

УДК 004.82:004:85

*Ярош О. Л., студент 2 курсу
спеціальності 122 «Комп'ютерні науки»
Нескородєва Т. В., к.т.н., доцент, доцент
кафедри інформаційних технологій*

АНАЛІЗ ДАНИХ ПРО РИНОК ЖИТЛА В УКРАЇНІ ЗАСОБАМИ МОВИ R

Донецький національний університет імені Василя Стуса, м. Вінниця

На сьогоднішній день майже кожна людина збирає кошти на власну нерухомість або вже її має. При виборі житла люди зазвичай дивляться на головний критерій – ціну. Тому актуальним є питання, як зробити правильний вибір та при цьому заощадити кошти? Щоб дати відповідь на це питання ми провели дослідження на основі даного набору даних.

Аналіз проведемо за допомогою пакету R. Для дослідження був використаний набір даних Real Estate Price Prediction, що містить 414 спостережень по 8 показниках:

- [1] "No" – номер,
- [2] "X1.transaction.date" – дата транзакції,
- [3] "X2.house.age" – вік будинку,
- [4] "X3.distance.to.the.nearest.MRT.station" – дистанція до найближчого мед. пункту,
- [5] "X4.number.of.convenience.stores" – кількість магазинів поблизу,
- [6] "X5.latitude" – ширина,
- [7] "X6.longitude" – довжина,
- [8] "Y.house.price.of.unit.area" – ціна за квадратний метр.

Побудувавши графік залежності ціни за квадратний метр від відстані до найближчого медпункту та від кількості магазинів поблизу (рис. 1 та рис. 2) ми зробили висновок, що ціна в більшій мірі залежить від інфраструктури навколо. Наявність магазинів, аптек та лікарні поблизу напряму впливає на ціну нерухомості.

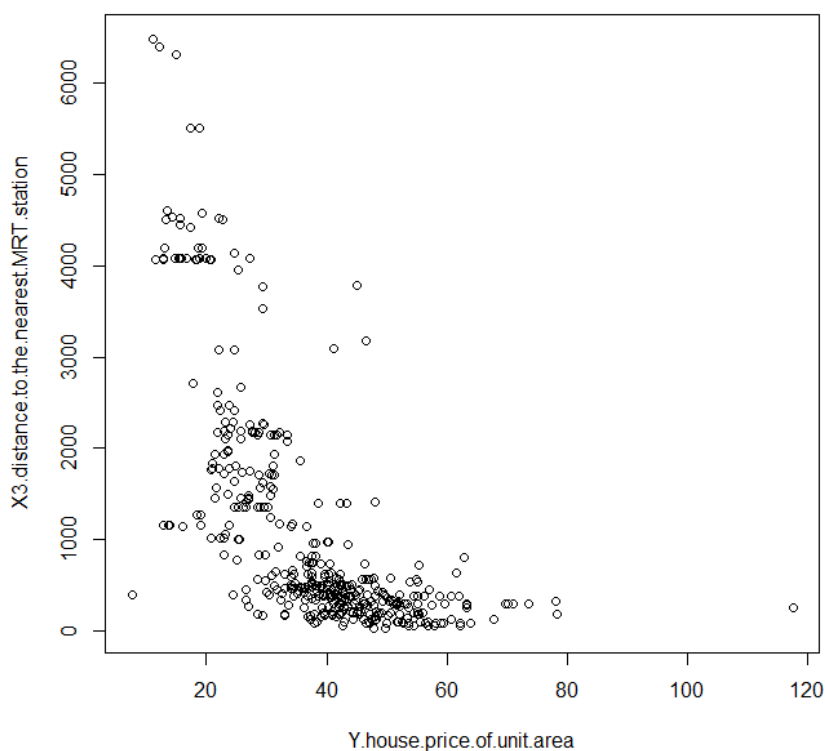


Рисунок 1 – Залежність ціни за квадратний метр від відстані до найближчого медпункту

