



ML FINAL PROJECT:

Real Estate Price Prediction

Team 16 :

0616022 張軒瑞

0616064 洪佑驊

0616242 唐飴萃

0616038 蔡雨恩

0616116 虞樸

TABLE OF CONTENTS

01

Introduction

03

Data Visualization

02

Data Pre-Processing

04

Model Construction
-Naive Bayes

05

Model Construction
-Decision Tree

07

Problems and Future Work

06

Model Construction
-KNN

08

Conclusion

01

INTRODUCTION



INTRODUCTION



我們想利用政府提供的不動
產實價登錄資訊，依照用途來
預測各地的房價

三峽區

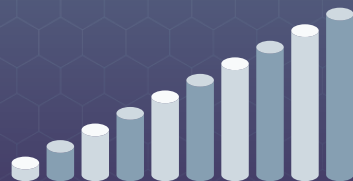
永和區

住

農

工

商



MOTIVATION

方便比價

方便我們在購屋或租屋時與周圍房屋做比較



快速搜索

能以過往的資料快速查找各功能性適合的房屋



預測房價走勢

透過比較歷年各地區房價做出預測的走勢(這次僅做當年度)



02

DATA

PRE-PROCESSING



資料來源

政府資料開放平台 > 資料集 > 不動產實價登錄資訊-買賣案件



政府資料開放平台

全部 新聞 圖片 影片 地圖

約有 9,340,000 項結果 (政府資料開)

政府資料開放平台

https://data.gov.tw

國家發展委員會-
https://www.ndc.gov.tw
一、背景: 政府施政透明
資料的開放, 可促使跨機關

國家發展委員會-
https://www.ndc.gov.tw
國家發展委員會資料開
放平台財產申報統計表、區

政府資料開放平臺

全部資料集 網站導覽 互動專區 資料故事館 NEW 最新消息 諮詢小組 關於平臺 EN 應用案例與活動

提供機關

基隆市政府地政處 59
新北市政府地政局 56
臺南市政府地政局 47
高雄市政府地政局 34
財政部國有財產署 29
行政院主計總處 22
桃園市政府地政局 15
經濟部商業司 14
臺中市政府地政局 10
地政司 8

▼ [顯示全部]

服務分類

公共資訊 160
購屋及遷徙 159
開創事業 15
投資理財 6
交通及通訊 4
生活安全及品質 1

不動產實價登錄資訊-買賣案件

月、建物總面積平方公尺、建物現況格局-房、建物現況格局-廳、建物現況格局-衛、建物現況格局-隔間、有無管理組織、有無附傢俱、總額 元、單價元平方公尺、車位類別、車位面積平方公尺、車位總額元、備註、編號、土地區段位置、土地移轉總面積平方公尺、使用分區或編定、屋齡、建物移轉面積平方公尺、主要用途、主要建材、建築完成日期、總層數、建物分層、車位類別、車位價格、車位面積平方公尺
地政司 / 詮釋資料更新時間: 2019/06/27 15:40
19173 15734 0
XML CSV

1. 不動產買賣案件實價登錄資訊, 包含標的位置(去識別化)、面積、總價等資訊。 2. 本資料集為每10日更新一次。
主要欄位說明:
district(鄉鎮市區)、rps01(交易標的)、rps02(土地區段位置或建物區門牌)、rps03(土地移轉總面積平方公尺)、rps04(都市土地使用分區)、rps05(非都市土地使用分區)、rps06(非都市土地編定)、rps07(交易年月)、rps08(交易筆棟數)、rps09(移轉層次)、rps10(總樓層數)、rps11(建物型態)、rps12(主要用途)、rps13(主要建材)、rps14(建築完成年月)、rps15(建物移轉總面積平方公尺)、rps16(建物現況格局-房)、rps17(建物現況格局-廳)、rps18(建物現況格局-衛)、rps19(建物現況格局-隔間)、rps20(有無管理組織)、rps21(總價元)、rps22(單價每平方公尺)、rps23(車位類別)、rps24(車位移轉總面積平方公尺)、rps25(車位總價元)、rps26(備註)、rps27(編號)
新北市政府地政局 / 詮釋資料更新時間: 2019/12/30 13:50
8307 4321 2

