

資料探勘預測分析— 以大台北地區房地產為例

專題第五組

目錄

- ◎ 研究動機與目的
- ◎ 研究流程
- ◎ 研究方法
- ◎ 群集分析
- ◎ 預測分析

研究動機

- ◎ 自 101 年 8 月起，實價登錄正式執行，並公開讓民眾查詢，同時也展現諸多打房政策，於是掀起一股對房價起伏的熱烈探討。
- ◎ 因此，我們決定以資料探勘方法來找出房地產暗藏的秘密，以及潛在的規則。

研究目的

1. 了解大台北地區房屋交易類型
2. 掌握影響房價的強力因素
3. 分析未來房價可能的趨勢，以此創造投資優勢

研究流程

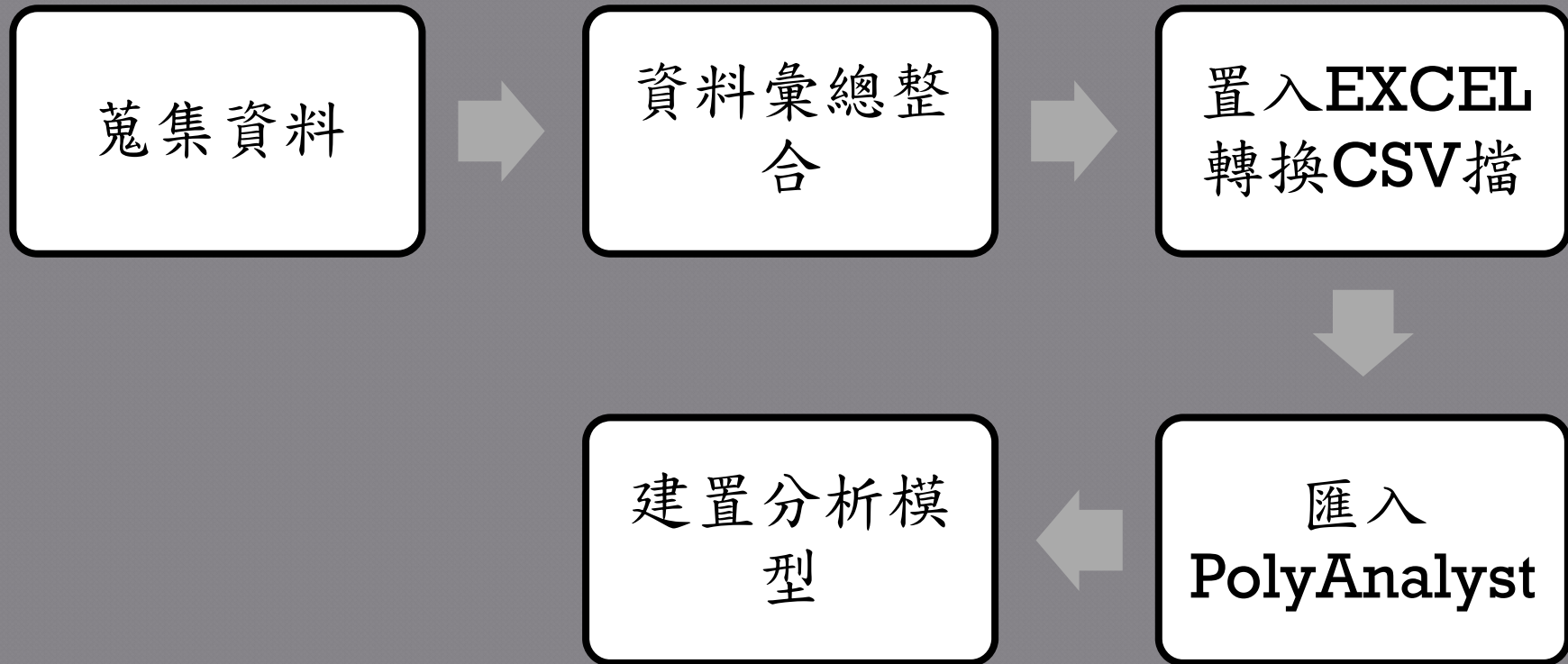
蒐集資料

資料彙總整合

置入EXCEL
轉換CSV檔

建置分析模型

匯入
PolyAnalyst



原始資料來源

Raw Data

原始資料

- 一、實價登錄新制自101年8月1日起施行，相關不動產買賣、預售屋及租賃價格自101年10月16日起開放查詢，查詢連結網址：<http://lvr.land.moi.gov.tw/>
- 二、以下房地產交易價格資料僅提供至101年7月底止。