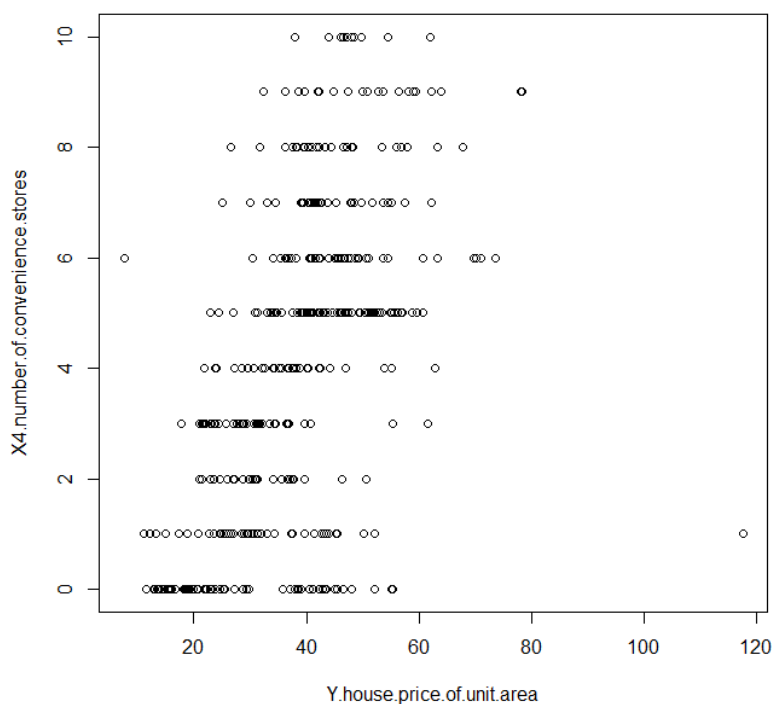


Рисунок 2 – Залежність ціни за квадратний метр від кількості магазинів поблизу

Щоб з'ясувати варіацію ціни Y в залежності від відстані, побудуємо лінійну регресію, в якій предикторами будуть виступати $X1.transaction.date$, $X2.house.age$, $X3.distance.to.the.nearest.MRT.station$, $X4.number.of.convenience.stores$, $X5.latitude$, $X6.longitude$.

```

Call:
lm(formula = Y.house.price.of.unit.area ~ X1.transaction.date +
    X2.house.age + X3.distance.to.the.nearest.MRT.station + X4.number.of.convenience.stores +
    X5.latitude + X6.longitude)

Residuals:
    Min       1Q   Median       3Q      Max
-35.664  -5.410  -0.966   4.217  75.193

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -1.444e+04  6.776e+03  -2.131  0.03371 *
X1.transaction.date  5.146e+00  1.557e+00   3.305  0.00103 **
X2.house.age      -2.697e-01  3.853e-02  -7.000  1.06e-11 ***
X3.distance.to.the.nearest.MRT.station -4.488e-03  7.180e-04  -6.250  1.04e-09 ***
X4.number.of.convenience.stores  1.133e+00  1.882e-01   6.023  3.84e-09 ***
X5.latitude       2.255e+02  4.457e+01   5.059  6.38e-07 ***
X6.longitude     -1.242e+01  4.858e+01  -0.256  0.79829
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.858 on 407 degrees of freedom
Multiple R-squared:  0.5824,    Adjusted R-squared:  0.5762
F-statistic: 94.59 on 6 and 407 DF,  p-value: < 2.2e-16

```

Рисунок 3 – Звіт регресії

Стандартне значення помилки коливається від 45.8514271 до 0.6526105. Остаточка стандартна помилка 8.858 на 407. Скоректований $R^2 = 0.5762$ і множинний $R^2 = 0.5824$. Після даного висновку стає зрозумілим, що ціна на будинок в більшій степені залежить від інфраструктури, яка включає в себе X3.distance.to.the.nearest.MRT.station та X4.number.of.convenience.stores, та від розміру будинку який залежить від X5.latitude та X6.longitude. За допомогою цих даних ми можемо без проблем проаналізувати ринок нерухомості та скорегувати ціни, відібрати найвигідніші варіанти. Можемо передбачити вартість потрібного будинку.

Список літератури

1. Набір даних *Real Estate Price Prediction* [Електронний ресурс]. Режим доступу – <https://www.kaggle.com/quantbruce/real-estate-price-prediction>
2. Джеймс Г., Уиттон Д., Тибишани Р. Введение в статистическое обучение с примерами R. Изд. Второе, испр. Пер с англ. С.Э. Мاستицкого – М. ДМК Пресс, 2017. -456с.