

數據分析報告

影響高雄地區房價因素之分析

Chia-Wei Chen

Jun 16, 2016

研究動機與目的

▶ 研究動機

目前絕大部分房價數據分析著重於台北市及新北市，但高雄市因工業及科技迅速發展也導致房價上升。人民在衡量房價時，也會同時考慮許多層面，因此影響房價的因素更顯得重要。

▶ 目的

利用高雄市實價登錄資料，來找出影響房價最顯著因子，提供房地產購買者或投資者做為參考。

資料

▶ 資料取得

來源：由內政部不動產交易實價查詢服務網取得高雄市交易內容。

時間：104/1~104/12

<http://plvr.land.moi.gov.tw/DownloadOpenData>

▶ 資料型態

共15794筆資料(取70%作為訓練集,30%為測試集)

▶ 應變數

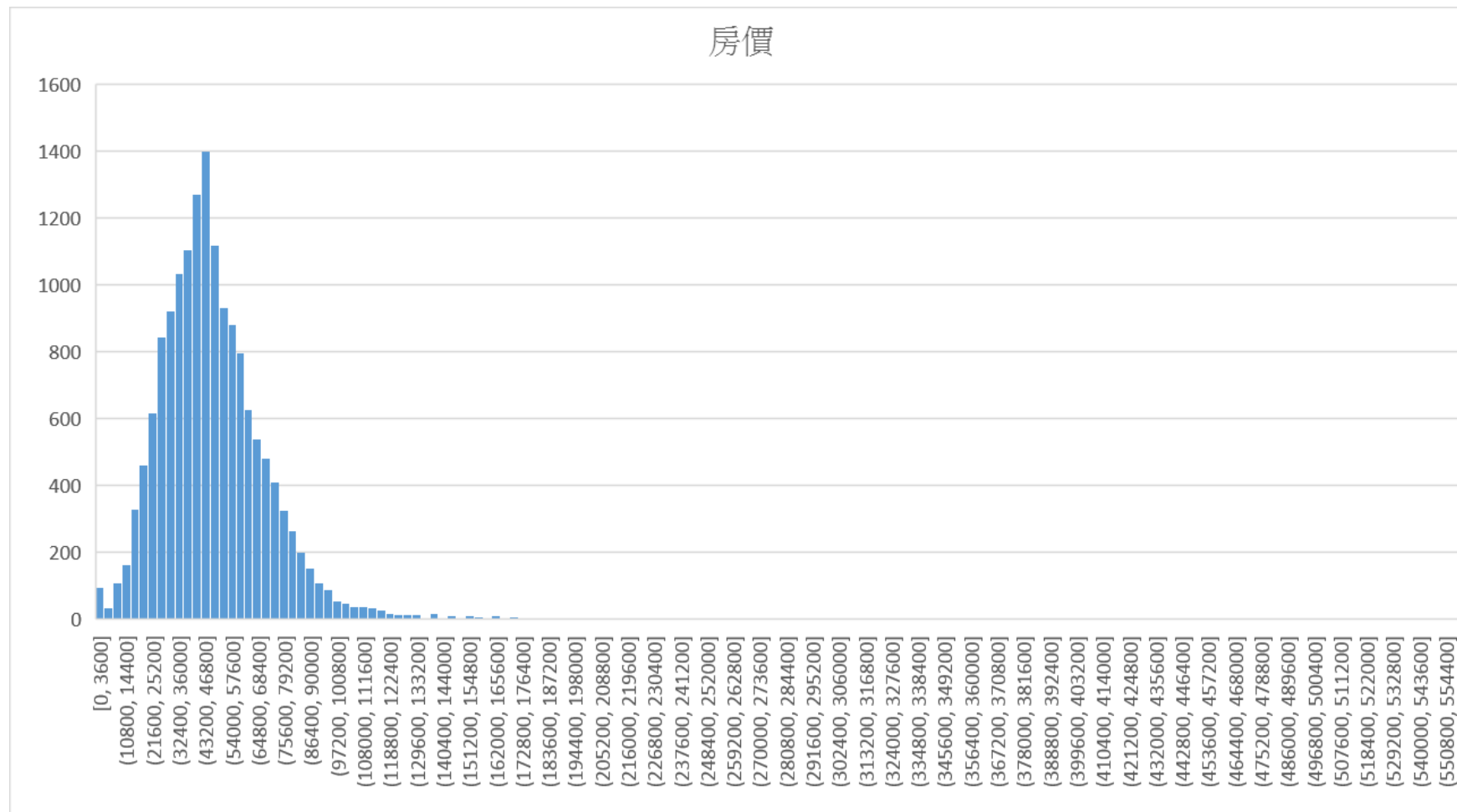
單價(元/平方公尺)