文字模式

圖形模式

交易資料概況

請輸入您要找尋的縣市等相關資料

- × 交易類別：
- × 縣市：
- × 交易季別： 年第 季至 年第 季
101年第3季資料，僅提供至101年7月底止調查之交易資料
- 街道名稱：
- 面積單位： 平方公尺 坪
- 用途類別：
- 使用分區：
- 交易總價： 萬元至 萬元

原始資料

房地交易價格問訊查詢結果

臺北市/大同區(民國89年第一季~民國101年第三季)

◎本文件列印時,印表機請採 A4 橫向列印

共有1654筆資料 [資料說明](#)

民國	季別	街道名稱及範圍	路寬(m)	臨街關係	用途類別	構造種類	建築完成年月	移轉層次	總樓層數	土地使用分區	移轉土地面積(坪)	移轉房屋面積(坪)	總價(萬元)	單價(萬元)
101	三	延平北路2段 延平北路2段1號至100號	16	臨街地	住宅大樓	鋼筋混凝土造	09809	15	15	商業區	2.65	24.38	2120.00	86.96
101	三	承德路3段 承德路3段101號至200號	40	路角地	住宅大樓	鋼筋混凝土造	09608	4	12	商業區	1.17	12.63	638.00	50.51
101	三	承德路3段 承德路3段101號至200號	40	路角地	住宅大樓	鋼筋混凝土造	09608	3	12	商業區	1.17	12.63	620.00	49.09
101	三	承德路3段 承德路3段101號至200號	40	路角地	住宅大樓	鋼筋混凝土造	09608	6	12	商業區	1.98	21.88	988.00	45.16
101	三	承德路3段 承德路3段101號至200號	40	臨街地	公寓	鋼筋混凝土造	06111	3	4	商業區	12.55	37.5	1760.00	46.93
101	三	承德路3段 承德路3段208巷1號至100號	11	臨街地	住宅大樓	鋼筋混凝土造	09605	5	11	商業區	2.06	13.4	696.00	51.94
101	三	承德路3段 承德路3段208巷1號至100號	40	臨街地	住宅大樓	鋼筋混凝土造	09605	6	11	商業區	1.16	7.48	430.00	57.49
101	三	南京西路 南京西路201號至300號	16	臨街地	住宅大樓	鋼筋混凝土造	08911	8	13	商業區	1.78	28.3	1100.8	38.9
101	三	南京西路 南京西路301號至400號	20	臨街地	住宅大樓	鋼筋混凝土造	06711	12	14	商業區	0.94	10.05	385.00	38.31
101	三	南京西路 南京西路301號至400號	20	路角地	住宅大樓	鋼筋混凝土造	06910	12	12	商業區	0.83	12.44	490.00	39.39
101	三	哈密街 哈密街23巷1號至100號	8	裡地	公寓	鋼筋混凝土造	08406	4	5	住宅區	11.25	33.52	1475.00	44
101	三	哈密街 哈密街59巷1號至100號	6	裡地	公寓	鋼筋混凝土造	06201	3	4	住宅區	5.52	15.51	650.00	41.91