資料集

全部資料集

最新資料集

互動專區

我有話要說

我想要更多

資料故事館

應用展示專區

教育資源專區

統計資料

網站使用統計

服務分類統計

諮詢小組

諮詢小組

關於平臺

關於我們

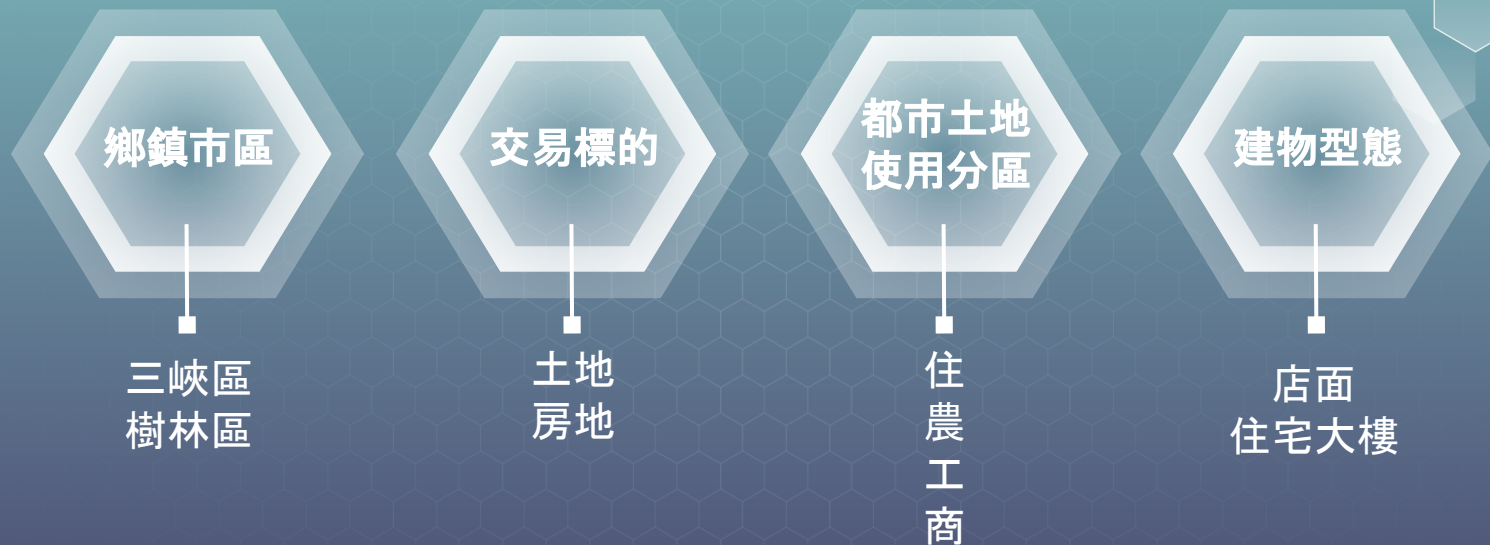
規範&授權

授權條款

M2M專區

district	rps01	rps02	rps03	rps04	rps05	rps06	rps07	rps08	rps09	rps10	rps11	rps12	rps13	rps14	rps15	rps16	rps17	rps18	rps19	rps20	rps21	rps22	rps23	rps24	rps25	rps26	rps27			
板橋區	房地(土地、建物)	新北市板橋區新莊街23巷1-30號	30.44	住			1080630	土地1建物1單位0	四層	五層	公寓(5樓含以下無電梯)	住家用	鋼筋混凝土造	770908	112.81	3	2	2	有	無	9090000	80578	0	0	0	含增建或未登記建物。;	RP9PMLMLP			
板橋區	房地(土地、建物)・車位	新北市板橋區文化路一段188巷31-60號	22.6	住			1080417	土地1建物1單位1	三層	十五層	住宅大樓(11層含以上有電梯)	見其他登記事項	鋼筋混凝土造	1051004	243.75	3	2	2	有	有	40470000	190149	坡道平面	44.59	2600000		RP9PMLMLP			
板橋區	土地	文化段2191-2220地號	0.3	住			1080417	土地1建物0單位0			其他				0	0	0	0	有	有	120000	0	0	0	0	0	畸零地或有合併使用之交易。;	RP9QMLMLP		
土城區	房地(土地、建物)	新北市土城區學府路一段98巷1-30號	4.19	住			1080627	土地1建物1單位0	九層	十三層	寮房(1房1廳1衛)	住家用	鋼筋混凝土造	821122	29.23	1	0	1	有	有	3600000	123161	0	0	0	0	0	含增建或未登記建物。;	RP9SMLMLP	
板橋區	房地(土地、建物)・車位	新北市板橋區縣民大道一段271-300號	41.33	住			1080114	土地3建物1單位2	九層	二十九層	住宅大樓(11層含以上有電梯)	見其他登記事項	鋼骨鋼筋混凝土造	1040505	520.65	0	0	0	無	有	89680000	179352	坡道平面	49.62	5200000	建商與地主合建案。;	RP9TMLMLP			
土城區	房地(土地、建物)・車位	新北市土城區中州路30巷31-60號	28.73	住			1080602	土地5建物1單位1	二層	十九層	住宅大樓(11層含以上有電梯)	住家用	鋼筋混凝土造	1020503	203.2	4	2	2	有	有	13950000	68652	坡道平面	32.35	0	0	共有人數2人。含增建或未登記建物。;	RP9WMLMLP		
板橋區	土地	成功段121-150地號	710.04	住			1080413	土地3建物0單位0			其他				0	0	0	0	無	有	91200000	128443	0	0	0	0	0		RP9TMLMLP	
土城區	房地(土地、建物)	新北市土城區日新街33巷1-30號	946.81	工			1080528	土地1建物4單位0	一層	二層	廠房	工業用	加強磚造	641117	1890.75	32	5	32	有	無	200000000	105778	0	0	0	0	0	含增建或未登記建物。權利人共11人。現況整棟為安養院使用。;	RP9WMLMLP	
土城區	房地(土地、建物)・車位	新北市土城區立仁街31-60號	5.39	商			1080622	土地1建物1單位1	二十層	二十層	住宅大樓(11層含以上有電梯)	住家用	鋼筋混凝土造	820726	65.51	2	1	1	有	有	7300000	111433	坡道平面	0	0	0	0	0		RP9TMLMLP
土城區	土地	頂福段991-1020地號	1	其他			1080618	土地2建物0單位0			其他				0	0	0	0	有	無	195239	0	0	0	0	0	0	本件為協議價購案件	RP9TMLMLP	
土城區	土地	頂福段991-1020地號	1	其他			1080618	土地2建物0單位0			其他				0	0	0	0	有	無	195239	0	0	0	0	0	0	本件為協議價購案件	RP9TMLMLP	
