



# Analyzing house prices in District 7 HCMC using Linear Regression Model

**COURSE: FINANCIAL ECONOMETRICS**

Presented by Tran Thanh Dat (MAMAIU17036)

Tran Viet Hang (MAMAIU18079)

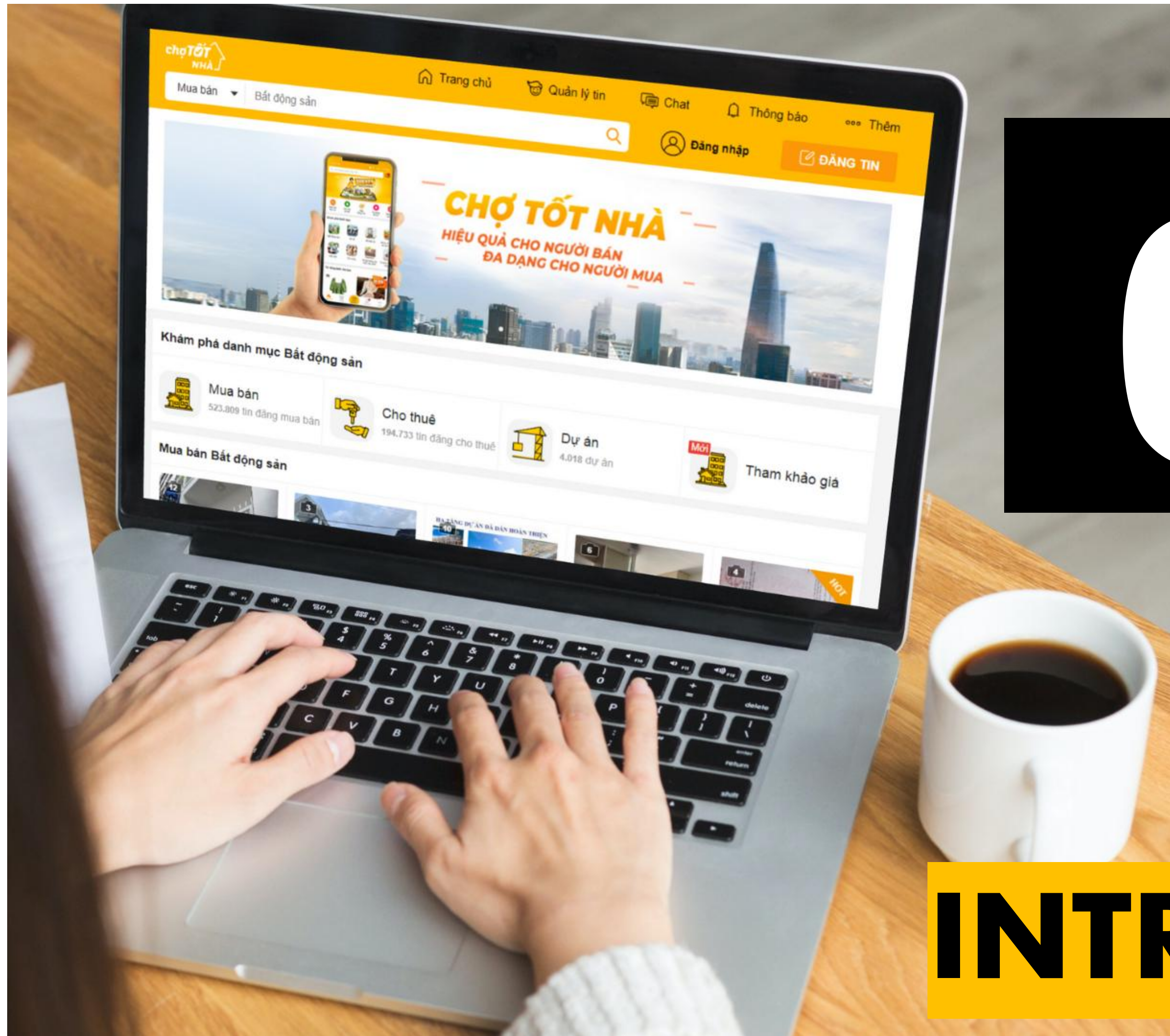
Instructed by Dr Nguyen Phuong Anh





# TABLE OF CONTENTS

1. INTRODUCTION
2. DATA OVERVIEW
3. MODEL AND ACCURACY
4. RESULTS AND FINDINGS



01

# INTRODUCTION



# 01. INTRODUCTION

- **Motivation:** Deciding to buy a house is more of a financial investment than a personal consumption decision, the research of which resulting in interesting conclusions



# 01. INTRODUCTION

- **Motivation:** Deciding to buy a house is more of a financial investment than a personal consumption decision, the research of which resulting in interesting conclusions
- **Research focus:** House price ranging from 18 billion to 28 billion VND in District 7, Ho Chi Minh City





# 01. INTRODUCTION

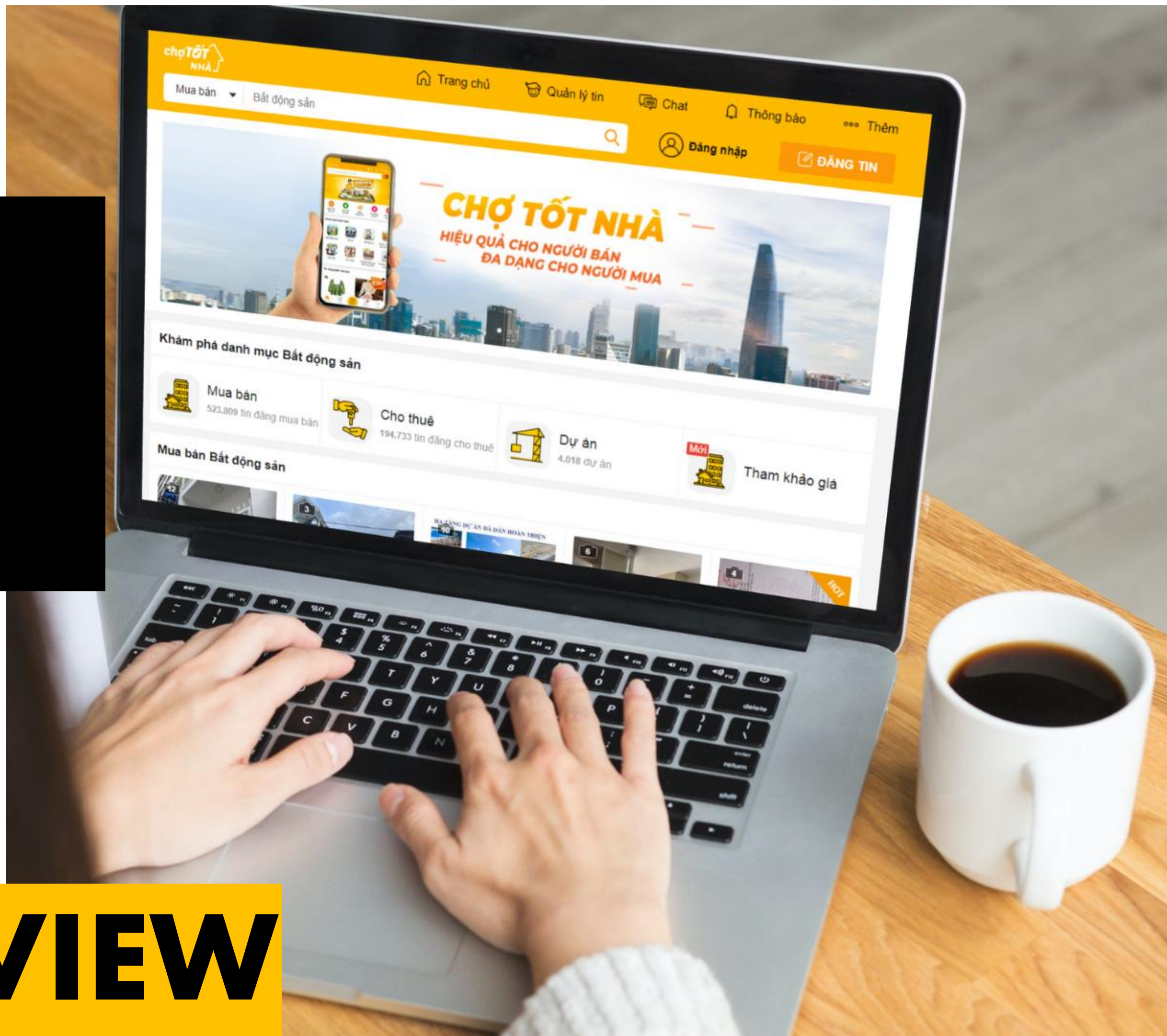
- **Motivation:** Deciding to buy a house is more of a financial investment than a personal consumption decision, the research of which resulting in interesting conclusions
- **Research focus:** House price ranging from 18 billion to 28 billion VND in District 7, Ho Chi Minh City
- **Research aims:** Identifying factors contribute to pricing houses using certain provided and collected data