原始資料

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1																	
2	89	—	中山區	明水路	明水公園	30	裡地	住宅大樓	鋼筋混凝	8810	4	7	住宅區	9.18	50.16	1360	0
3	89	—	內湖區	金湖路	內湖圍管	20	臨街地	住宅大樓	鋼筋混凝	8810	4	9	住宅區	14.81	51.12	1285	0
4	89	—	北投區	大業路	淡水捷運	30	臨街地	住宅大樓	鋼筋混凝	8808	10	14	商業區	2.87	22.9	608	0
5	89	—	中正區	南昌路二	市圖古亭	15	臨街地	辦公商業	鋼筋混凝	8805	5	12	商業區	1.58	10.01	370	0
6	89	—	中正區	羅斯福路	三軍總醫	20	裡地	住宅大樓	鋼筋混凝	8708	5	6	住宅區	12.12	44.17	960	0
7	89	—	大同區	伊寧街	景化公園	6	裡地	住宅大樓	鋼筋混凝	8706	6	6	商業區	4.45	16.05	360	0
8	89	—	大同區	哈密街	慶昌橋附	8	臨街地	住宅大樓	鋼筋混凝	8606	6	6	住宅區	6.74	26.87	550	0
9	89	—	中正區	寧波西街	中正區行	13	臨街地	住宅大樓	鋼筋混凝	8606	8	12	商業區	0.96	7.54	308	0
10	89	—	中山區	松江路	建國啤酒	40	臨街地	辦公商業	鋼筋混凝	8605	2	13	商業區	1.31	23.98	650	0
11	89	—	北投區	吉利街	吉利超市	10	臨街地	住宅大樓	鋼筋混凝	8601	3	9	商業區	4.59	50.16	810	0
12	89	—	文山區	辛亥路四	辛亥國小	20	臨街地	住宅大樓	鋼筋混凝	8511	10	14	住宅區	17.72	24.46	750	0
13	89	—	南港區	南港路二	東明公園	20	臨街地	住宅大樓	鋼筋混凝	8510	10	13	住宅區	13.92	52.85	1335	0
14	89	—	文山區	忠順街二	忠順市場	12	臨街地	住宅大樓	鋼筋混凝	8509	2	8	商業區	4.8	17.36	570	0
15	89	—	文山區	木柵路三	木柵國中	20	裡地	住宅大樓	鋼筋混凝	8508	1	10	住宅區	11.5	34.94	1160	0
16	89	—	中山區	長春路	長春市場	20	臨街地	住宅大樓	鋼筋混凝	8507	16	16	商業區	4.49	41.56	980	0
17	89	—	大同區	歸綏街	靜修女中	6	臨街地	住宅大樓	鋼筋混凝	8505	7	13	商業區	5.4	31.3	535	0
18	89	—	北投區	文林北路	自強街與	42	臨街地	住宅大樓	鋼筋混凝	8502	17	23	住宅區	8.13	33.59	1250	0
19	89	—	文山區	試院路	華興公園	12	裡地	住宅大樓	鋼筋混凝	8410	3	6	住宅區	11.02	37.92	840	0
20	89	—	文山區	新光路一	木柵動物	30	臨街地	住宅大樓	鋼筋混凝	8407	8	12	住宅區	9.11	36.01	920	0
21	89	—	北投區	石牌路二	嘿哩岸公	12	裡地	住宅大樓	鋼筋混凝	8406	6	11	住宅區	3.95	14.14	410	0
22	89	—	文山區	興隆路二	興隆公園	22	臨街地	透天住宅	鋼筋混凝	8405	全	2	住宅區	33.58	40.88	1430	0
23	89	—	文山區	和興路	考試院附	8	臨街地	住宅大樓	鋼筋混凝	8404	7	7	住宅區	11.3	30.26	800	0
24	89	—	內湖區	成功路四	碧湖國小	30	裡地	住宅大樓	鋼筋混凝	8402	8	13	商業區	1.84	13.82	340	0
25	89	—	文山區	福興路	興福國中	10	臨街地	住宅大樓	鋼筋混凝	8402	8	10	住宅區	3.63	12.93	350	0
26	89	—	內湖區	成功路四	康寧國小	30	裡地	住宅大樓	鋼筋混凝	8401	10	18	住宅區	7.01	33.94	1100	0
27	89	—	文山區	下崙路	木柵游泳	10	裡地	住宅大樓	鋼筋混凝	8401	3	6	住宅區	2.77	9.68	240	0
28	89	—	南港區	興中路	南港區行	15	裡地	住宅大樓	鋼筋混凝	8312	5	17	商業區	2.59	24.96	515	0
29	89	—	北投區	東華街二	東華公園	10	臨街地	住宅大樓	鋼筋混凝	8310	13	16	住宅區	29.02	84.78	1500	0
30	89	—	中山區	明水路	明水公園	30	裡地	住宅大樓	鋼筋混凝	8309	1	6	住宅區	8.33	28.41	1350	0
31	89	—	內湖區	康寧路三	南湖國小	11	臨街地	住宅大樓	鋼筋混凝	8307	7	9	住宅區	4.48	27.65	470	0
32	89	—	南港區	興南路	東新國小	12	裡地	住宅大樓	鋼筋混凝	8302	6	11	住宅區	11.11	44.77	775	0
33	89	—	內湖區	康寧路三	明湖國小	20	裡地	公寓	鋼筋混凝	8212	1	5	住宅區	9.06	26.16	620	0
34	89	—	北投區	石牌路二	榮民總醫	12	裡地	住宅大樓	鋼筋混凝	8212	6	6	住宅區	6.49	27.95	700	0