土城區	土地	頂福段991-1020地號	1	其他			1080618	土地2建物0單位0			其他				0	0	0	0	有	無	195239	0	0	0	0	0	0	本件為協議價購案件	RP9TMLMLP	
土城區	土地	頂福段991-1020地號	4.12	其他			1080618	土地2建物0單位0			其他				0	0	0	0	有	無	807739	196053	0	0	0	0	0	0	本件為協議價購案件	RP9TMLMLP
土城區	土地	頂福段1051-1080地號	1.01	其他			1080618	土地1建物0單位0			其他				0	0	0	0	有	無	197890	195931	0	0	0	0	0	0	本件為協議價購案件	RP9TMLMLP
土城區	土地	頂福段1051-1080地號	1.01	其他			1080618	土地1建物0單位0			其他				0	0	0	0	有	無	197890	195931	0	0	0	0	0	0	本件為協議價購案件	RP9TMLMLP
土城區	土地	頂福段1171-1200地號	250.84	其他			1080618	土地10建物0單位0			其他				0	0	0	0	有	無	49164640	196000	0	0	0	0	0	0	本件為協議價購案件	RP9TMLMLP
板橋區																														

使用到的特徵(Feature)



資料處理

STEP I.挑出需要的特徵並將中文轉為英文方便後處理可視化

1	district	type	usage	building_type	price
2	Ban_Chiao	building_land	live	noelev_5down	80.578
3	Ban_Chiao	building_land_parkspace	live	elev_11up	190.149
4	Tu_Cheng	building_land	live	suite	123.161
5	Ban_Chiao	building_land_parkspace	live	elev_11up	179.352
6	Tu_Cheng	building_land_parkspace	live	elev_11up	68.652
7	Ban_Chiao	land	live	other	128.443
8	Tu_Cheng	building_land	industry	factory_office	105.778
9	Tu_Cheng	building_land_parkspace	business	elev_11up	111.433
10	Tu_Cheng	land	other	other	196.053
11	Tu_Cheng	land	other	other	195.931
12	Tu_Cheng	land	other	other	195.931
13	Tu_Cheng	land	other	other	196
14	Ban_Chiao	building_land	live	noelev_5down	59.411
15	Ban_Chiao	building_land	live	noelev_5down	184.332
16	Ban_Chiao	building_land_parkspace	live	elev_10down	75.481
17	Ban_Chiao	land	other	other	73.03
18	Ban_Chiao	building_land	live	noelev_5down	110.167
19	Tu_Cheng	building_land	live	elev_11up	104.617
20	Tu_Cheng	building_land	live	noelev_5down	75.11
21	Tu_Cheng	building_land_parkspace	business	elev_11up	124.035
22	Tu_Cheng	building_land_parkspace	business	elev_11up	124.378
23	Tu_Cheng	building_land_parkspace	business	elev_11up	127.28
24	Tu_Cheng	building_land_parkspace	business	elev_11up	129.761
25	Tu_Cheng	building_land_parkspace	business	elev_11up	122.119
26	Tu_Cheng	building_land_parkspace	business	elev_11up	129.154

STEP 2.