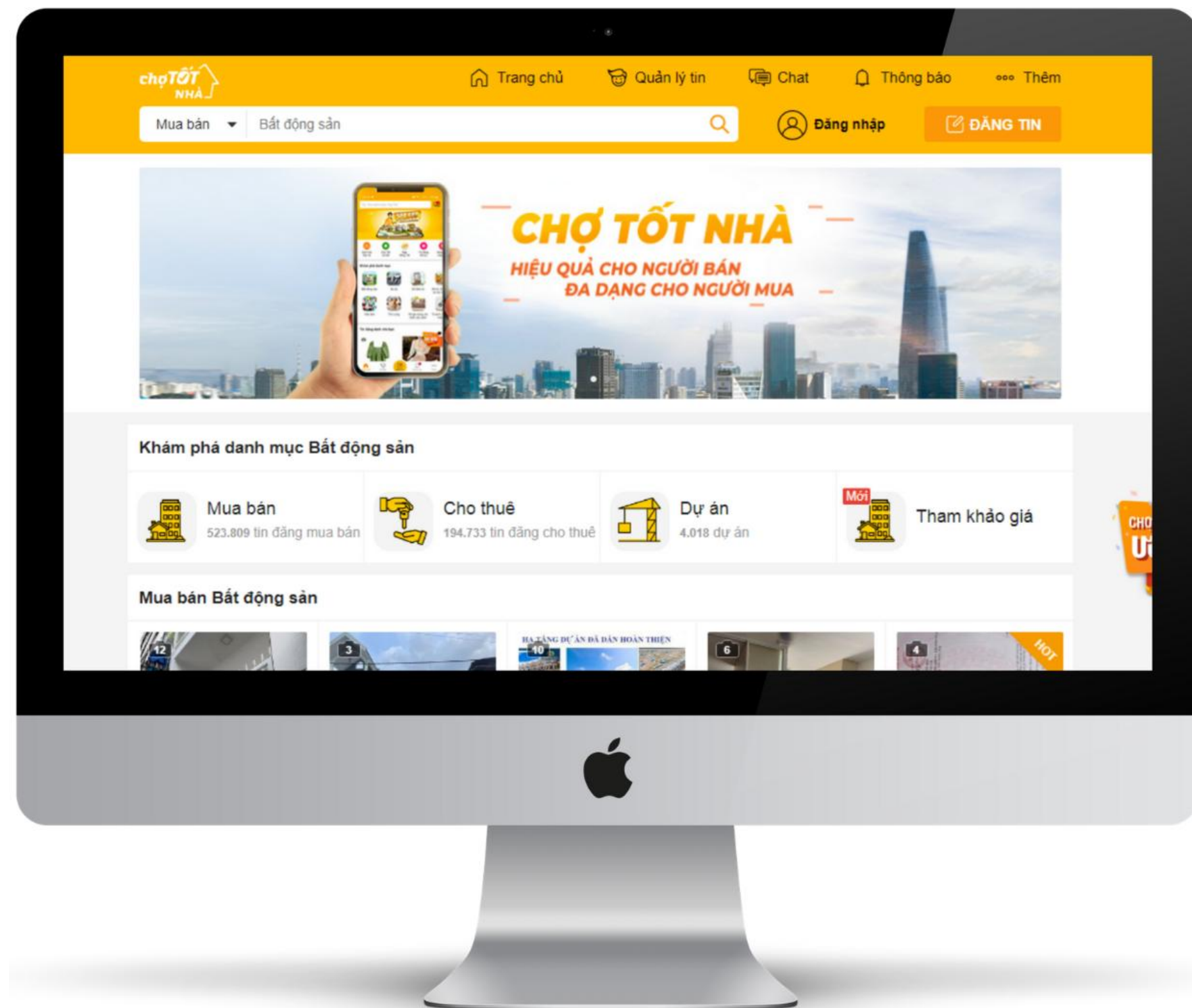


02

# DATA OVERVIEW



# 02. DATA OVERVIEW

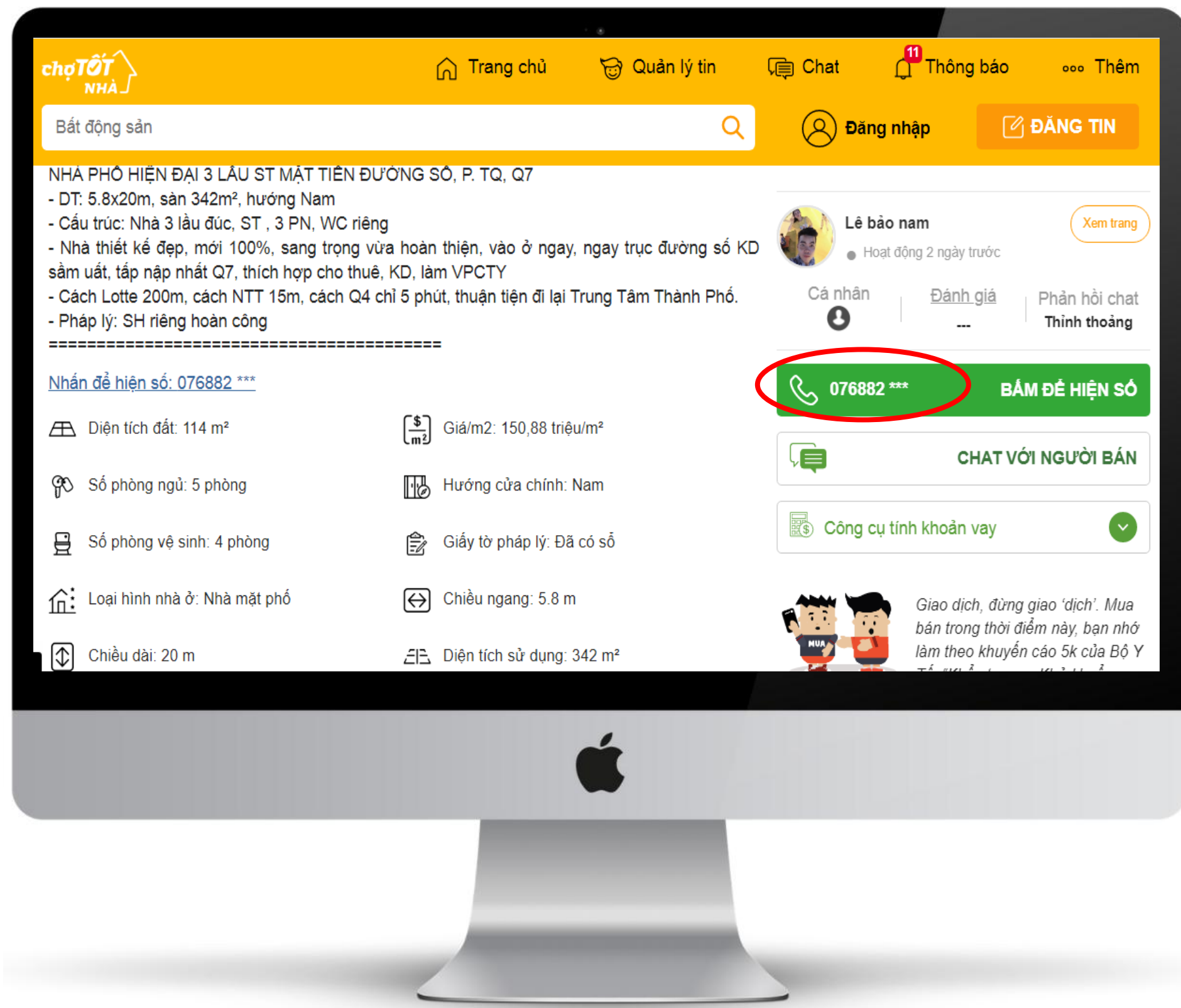


## 1. Sources

- Chotot.com



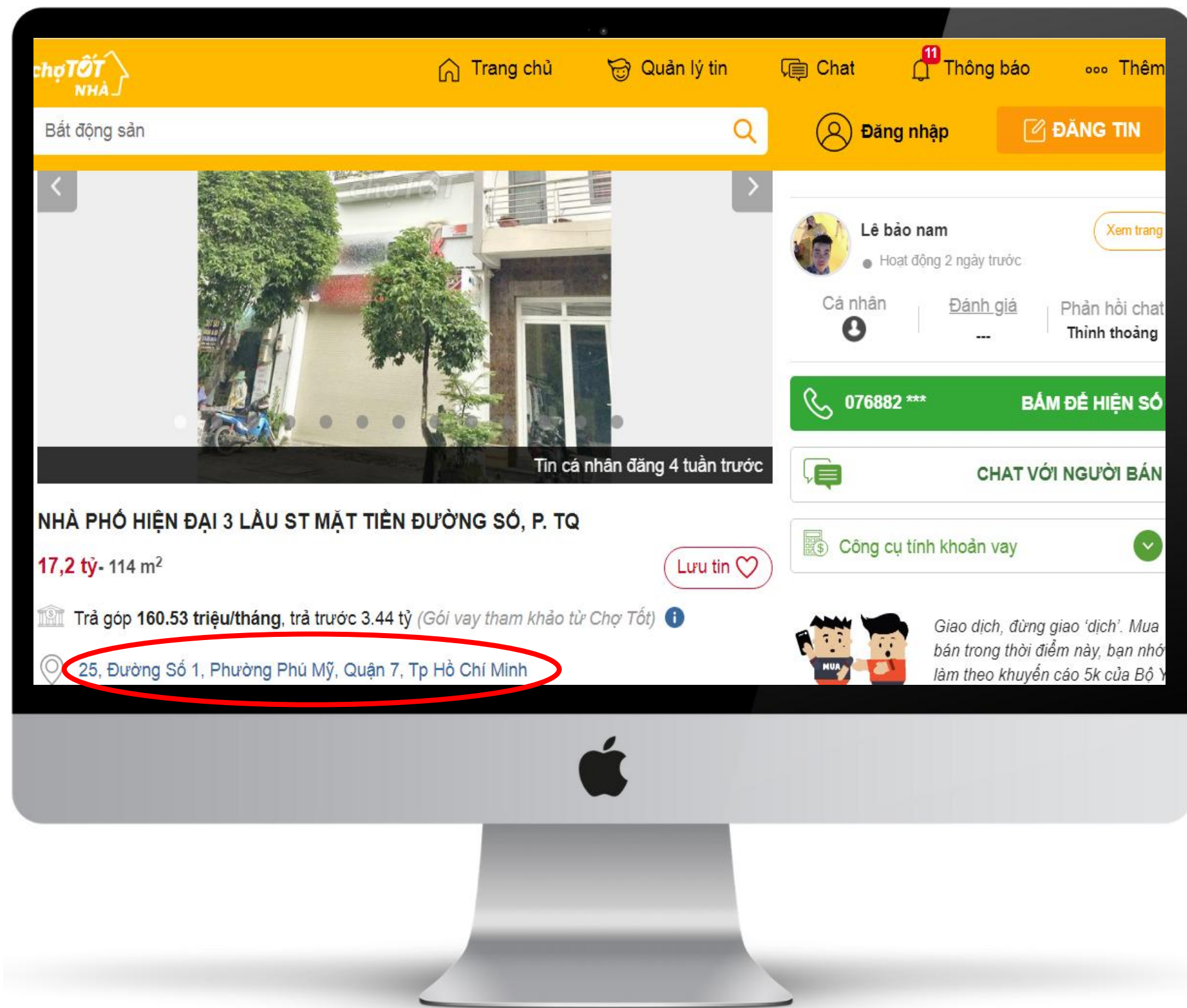
# 02. DATA OVERVIEW



## 1. Sources

- Chotot.com
- Phone number

# 02. DATA OVERVIEW

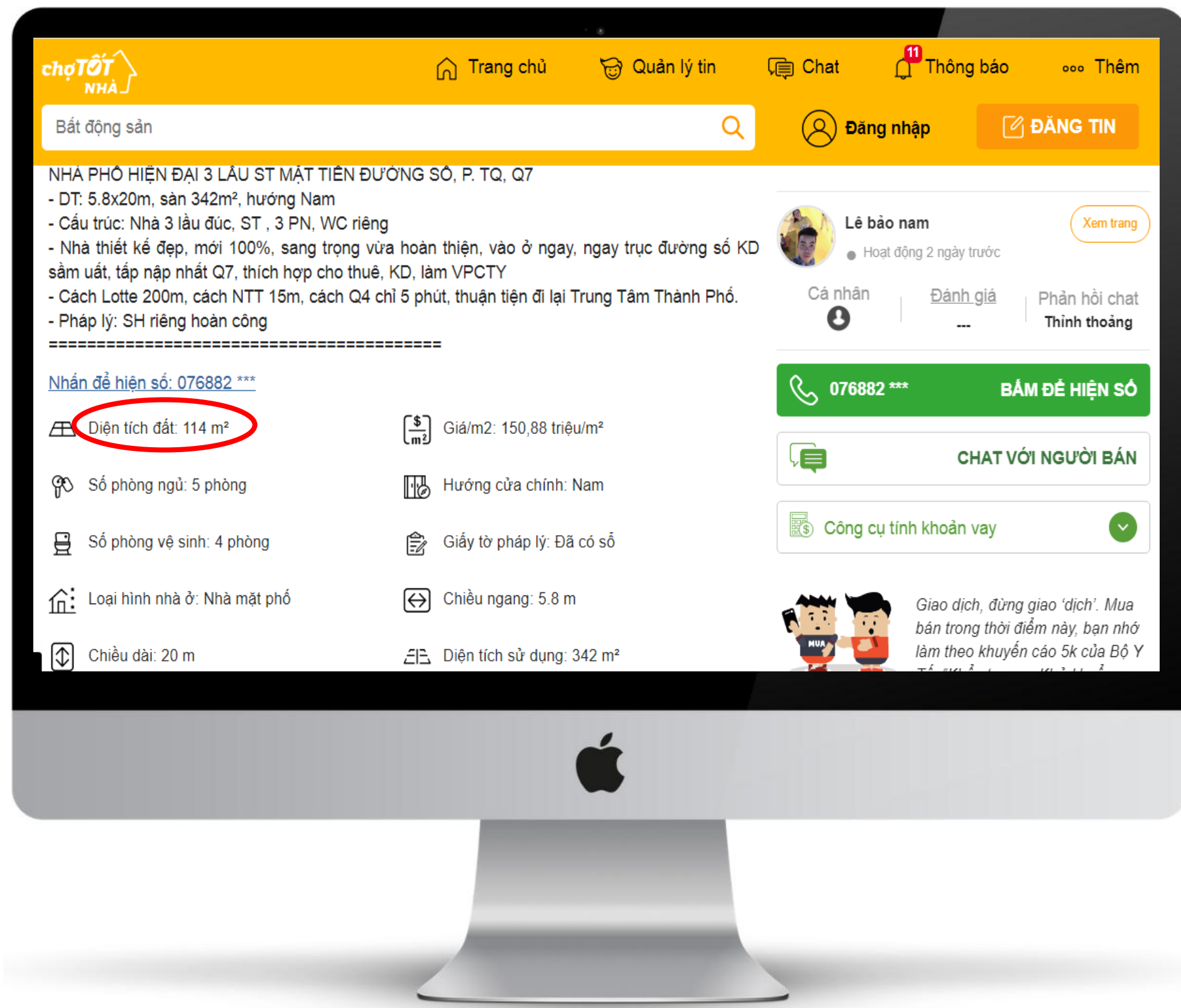