研究方法

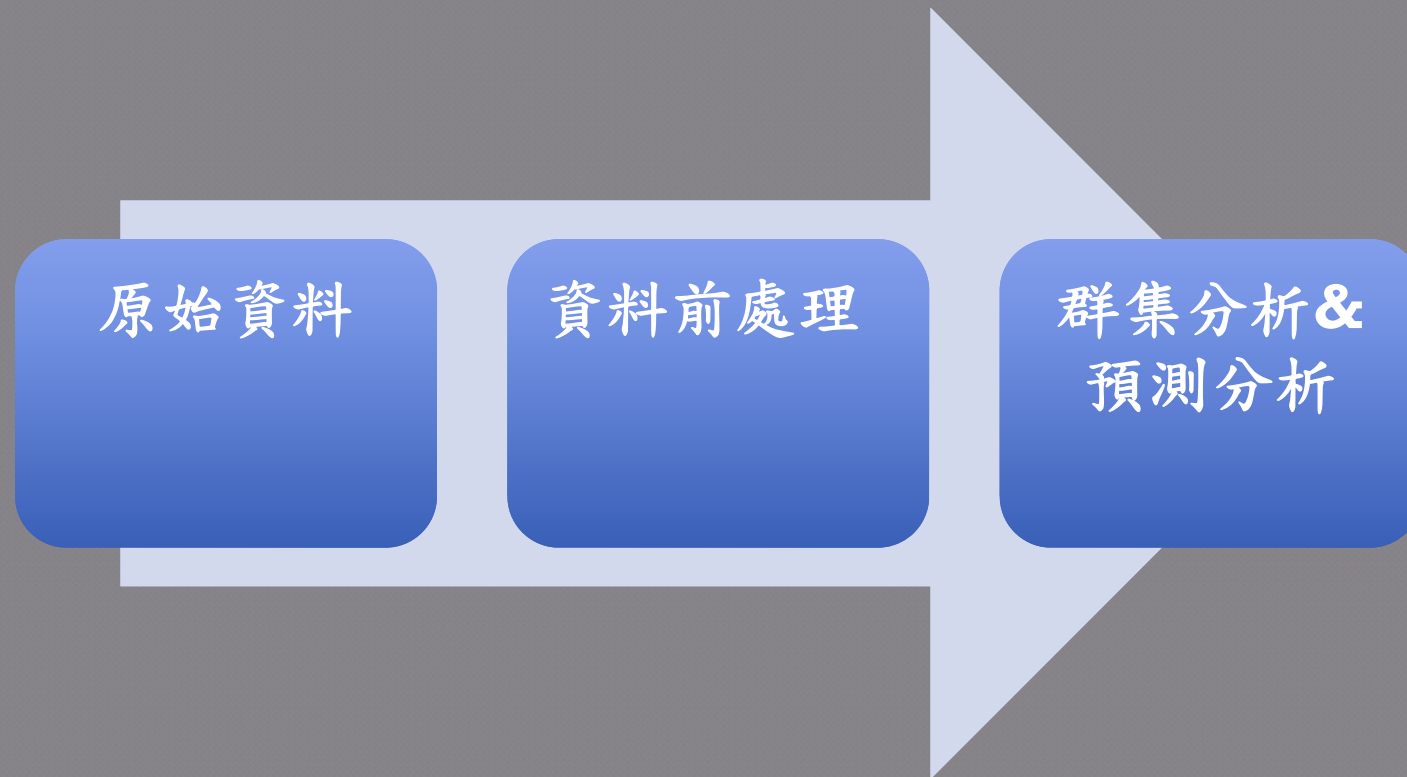
◎ Clustering 群集分析

- 台北市與新北市分別作群集分析

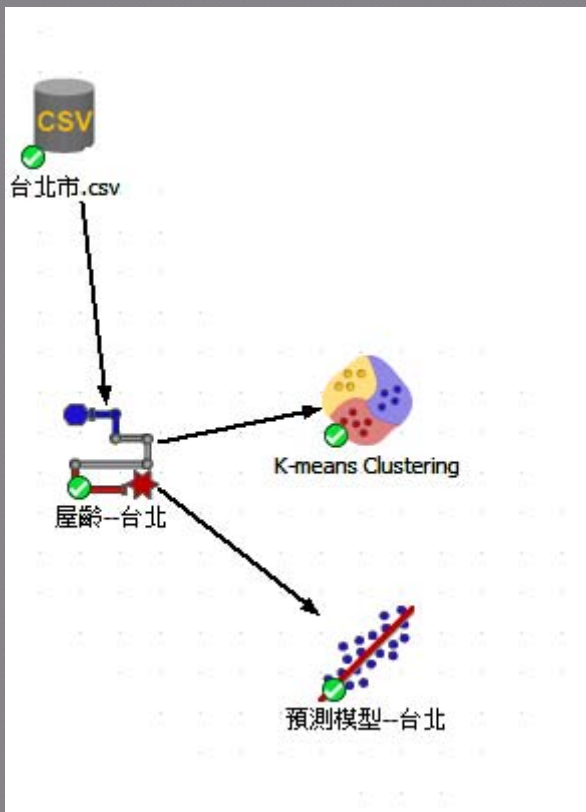
◎ Prediction 預測分析

- 建立房價預測模型

資料探勘流程

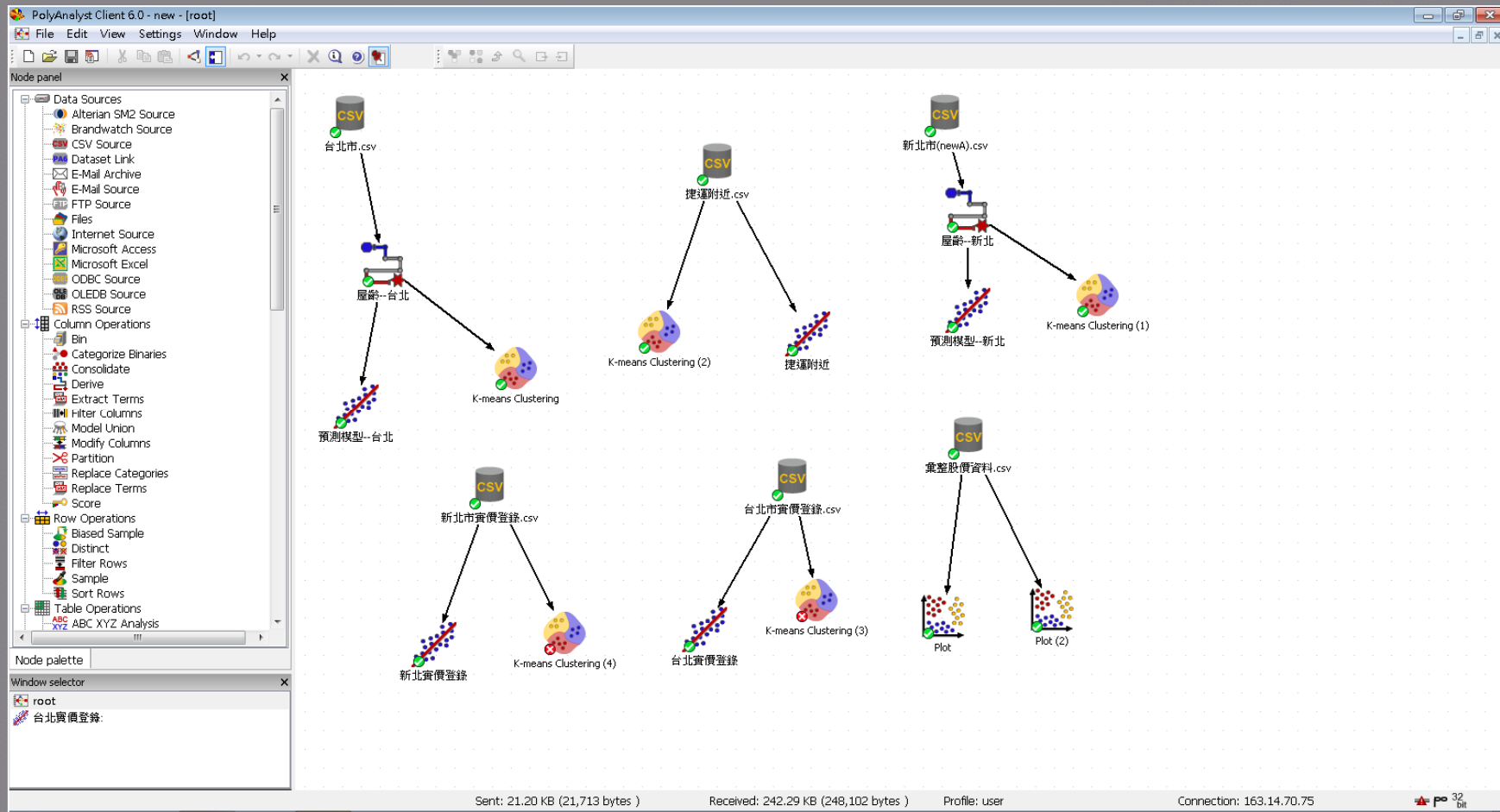


使用工具



- ◎ **CSV**—用於分析的資料格式
- ◎ **Derive**—用於欄位間的計算處理(Ex：以求屋齡)
- ◎ **K-means Clustering**
--群集分析，找出不同的群集，以求群集間的特徵
- ◎ **Liner Regression**—線性回歸，用於預測分析，目的為找出影響房價之因素

系統畫面



群集分析

Clustering

群集分析是什麼?

- ◎ 一種劃分方法，目的在將一些事物歸併在一起，利用群體中各事物都具有相同的特性，而在群體與群體之間卻有顯著的差異。

群集分析

- ◎ 臺北市和新北市作為群集分析項目，再運用資料探勘**K-Means**群集分析方法，交互分析探討臺北市和新北市房地產交易特性之群集關聯性，提供買賣房屋屋時作為參考

Clustering

◎ 台北市:

