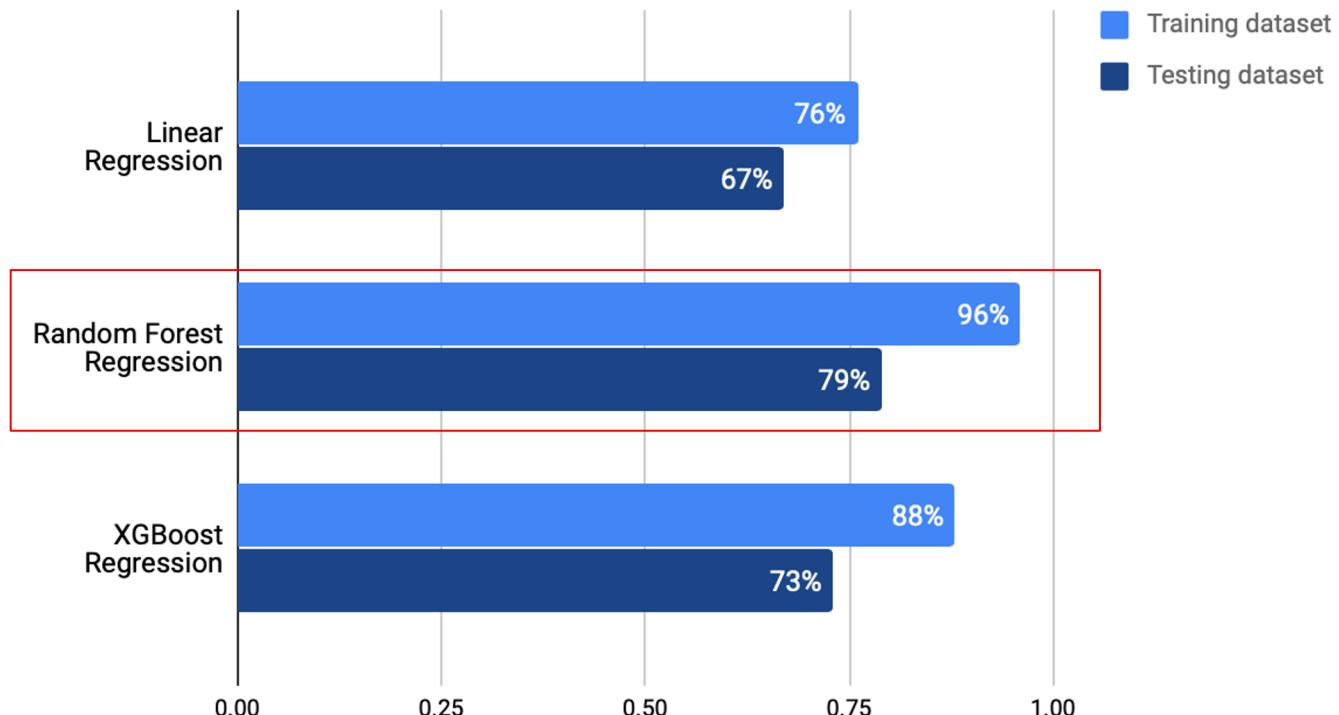
Split the dataset

Training dataset : 80%, Testing dataset : 20%.