## 1. Sources

- Chotot.com
- Phone number
- Address



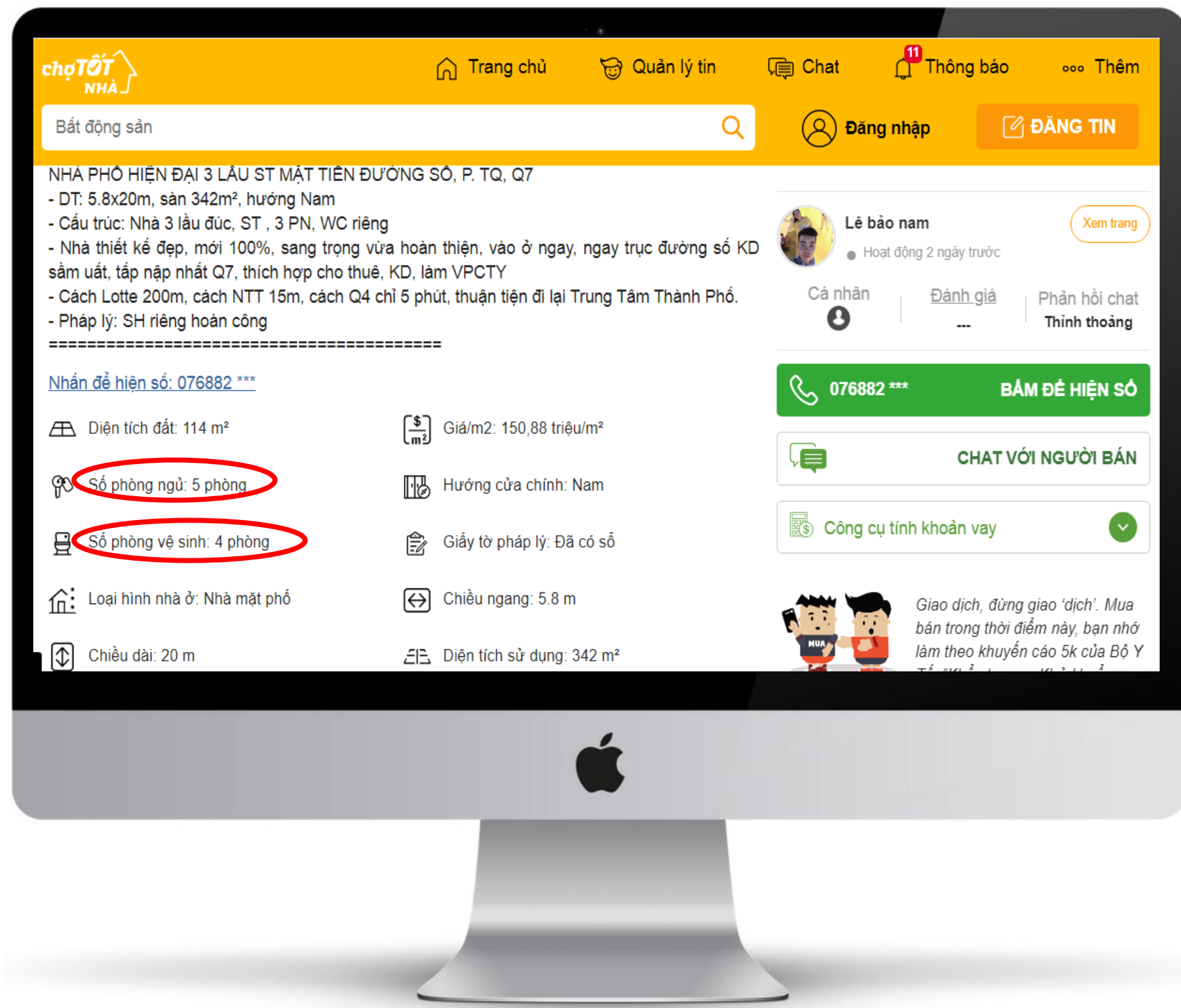
# 02. DATA OVERVIEW



## 1. Sources

- Chotot.com
- Phone number
- Address
- Area

# 02. DATA OVERVIEW

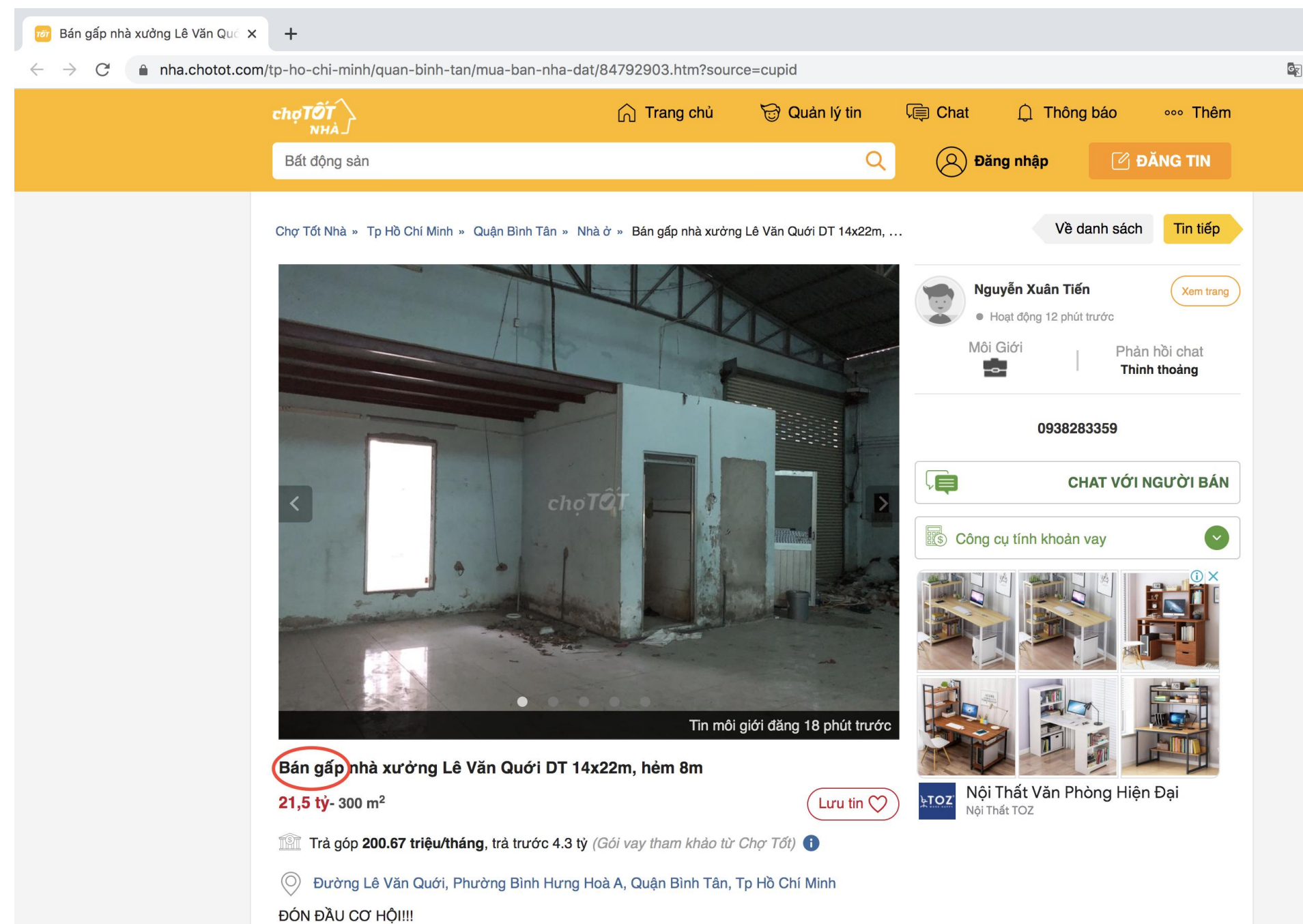


## 1. Sources

- Chotot.com
- Phone number
- Address
- Area
- Number of bedrooms
- Number of bathrooms