- 最大差異在於住宅區/商業區、用途類別
- 交易集中在第四季

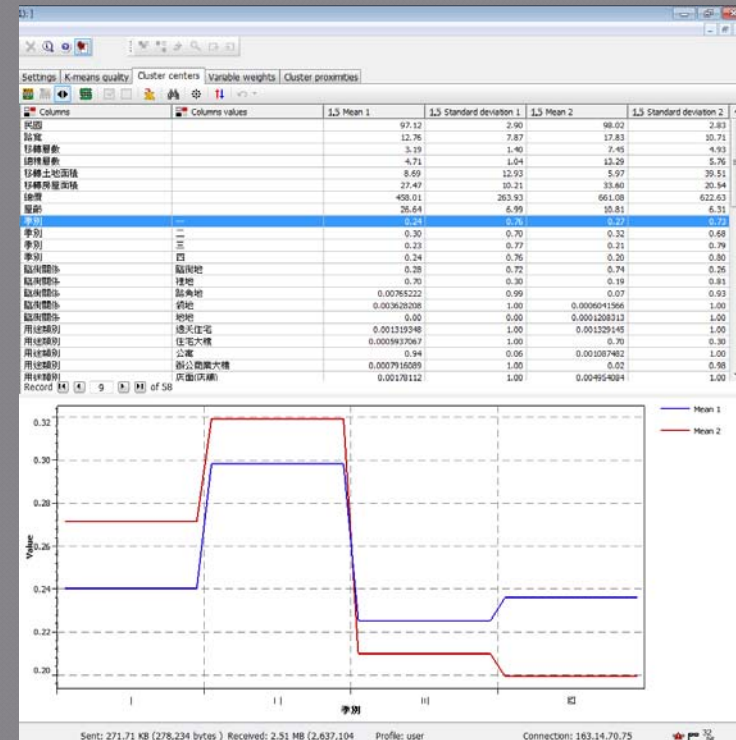
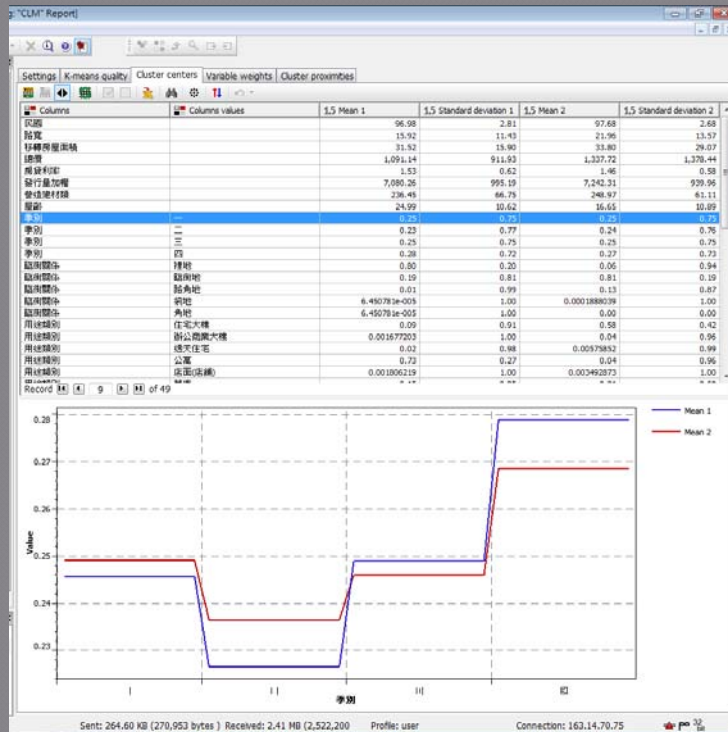
◎ 新北市:

- 最大差異在於房價、屋齡、用途類別
- 交易集中在第二季

群集分析-季別

台北市

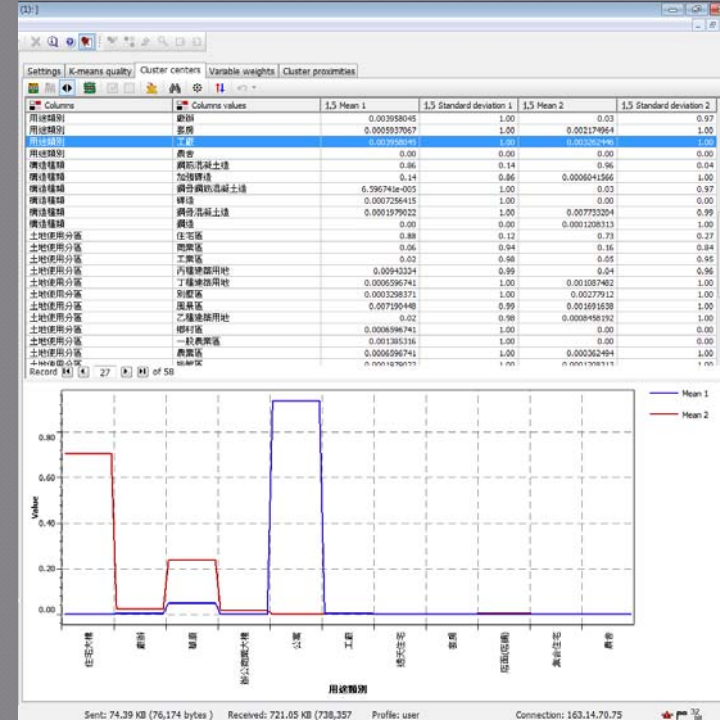
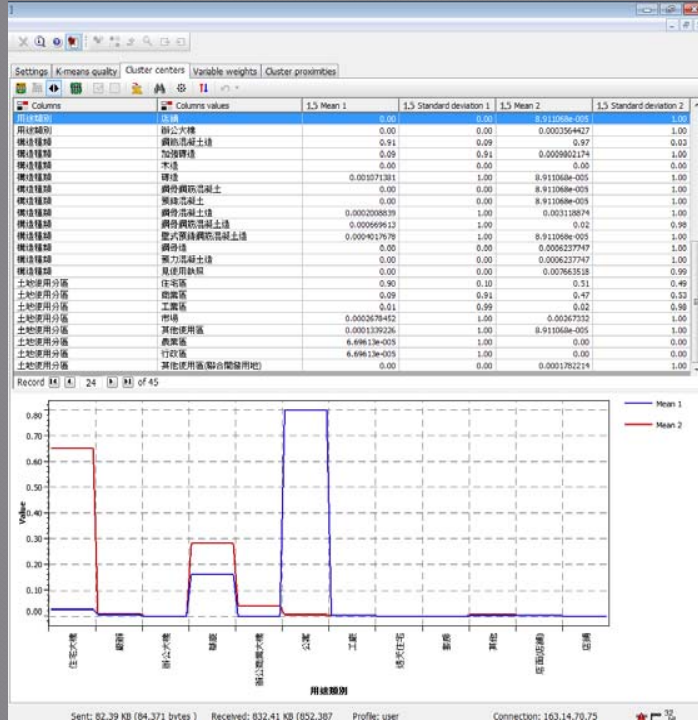
新北市