將連續資料離散化

	A	B	C	D	E	F
1	district	type	usage	building_type	price	
2	Ban_Chiao	building_land	live	noelev_5down	100	
3	Ban_Chiao	building_land_parkspace	live	elev_11up	200	
4	Tu_Cheng	building_land	live	suite	200	
5	Ban_Chiao	building_land_parkspace	live	elev_11up	200	
6	Tu_Cheng	building_land_parkspace	live	elev_11up	100	
7	Ban_Chiao	land	live	other	200	
8	Tu_Cheng	building_land	industry	factory_office	200	
9	Tu_Cheng	building_land_parkspace	business	elev_11up	200	
10	Tu_Cheng	land	other	other	200	
11	Tu_Cheng	land	other	other	200	
12	Tu_Cheng	land	other	other	200	
13	Tu_Cheng	land	other	other	200	
14	Ban_Chiao	building_land	live	noelev_5down	100	
15	Ban_Chiao	building_land	live	noelev_5down	200	
16	Ban_Chiao	building_land_parkspace	live	elev_10down	100	
17	Ban_Chiao	land	other	other	100	
18	Ban_Chiao	building_land	live	noelev_5down	200	
19	Tu_Cheng	building_land	live	elev_11up	200	
20	Tu_Cheng	building_land	live	noelev_5down	100	
21	Tu_Cheng	building_land_parkspace	business	elev_11up	200	
22	Tu_Cheng	building_land_parkspace	business	elev_11up	200	
23	Tu_Cheng	building_land_parkspace	business	elev_11up	200	
24	Tu_Cheng	building_land_parkspace	business	elev_11up	200	

STEP 3.

將名稱的資訊轉為數字代號

適意實際上先轉成英文

```
cn_dis = {'三峽區' : 0, '三芝區' : 1, '三重區' : 2, '中和區' : 3, '五股區' : 4, '八里區' : 5, '土城區' : 6, '平溪區' : 7, '新店區' : 8, '新莊區' : 9, '板橋區' : 10, '林口區' : 11, '樹林區' : 12, '永和區' : 13, '汐止區' : 14, '泰山區' : 15, '淡水區' : 16, '深坑區' : 17, '烏來區' : 18, '瑞芳區' : 19, '基隆區' : 20, '宜蘭區' : 21, '花蓮區' : 22, '澎湖區' : 23, '金門區' : 24, '馬祖區' : 25, '屏東區' : 26, '嘉義區' : 27}
cn_type = {'土地' : 0, '建物' : 1, '其他' : 2, '套房' : 3, '其他' : 4, '住' : 5, '住' : 6, '住' : 7, '住' : 8, '住' : 9, '住' : 10, '住' : 11, '住' : 12, '住' : 13, '住' : 14, '住' : 15, '住' : 16, '住' : 17, '住' : 18, '住' : 19, '住' : 20, '住' : 21, '住' : 22, '住' : 23, '住' : 24, '住' : 25, '住' : 26, '住' : 27}
use_type = {'住' : 0, '住' : 1, '住' : 2, '住' : 3, '住' : 4, '住' : 5, '住' : 6, '住' : 7, '住' : 8, '住' : 9, '住' : 10, '住' : 11, '住' : 12, '住' : 13, '住' : 14, '住' : 15, '住' : 16, '住' : 17, '住' : 18, '住' : 19, '住' : 20, '住' : 21, '住' : 22, '住' : 23, '住' : 24, '住' : 25, '住' : 26, '住' : 27}
build = {'住宅大樓(11~15層)' : 1, '其他' : 2, '套房' : 3, '其他' : 4, '住' : 5, '住' : 6, '住' : 7, '住' : 8, '住' : 9, '住' : 10, '住' : 11, '住' : 12, '住' : 13, '住' : 14, '住' : 15, '住' : 16, '住' : 17, '住' : 18, '住' : 19, '住' : 20, '住' : 21, '住' : 22, '住' : 23, '住' : 24, '住' : 25, '住' : 26, '住' : 27}
```