# 02. DATA OVERVIEW

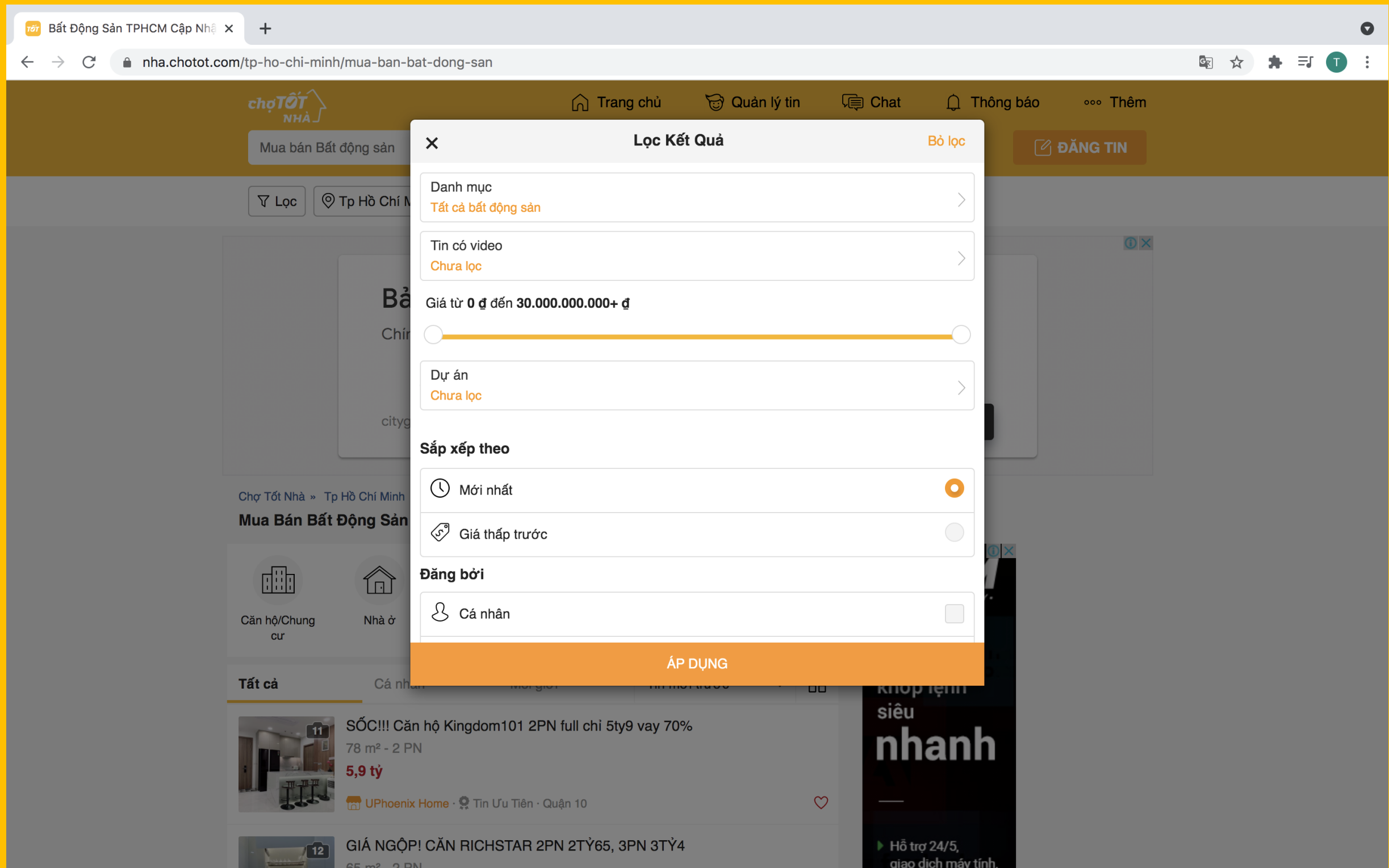


## 1. Sources

- Chotot.com
- Phone number
- Address
- Area: Size of the house
- Number of bedrooms
- Number of bathrooms
- Level of urgency that sellers want to sell