Regression Model

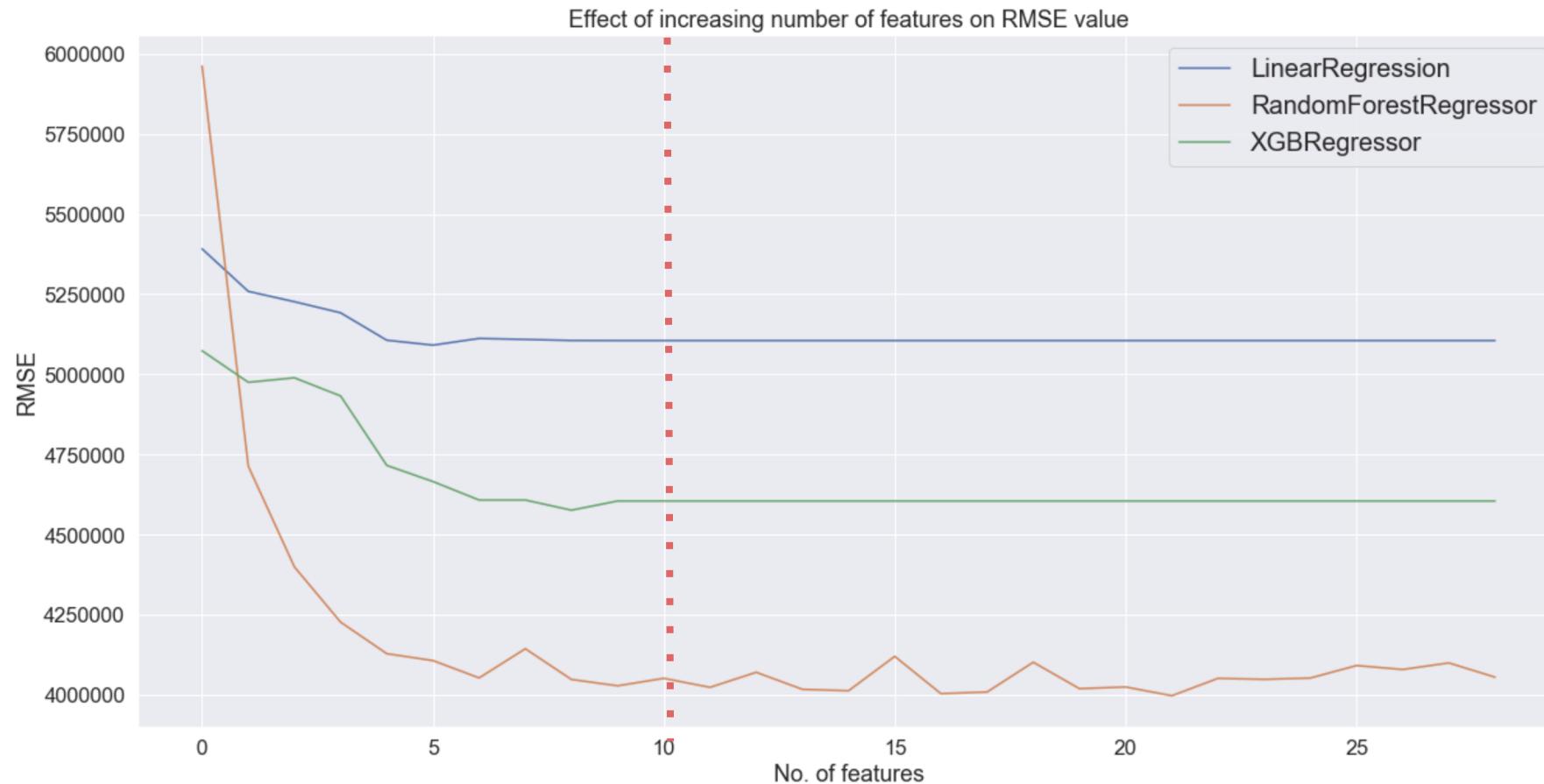
R square

R square



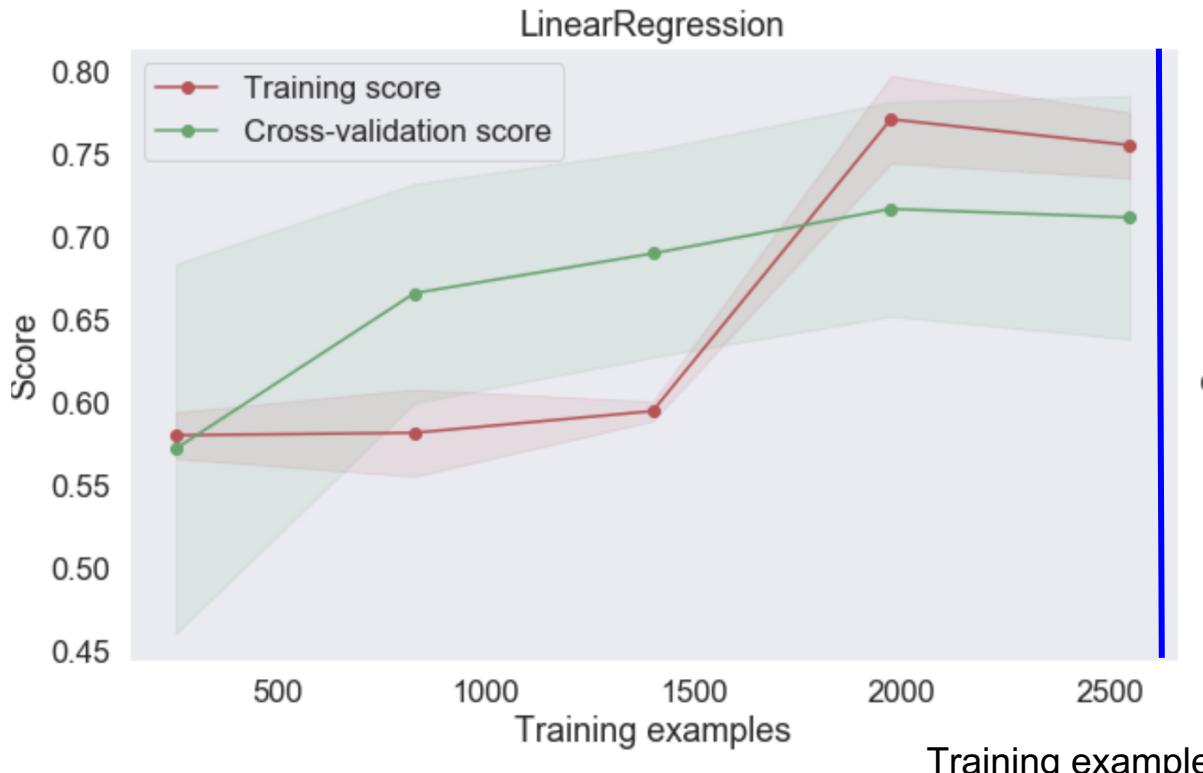
Regression Model

RMSE



Regression Model

Score(R square) with Training samples

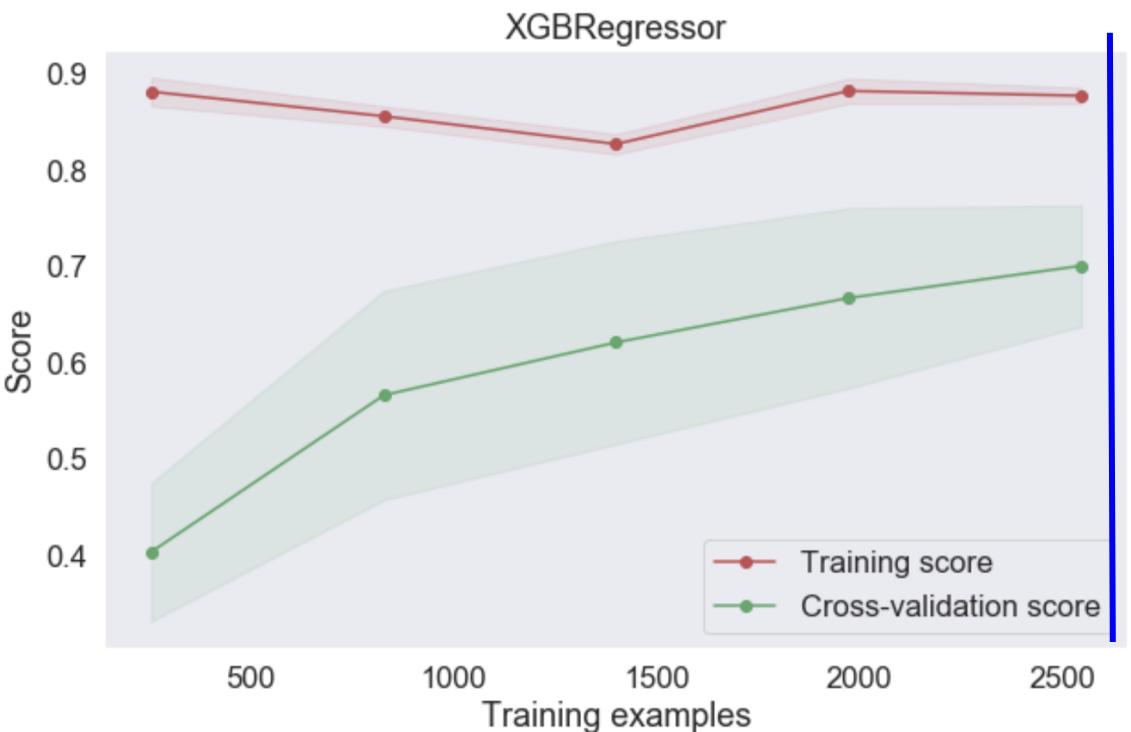


Regression Model

Score(R square) with Training samples



Training example:2900



Training example:2900 35



Conclusion

05

Conclusion

Flow Chart



What issue?

We cannot know how much budget we should prepare.



Measuring the budget

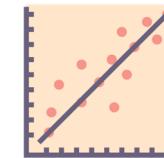
How much money to buy a ideal house.

Import recent dataset



Visualizing data

Find some interesting features in our dataset.



Present on reality

Predicting total price

R square 79%

We have 79% of the observed variation can be explained in this model.



**Thanks for your
watching and listening.**

CHIA-HAO HSU
Haoran Gao
Md Abu Osman