▶ 自變數(7項)(代碼見下頁)

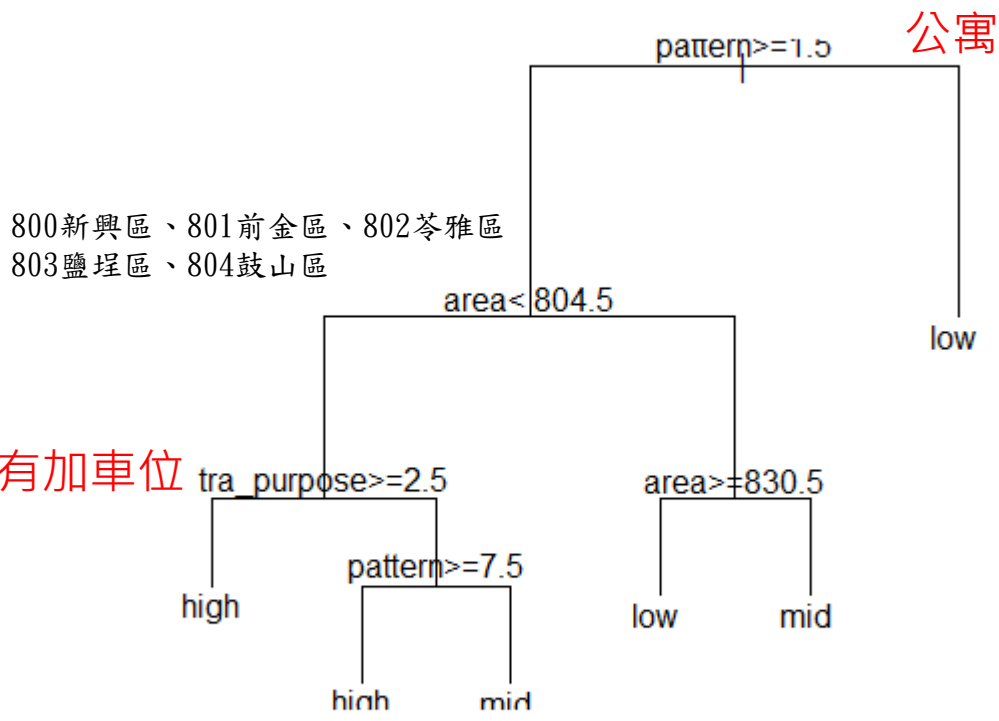
鄉鎮市區、交易標的、使用分區或編定、建物型態、建物移轉總面積、有無隔間
及有無管理組織。

數據檢視

▶ 房價 (依大小排序做長條圖)



CART分析演算法



SVM分析演算法

▶ 取70%作為訓練集,30%為測試集

▶ 訓練集正確率 **61.23%**

```
testingtarget
test.pred high low mid
high 483 80 174
low 63 450 207
mid 636 677 1968
```

```
> accuracy = sum(testingtarget == test.pred) /length(testingtarget)
```

```
> sprintf("%.2f%%", accuracy * 100)
```

```
[1] "61.23%"
```

▶ 測試集正確率 **61.00%**

```
trainingtarget
train.pred high low mid
high 1087 169 445
low 113 1076 522
mid 1567 1496 4581
```

```
> accuracy = sum(trainingtarget == train.pred) /length(trainingtarget)
```

```
> sprintf("%.2f%%", accuracy * 100)
```

```
[1] "61.00%"
```

比較與結論

▶ 2015年1~12月資料訓練集與測試集正確率

正確率	CART演算法	C5.0演算法	SVM演算法	Naive-Bayes演算法
訓練集	59.2%	68.9%	61.23%	51.88%
測試集	58.4%	65.5%	61.00%	51.71%

正確率	CART演算法	C5.0演算法	SVM演算法	Naive-Bayes演算法
測試資料	54.9%	61.1%	56.85%	51.71%

▶ 結論

1. 從分類迴歸樹可知,地區(area)、建物型態(pattern)以及交易標的(tra_purpose)為主要影響房價因素。
2. C5.0演算法較其他演算法來得好些
3. 年份差別所測試結果並無太大差異