# 02. DATA OVERVIEW

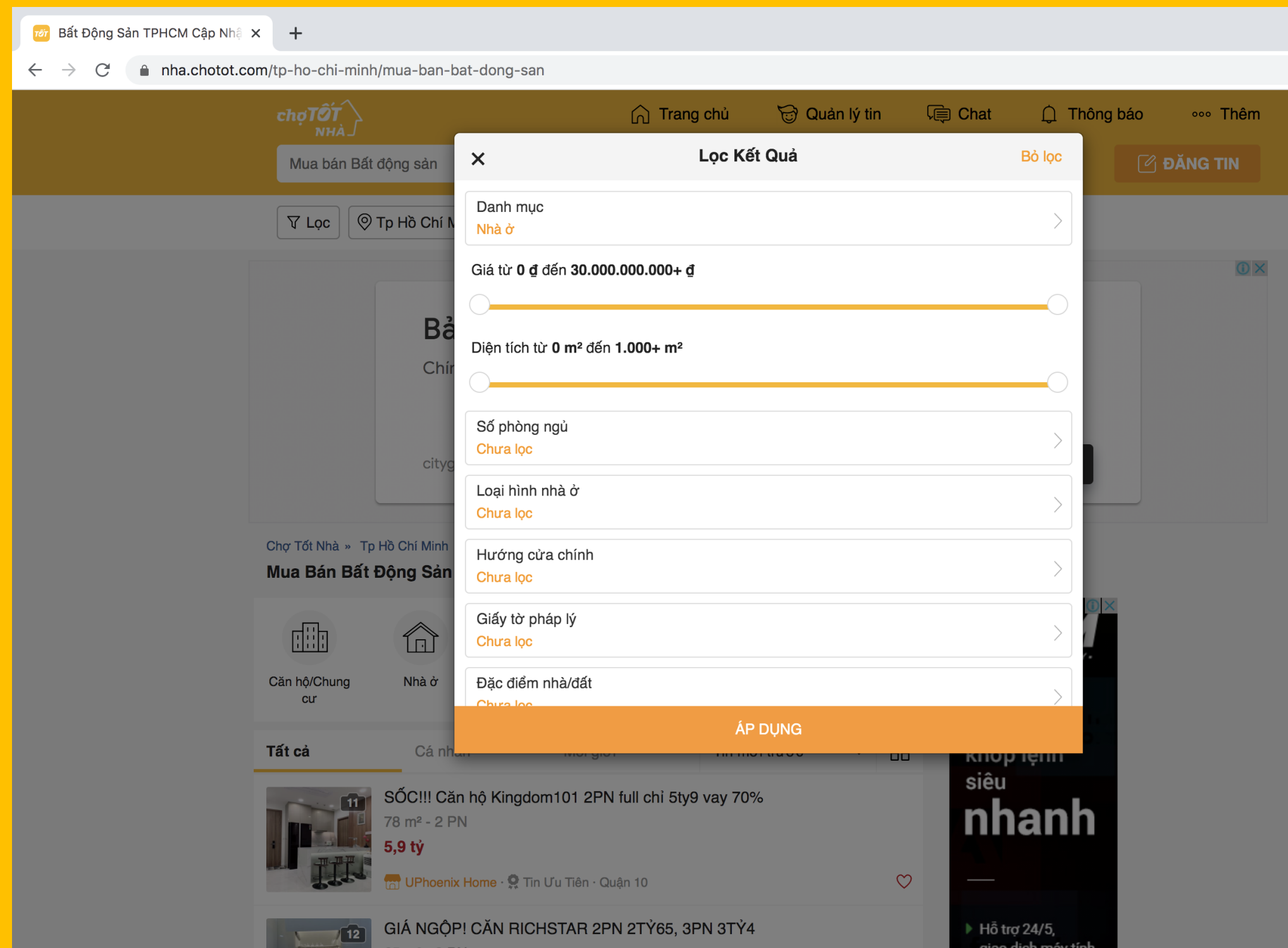
## 1. Sources





# 02. DATA OVERVIEW

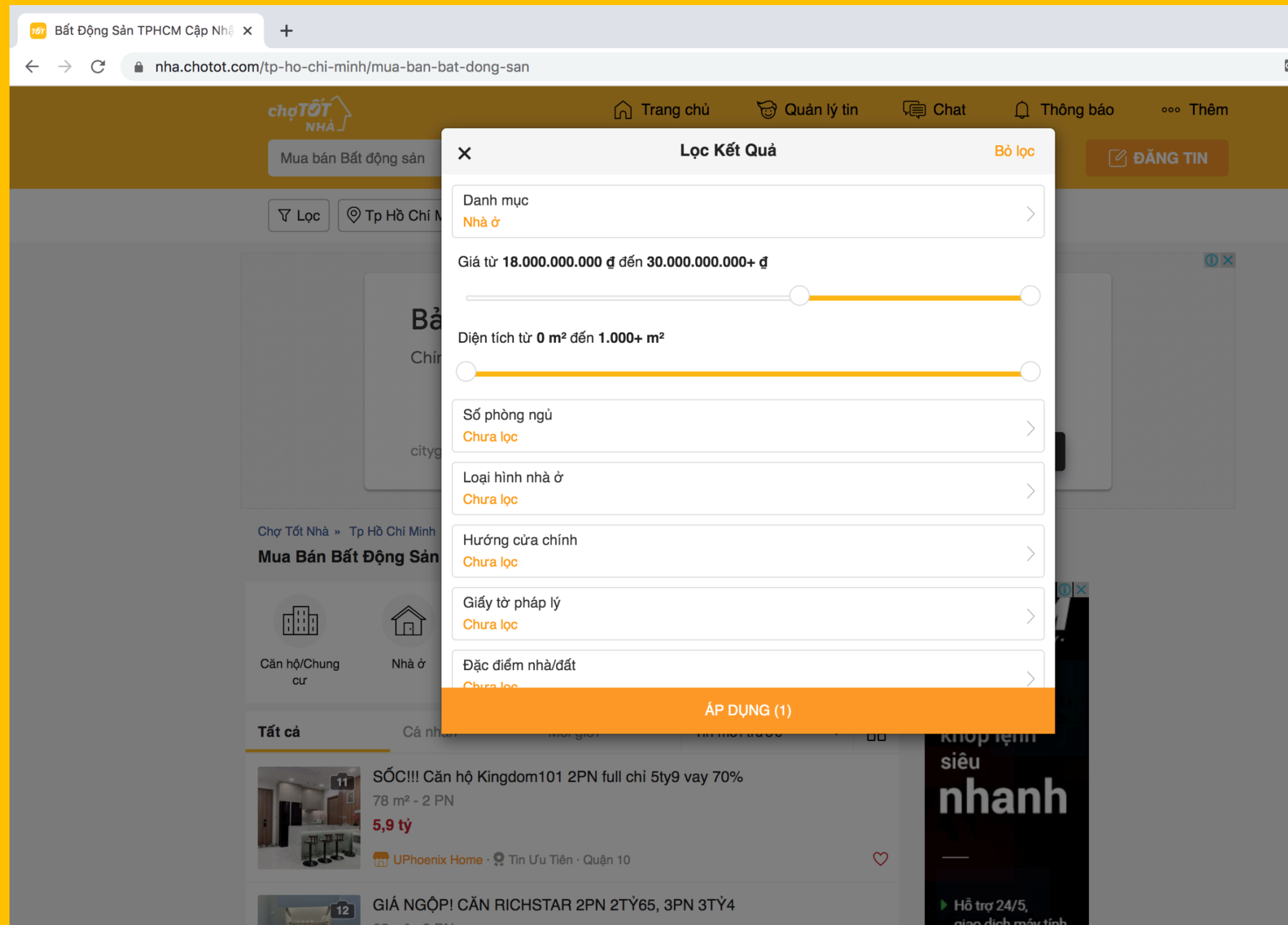
## 1. Sources



- House

# 02. DATA OVERVIEW

## 1. Sources



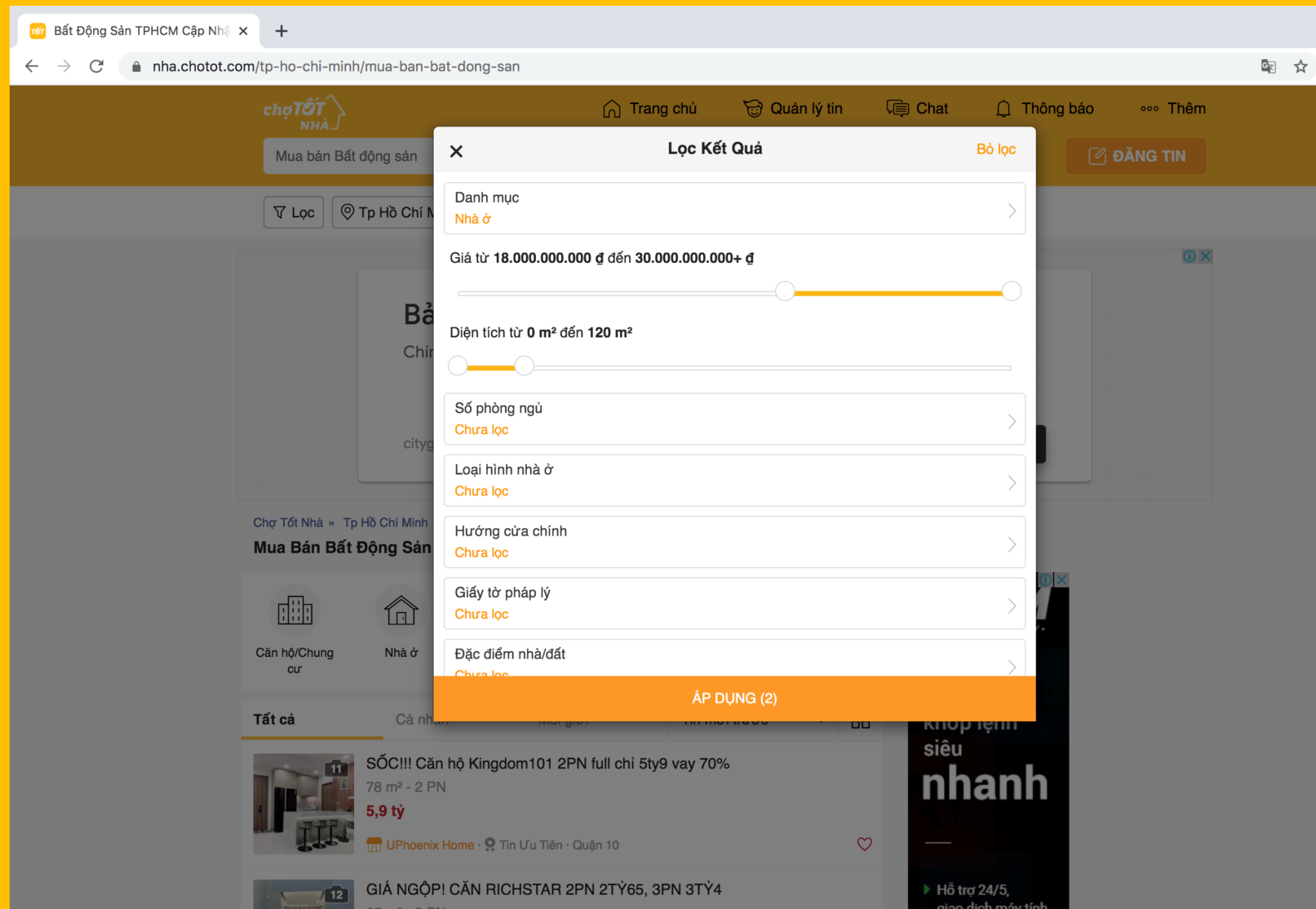
- House

- 18B – 30B



# 02. DATA OVERVIEW

## 1. Sources



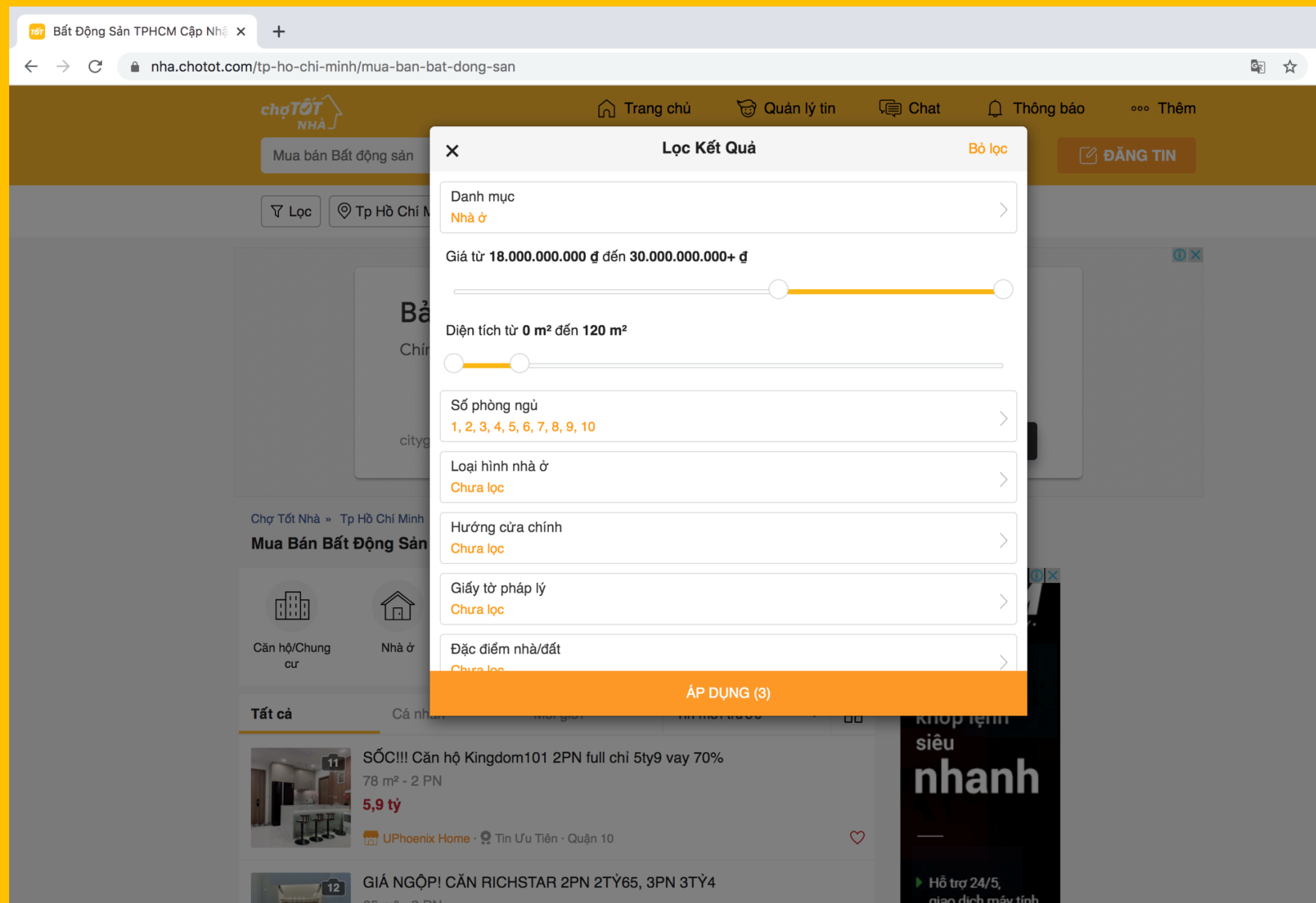
- House

- 18B – 30B

- Area 0 – 120m<sup>2</sup>

# 02. DATA OVERVIEW

## 1. Sources



- House
- 18B – 30B
- Area: 0 – 120m<sup>2</sup>
- Number of bedrooms

# 02. DATA OVERVIEW

## 2. Attributes

### Additional information

**WARDPRICE:** Price of each meter square of the house in a specific ward of District 7

**WARDPOP:** Population of the ward in District 7

	A	B	C	D	E	F	G	H	I	J
1	ID	Phone Number	Ward	ASAP	BEDRMS	BATHS	AREA	WARDPRICE	WARDPOP	PRICE
2	1	0936512261	Tân Thuận Tây	0	4	3	86	136,000,000	25,966	25,224,212,794
3	2	0938736912	Tân Phú	0	7	6	86	110,000,000	19,997	22,352,165,213
4	3	0913886619	Tân Thuận Đông	0	3	4	98	126,000,000	34,577	28,036,488,004
5	4	0904497992	Tân Thuận Đông	0	4	3	87	126,000,000	34,577	27,482,854,931
6	5	0907955123	Bình Thuận	0	9	7	98	132,000,000	30,949	21,076,485,160
7	6	0968026000	Tân Phong	0	10	8	75	132,000,000	12,221	17,547,743,632
8	7	0948303223	Tân Kiểng	0	8	6	92	111,000,000	26,535	21,709,250,747
9	8	0769887123	Tân Kiểng	1	7	5	85	111,000,000	26,535	22,305,063,636
10	9	0769887123	Tân Thuận Tây	0	9	7	72	136,000,000	25,966	20,771,634,183
11	10	0981662429	Bình Thuận	0	6	4	67	132,000,000	30,949	22,433,716,522
12	11	0909506698	Tân Phú	0	10	8	86	110,000,000	19,997	18,792,166,199
13	12	0909506698	Phú Thuận	0	5	6	98	120,000,000	26,042	23,019,444,076
14	13	0902501544	Tân Phong	0	9	7	91	132,000,000	12,221	20,561,302,803
15	14	0908686906	Phú Thuận	0	7	5	106	120,000,000	26,042	25,311,373,898
16	15	0908686906	Tân Quy	0	5	4	97	127,000,000	21,718	23,597,387,080
17	16	07777666610	Tân Thuận Đông	0	5	3	62	126,000,000	34,577	24,807,339,830
18	17	0931327929	Tân Quy	1	7	5	98	127,000,000	21,718	22,413,014,387
19	18	0931004612	Tân Hưng	0	5	4	86	115,000,000	31,572	24,664,631,936



# 02. DATA OVERVIEW

## 3. Basic Illustrations

Price: 17.3B – 28.6m

ASAP: Only ~ 10% yes

BEDRMS: 3 -10

BATHS: 2 -10


AREA: 45 – 123m<sup>2</sup>

WARDPRICE: 110m – 136m

WARDPOP: 12221 - 34577

```
> summary(Data)
```

ID	Phone Number	ward	ASAP
Min. : 1	Length:81	Length:81	Min. :0.0000
1st Qu.:21	Class :character	Class :character	1st Qu.:0.0000
Median :41	Mode :character	Mode :character	Median :0.0000
Mean :41			Mean :0.1111
3rd Qu.:61			3rd Qu.:0.0000
Max. :81			Max. :1.0000
BEDRMS	BATHS	AREA	PRICE