	A	B	C	D	E	F
1	district	type	usage	building_type	price	
2	10	2	0	1	100	
3	10	3	0	0	200	
4	6	2	0	3	200	
5	10	3	0	0	200	
6	6	3	0	0	100	
7	10	0	0	2	200	
8	6	2	3	6	200	
9	6	3	2	0	200	
10	6	0	1	2	200	
11	6	0	1	2	200	
12	6	0	1	2	200	
13	6	0	1	2	200	
14	10	2	0	1	100	
15	10	2	0	1	200	
16	10	3	0	7	100	
17	10	0	1	2	100	
18	10	2	0	1	200	

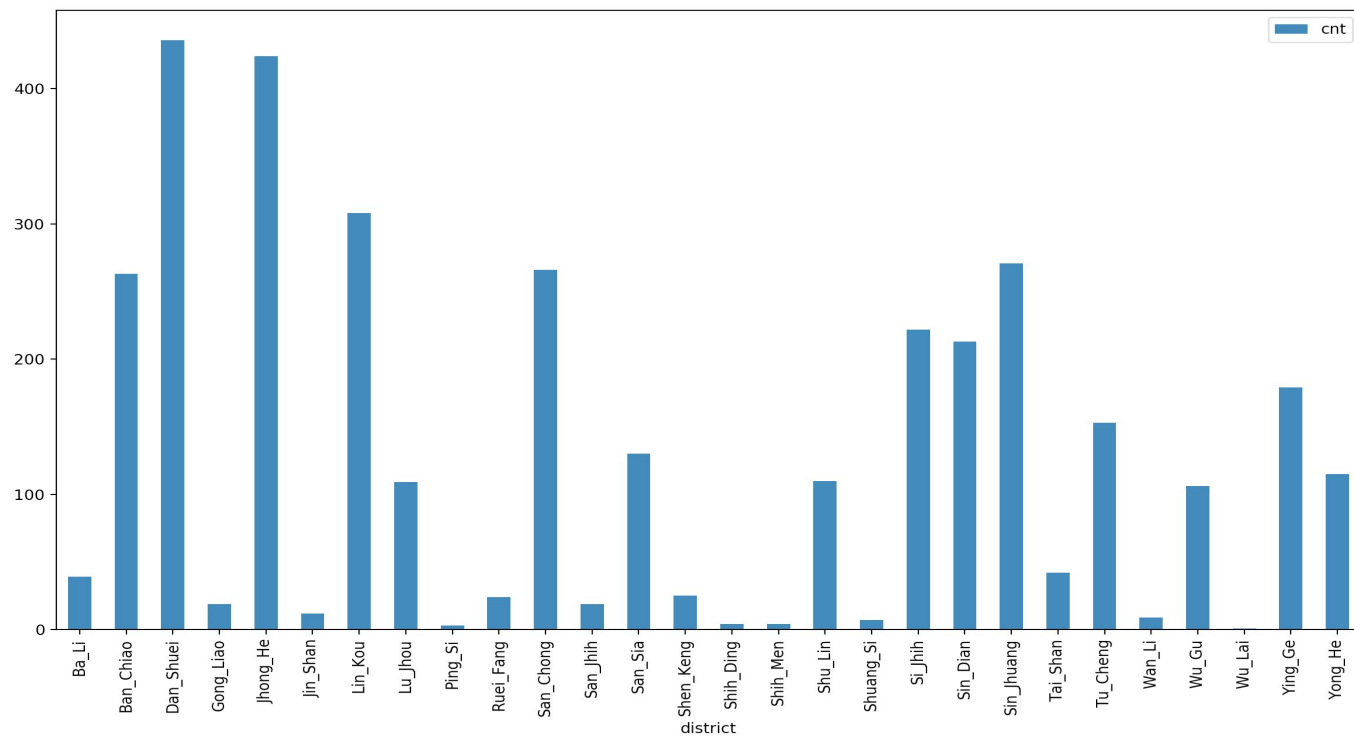
03

Data

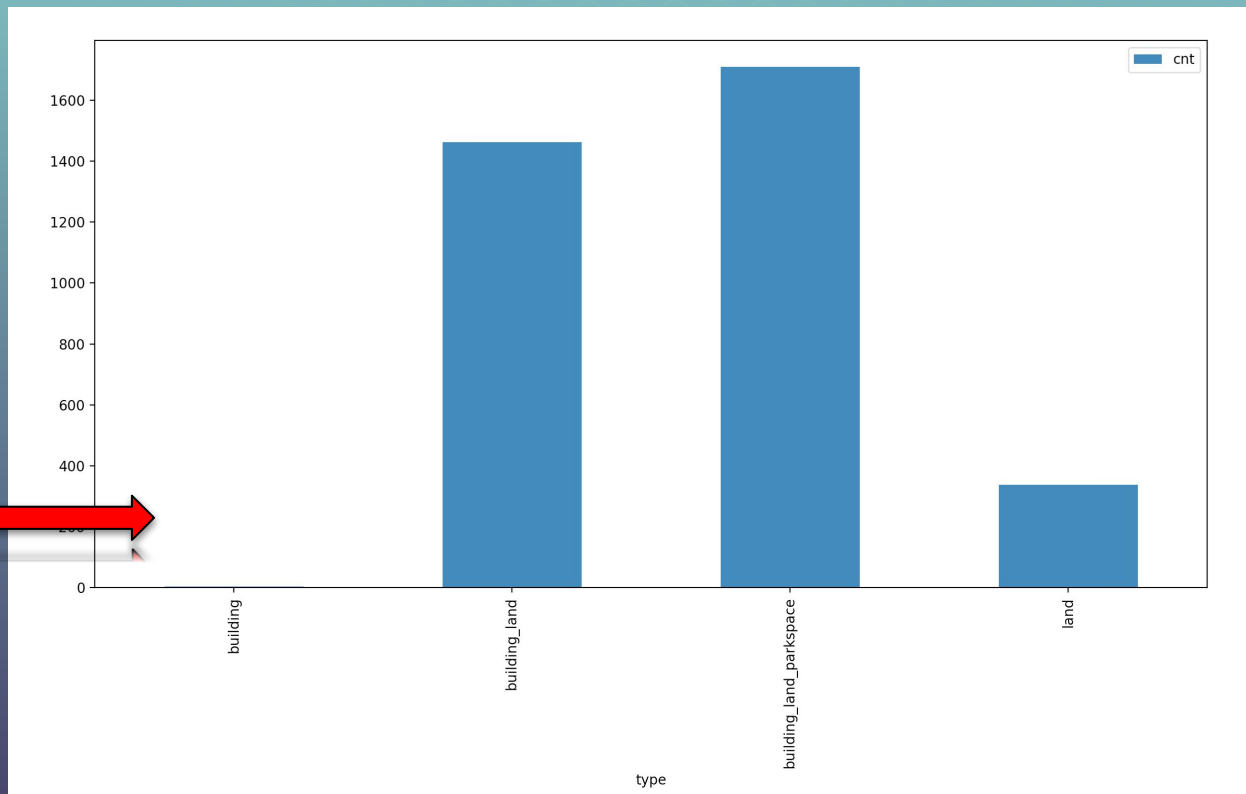
Visualization