Min. : 3.000	Min. : 2.000	Min. : 45.00	Min. :1.686e+10
1st Qu.: 5.000	1st Qu.: 4.000	1st Qu.: 81.00	1st Qu.:2.077e+10
Median : 7.000	Median : 6.000	Median : 88.00	Median :2.243e+10
Mean : 6.778	Mean : 5.778	Mean : 88.63	Mean :2.246e+10
3rd Qu.: 9.000	3rd Qu.: 7.000	3rd Qu.: 97.00	3rd Qu.:2.443e+10
Max. :10.000	Max. :10.000	Max. :123.00	Max. :2.804e+10
WARDPRICE	WARDPOP		
Min. :110000000	Min. :12221		
1st Qu.:114000000	1st Qu.:19997		
Median :120000000	Median :25966		
Mean :121827160	Mean :24136		
3rd Qu.:132000000	3rd Qu.:30949		
Max. :136000000	Max. :34577		



# MODEL AND ACCURACY



03

# 03. MODEL AND ACCURACY

```
Residuals:
    Min       1Q   Median       3Q      Max
-1.987e+09 -6.081e+08 -1.196e+08  6.414e+08  2.338e+09

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.190e+10  1.926e+09  11.372  < 2e-16 ***
AREA         4.204e+07  7.689e+06   5.467  5.92e-07 ***
BEDRMS      -7.247e+08  9.708e+07  -7.465  1.32e-10 ***
BATHS       -2.246e+08  9.666e+07  -2.324  0.022899 *
WARDPOP      8.124e+04  2.243e+04   3.622  0.000533 ***
ASAP         1.971e+08  3.582e+08   0.550  0.583788
WARDPRICE    8.767e+00  1.233e+01   0.711  0.479404
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 974900000 on 74 degrees of freedom
Multiple R-squared:  0.858,    Adjusted R-squared:  0.8465
F-statistic: 74.52 on 6 and 74 DF,  p-value: < 2.2e-16
```

## 1. Test for statistical significance

P-value of AREA, BEDRMS, BATHS, WARDPOP are less than  $\alpha = 0.05$

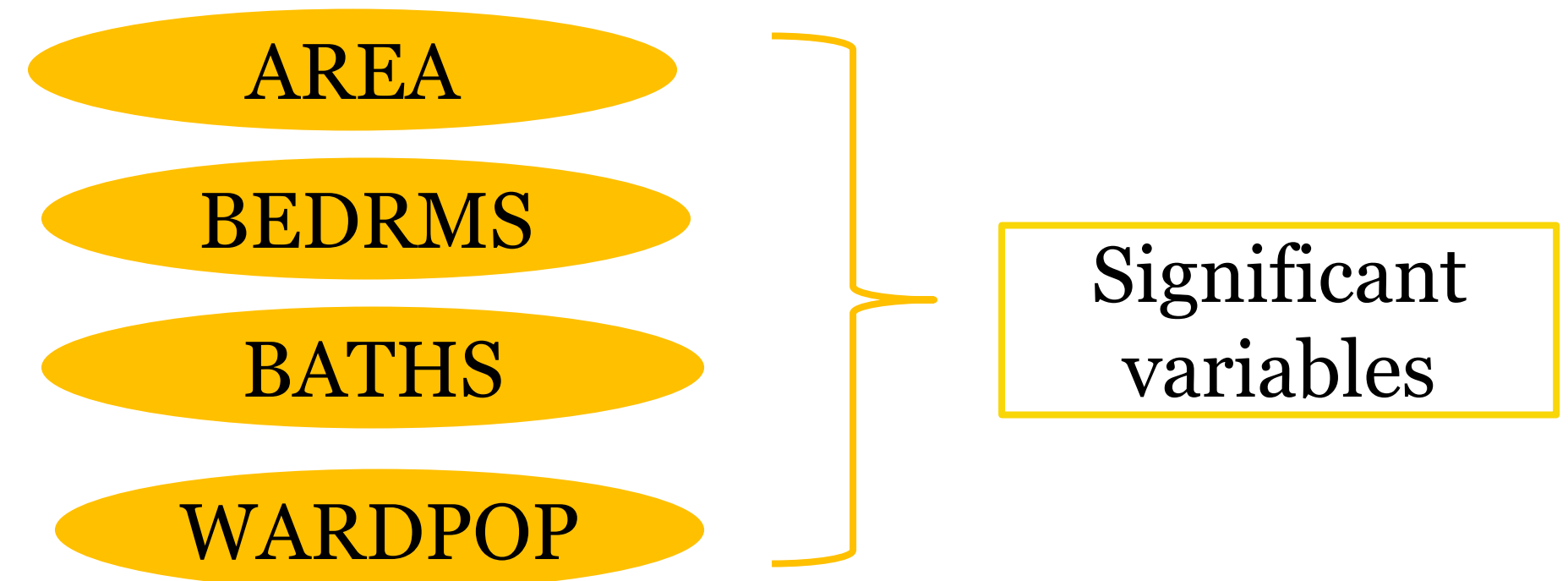


Figure 1: Multiple regression with six variables (*ASAP*, *AREA*, *BEDRMS*, *BATHS*, *WARDPRICE*, *WARDPOP*)



# 03. MODEL AND ACCURACY

```
Residuals:
    Min       1Q   Median       3Q      Max
-2.082e+09 -5.453e+08 -1.258e+08  6.603e+08  2.314e+09

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.303e+10  1.190e+09  19.349  < 2e-16 ***
AREA         4.136e+07  7.558e+06   5.473  5.48e-07 ***
BEDRMS      -7.184e+08  9.499e+07  -7.563  7.52e-11 ***
BATHS       -2.319e+08  9.298e+07  -2.494  0.014786 *
WARDPOP      8.192e+04  2.186e+04   3.748  0.000345 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 967200000 on 76 degrees of freedom
Multiple R-squared:  0.8565,    Adjusted R-squared:  0.8489
F-statistic: 113.4 on 4 and 76 DF,  p-value: < 2.2e-16
```

## 2. Test for heteroscedasticity

```
studentized Breusch-Pagan test

data:  ReturnReg
BP = 18.989, df = 19, p-value = 0.4576
```

$P\text{-value} = 0.4576 > \alpha = 0.05$

No heteroscedasticity exists

Figure 2: Multiple regression with 4 significant variables (*AREA*, *BEDRMS*, *BATHS*, *WARDPOP*)

# 03. MODEL AND ACCURACY

## 3. Test for multicollinearity

	AREA	BEDRMS	BATHS	WARDPOP
AREA	1.0000000	0.1214834	0.1829282	-0.1356163
BEDRMS	0.1214834	1.0000000	0.7832158	-0.7184542
BATHS	0.1829282	0.7832158	1.0000000	-0.6467865
WARDPOP	-0.1356163	-0.7184542	-0.6467865	1.0000000

Figure 3: Correlations of 4 significant variables (*AREA*, *BEDRMS*, *BATHS*, *WARDPOP*)

VIF of 4 variables are all below 5

```
> vif(ReturnReg)
      AREA  BEDRMS  BATHS  WARDPOP
1.037913  3.240699  2.735727  2.152160
```

No multicollinearity exists

Figure 4: VIF of 4 significant variables (*AREA*, *BEDRMS*, *BATHS*, *WARDPOP*)

# 03. MODEL AND ACCURACY

## 4. Test for assumption of normality

Jarque Bera Test

```
data: ReturnReg$residuals  
X-squared = 0.82274, df = 2, p-value = 0.6627
```

$JB = 0.82274$  and  $p\text{-value} = 0.6627 > 0.05$



The data is normally distributed

Figure 5: Jarque – Bera test conducted on R



# 03. MODEL AND ACCURACY

## 4. Test for autocorrelation

$$DW = 2.1817$$

Durbin-Watson test

```
data: ReturnReg  
DW = 2.1817, p-value = 0.7817  
alternative hypothesis: true autocorrelation is greater than 0
```

$k=4$  (number of variables),  $n=81$  (number of observations),  $d_L = 1.534$ ,  $d_U = 1.743$   
(Durbin-Watson table for  $\alpha = 0.05$ )

$$\rightarrow d_U < DW = 2.1817 < 4 - d_U$$

Figure 6: Durbin-Watson test conducted on R

Have positive autocorrelation	Inconclusive	No autocorrelation	Inconclusive	Have negative autocorrelation
0	$d_L$	$d_U$	$4 - d_U$	$4 - d_L$
		2		4



No autocorrelation

# 03. MODEL AND ACCURACY

## 2. Regression Model

Residuals:

Min	1Q	Median	3Q	Max
-2.082e+09	-5.453e+08	-1.258e+08	6.603e+08	2.314e+09

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	2.303e+10	1.190e+09	19.349	< 2e-16	***
AREA	4.136e+07	7.558e+06	5.473	5.48e-07	***
BEDRMS	-7.184e+08	9.499e+07	-7.563	7.52e-11	***
BATHS	-2.319e+08	9.298e+07	-2.494	0.014786	*
WARDPOP	8.192e+04	2.186e+04	3.748	0.000345	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 967200000 on 76 degrees of freedom  
Multiple R-squared: 0.8565, Adjusted R-squared: 0.8489  
F-statistic: 113.4 on 4 and 76 DF, p-value: < 2.2e-16

•BEDRMS Coef. =  $-7.184 \times 10^8$

•AREA Coef. =  $4.136 \times 10^7$

•BATHS Coef. =  $-2.319 \times 10^8$

•WARDPOP Coef. =  $8.192 \times 10^4$

• Intercept =  $2.303 \times 10^{10}$

# 03. MODEL AND ACCURACY

## Multiple regression model for house price in District 7

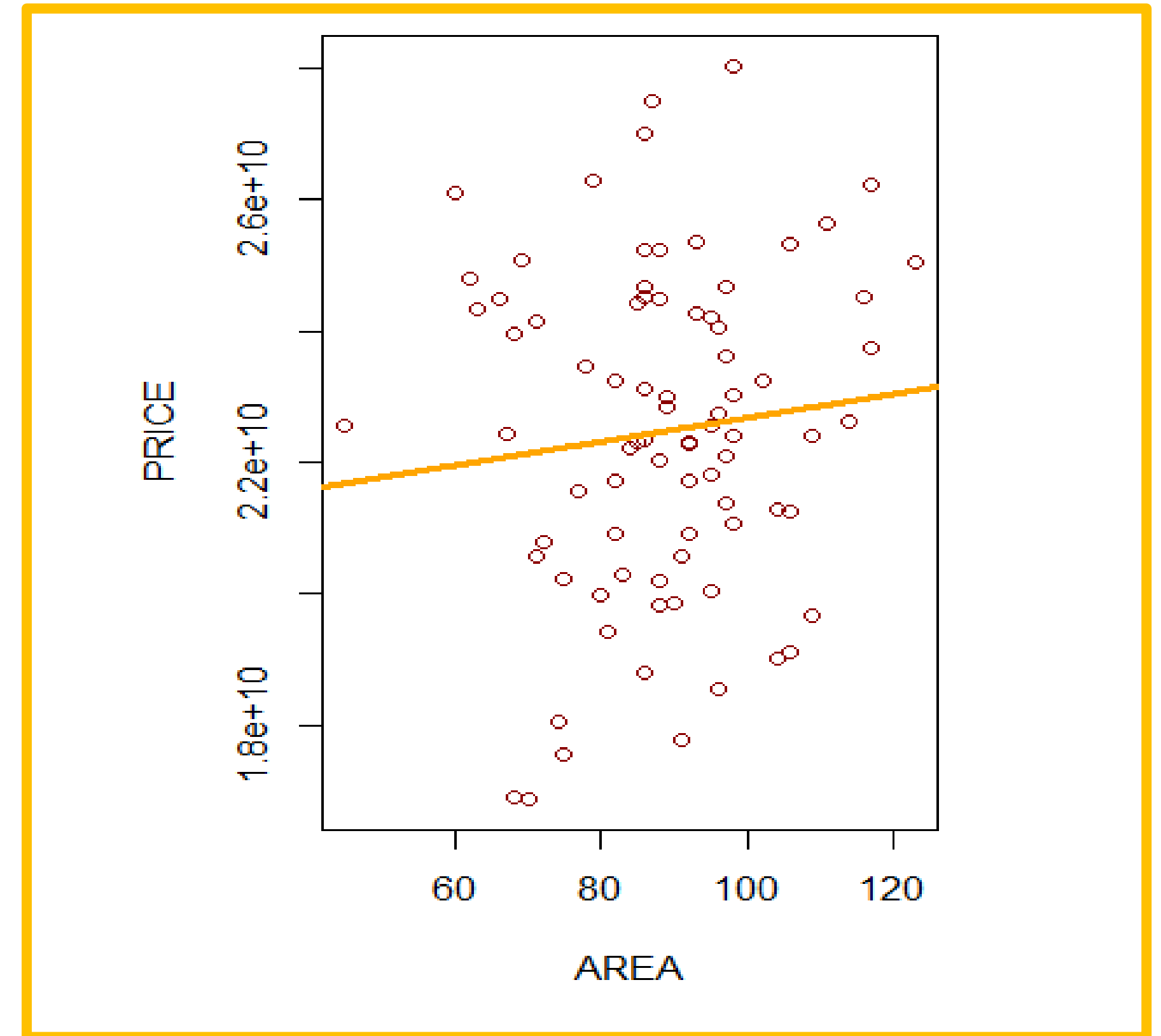
Phone Number	Ward	ASAP	BEDRMS	BATHS	AREA	PRICE	WARDPRICE	WARDPOP
0936512261	Tân Thuận Tây	0	4	3	86	25224212794	1.36e+08	25966
0938736912	Tân Phú	0	7	6	86	22352165213	1.10e+08	19997
0913886619	Tân Thuận Đông	0	3	4	98	28036488004	1.26e+08	34577
0904497992	Tân Thuận Đông	0	4	3	87	27482854931	1.26e+08	34577
0907955123	Bình Thuận	0	9	7	98	21076485160	1.32e+08	30949
0968026000	Tân Phong	0	10	8	75	17547743632	1.32e+08	12221
0948303223	Tân Kiểng	0	8	6	92	21709250747	1.11e+08	26535
0769887123	Tân Kiểng	1	7	5	85	22305063636	1.11e+08	26535
0769887123	Tân Thuận Tây	0	9	7	72	20771634183	1.36e+08	25966
0981662429	Bình Thuận	0	6	4	67	22433716522	1.32e+08	30949
0909506698	Tân Phú	0	10	8	86	18792166199	1.10e+08	19997

$$PRICE = 2.303 \times 10^{10} + 4.136 \times 10^7 AREA - 7.184 \times 10^8 BEDRMS - 2.319 \times 10^8 BATHS + 8.192 \times 10^4 WARDPOP$$



# 4. RESULTS AND FINDINGS

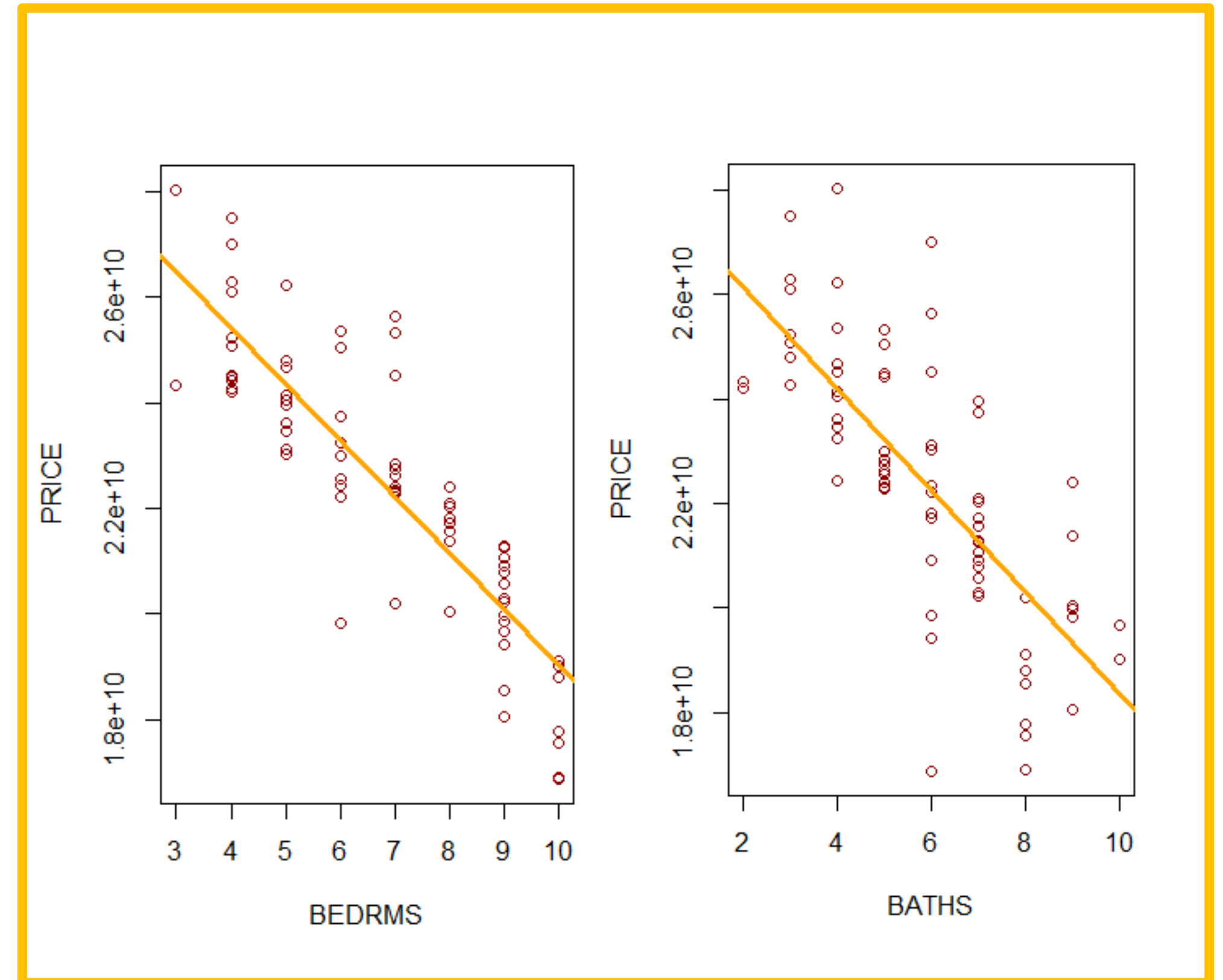
AREA and PRICE have a positive relationship:  
The higher the size of the house is, the higher its price is.