各鄉鎮市區本月成交之不動產

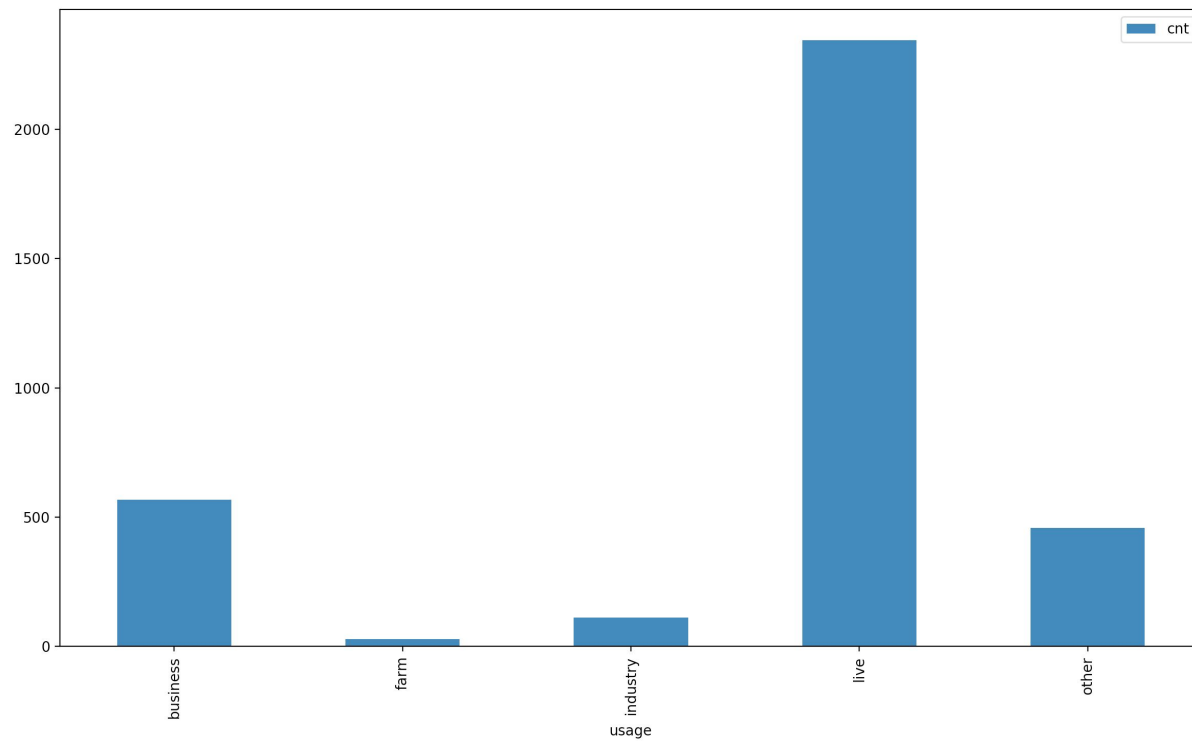


交易標的



本月純建物交易幾乎為無

都市土地使用分區(使用目的)



建物型態

旁邊依次為：

辦公商業大樓

華廈(10層含以下有電梯)

住宅大樓(11層含以上有電梯)

廠辦

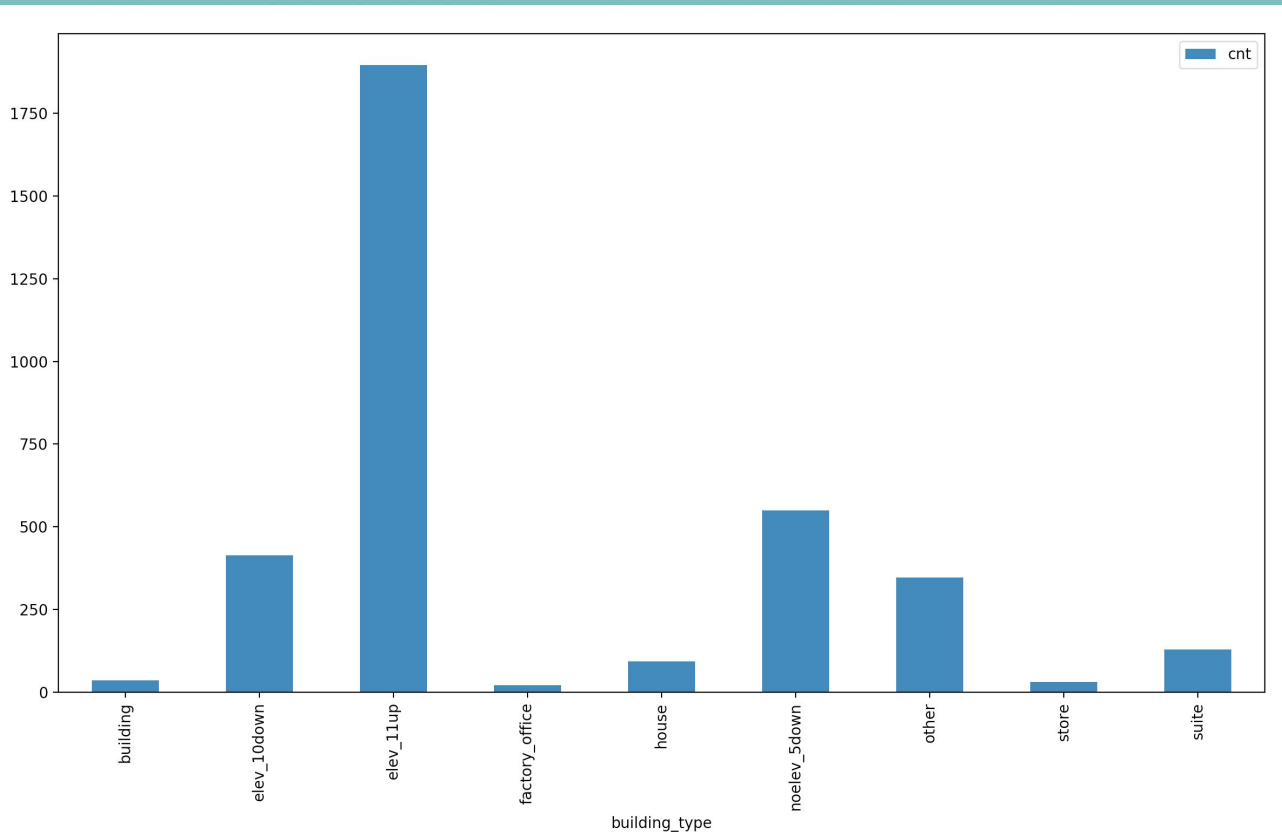
透天厝

公寓(5樓含以下無電梯)