# 4. RESULTS AND FINDINGS

Assume that we do not change the area and number of bathrooms, if we add one more bedroom, on average, the price of the house will decrease by \$0.7184B. The problem is that house's owners divide an unchangeable area into multiple bedrooms. As a result, each bedroom will become narrower and buyers do not like doing so. They will be willing to pay a lower price to buy that house. Likewise, if the number of bathrooms is increased by one while the number of bedrooms and house's area remain the same, the price of the house will decrease by \$0.2319B

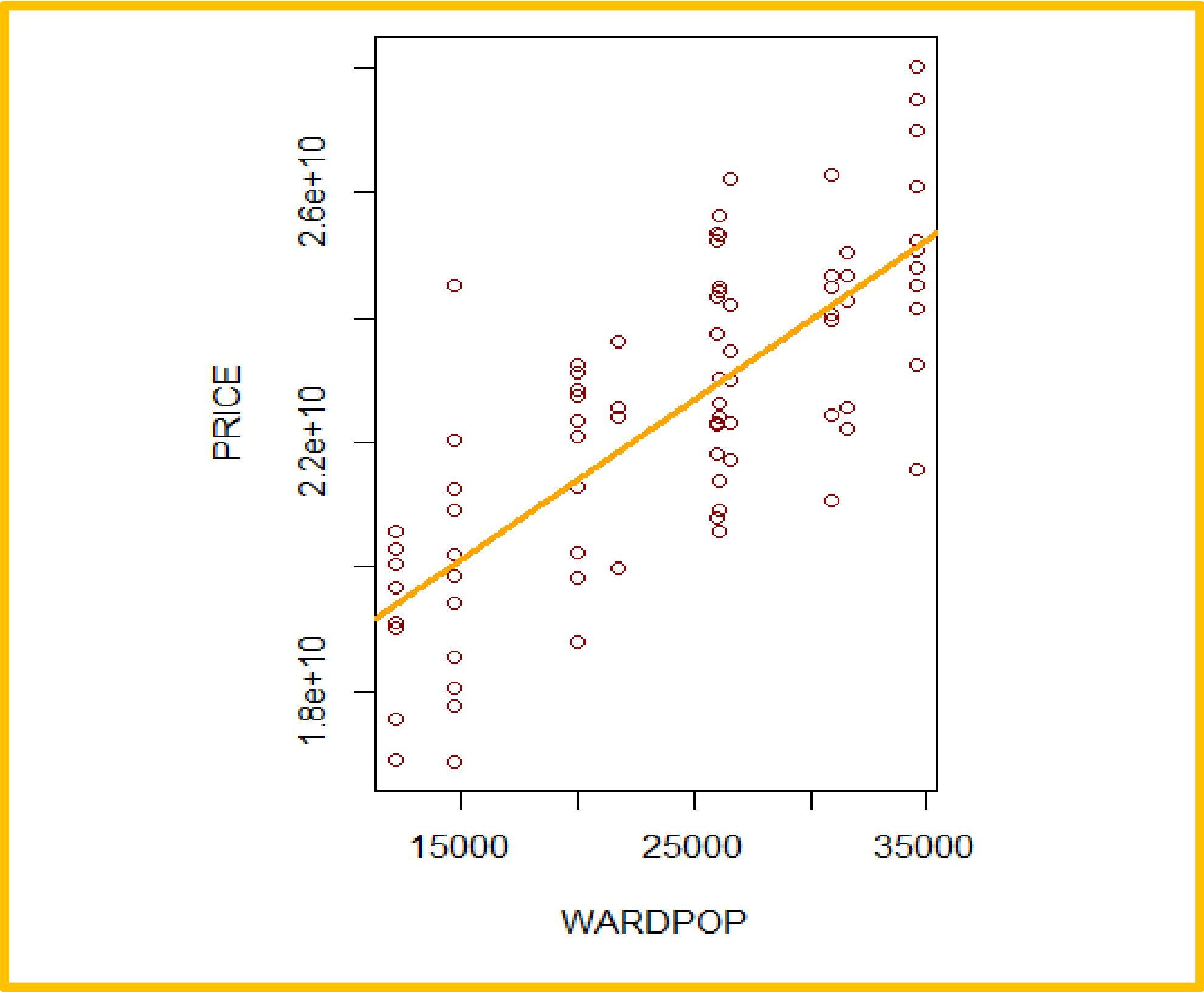
→ BEDRMS and BATHS have a negative relationship with PRICE



# 4. RESULTS AND FINDINGS

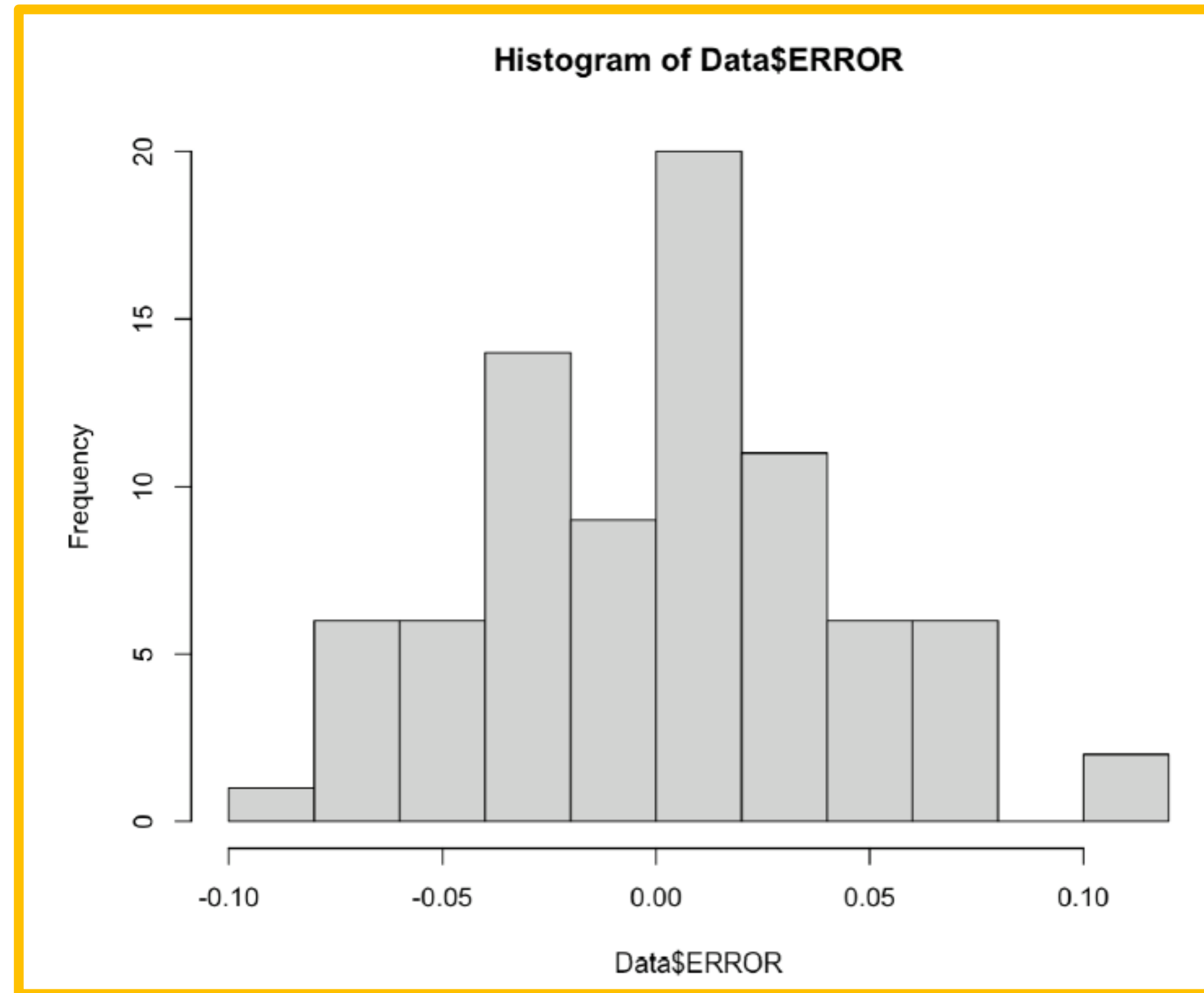


WARDPOP has a positive relationship with PRICE





# 4. RESULTS AND FINDINGS







Chợ Tốt

MUỐN LÀ CÓ

THANK YOU  
FOR LISTENING!