其他

店面(店鋪)

套房(1房1廳1衛)



04

Model Construction - Naive Bayes



Naive Bayes

```
1  from sklearn.naive_bayes import GaussianNB
2
3  clf = GaussianNB()
4  clf = clf.fit(trainData, trainTarget)
5
6  y_pred=clf.predict(testData)
7
8  total = testData.shape[0]
9  correct = (testTarget == y_pred).sum()
10 print("total: %d correct : %d" % (total, correct))
11 print("acc: %f" % (correct / total))
```

Result



```
~/workspace/machine learning/final/house_price ➤ python3 naive_bayes.py  
total: 1055 correct : 671  
acc: 0.636019
```

05

Model Construction – Decision Tree



Decision Tree

```
1  from sklearn.tree import DecisionTreeClassifier
2
3  clf = tree.DecisionTreeClassifier()
4  clf = clf.fit(trainData, trainTarget)
5  y_pred=clf.predict(testData)
6
7  total = testData.shape[0]
8  correct = (testTarget == y_pred).sum()
9  print("total: %d correct : %d" % (total, correct))
10 print("acc: %f" % (correct / total))
```

Result



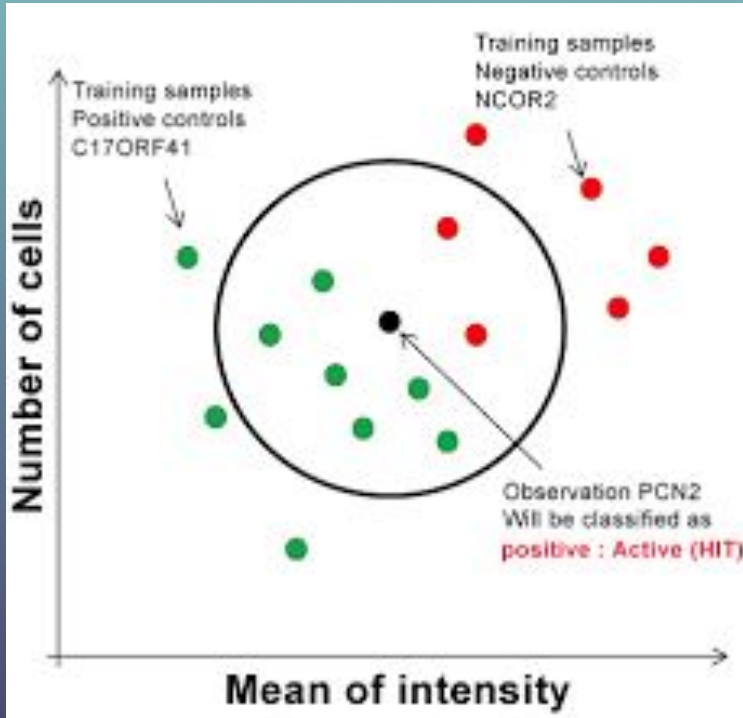
```
~/workspace/machine learning/final/house_price python3 decision_tree.py  
total: 1055 correct : 864  
acc : 0.818957
```


06

Model Construction -KNN



KNN(k-Nearest Neighbors)



```
def KNeighborsClassifier(n_neighbors = 5,  
                        weights='uniform',  
                        algorithm = '',  
                        leaf_size = '30',  
                        p = 2,  
                        metric = 'minkowski',  
                        metric_params = None,  
                        n_jobs = None  
                        )
```

`n_neighbors = 5`
`weights = 'uniform'`

Knn

```
1  from sklearn.neighbors import KNeighborsClassifier
2
3  clf = neighbors.KNeighborsClassifier()
4  clf = clf.fit(trainData, trainTarget)
5
6  y_pred=clf.predict(testData)
7
8  total = testData.shape[0]
9  correct = (testTarget == y_pred).sum()
10 print("total: %d correct : %d" % (total, correct))
11 print("acc: %f" % (correct / total))
```

Result



```
~/workspace/machine learning/final/house_price ➤ python3 knn.py
```

```
total: 1055 correct : 841
```

```
acc : 0.797156
```

07

Problem and Future Work



Problem



資料筆數不夠

以月為單位



手刻model

刻爆



資料分布不均

容易集中於
某區域、類型

Future Work

收集更多月份資料

->預測未來趨勢

->分區更細(路段、地址...)



08

Conclusion



Conclusion

1. 所在的鄉鎮市區
為影響建物價錢的主要因素
2. Decision Tree的準確率最高