群集分析-用途類別

台北市

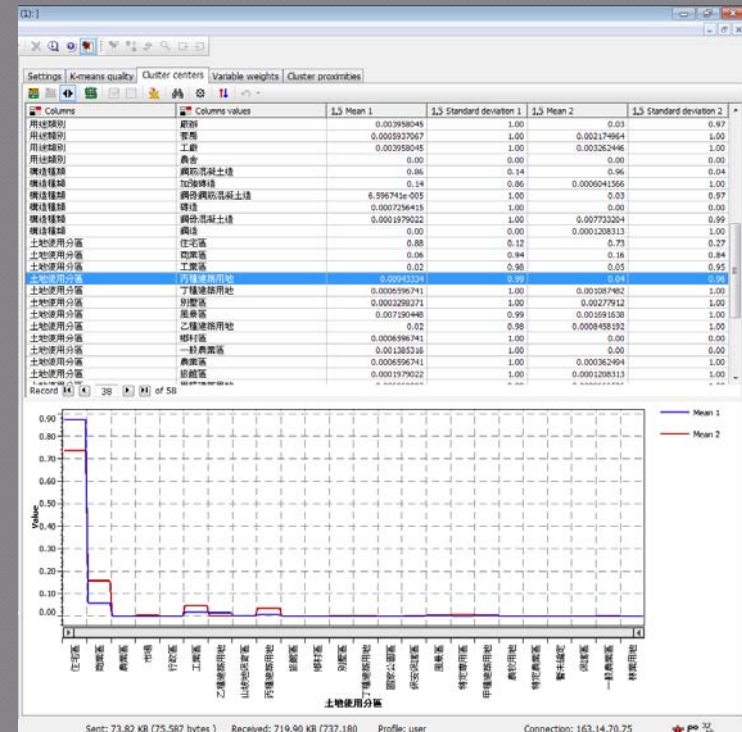
新北市



群集分析-土地使用分區

台北市

新北市



群集分析--結論

- ◎ 想要投資房地產的人們可以透過這樣的群集分析找到符合自己交易標的的群集，來掌握購屋或出售的時機

預測分析

Prediction

Prediction

◎ 訓練資料集:

- 89~100台北縣市房屋交易資料和股市資料

◎ 測試資料集:

- 101年的房屋交易資料
- 信義房屋靠近捷運站10分鐘
- 實價登錄網站上房屋交易資料

模型參考因素

房地產著名學者張金鶚教授

房仲業者資訊



房價預測Model

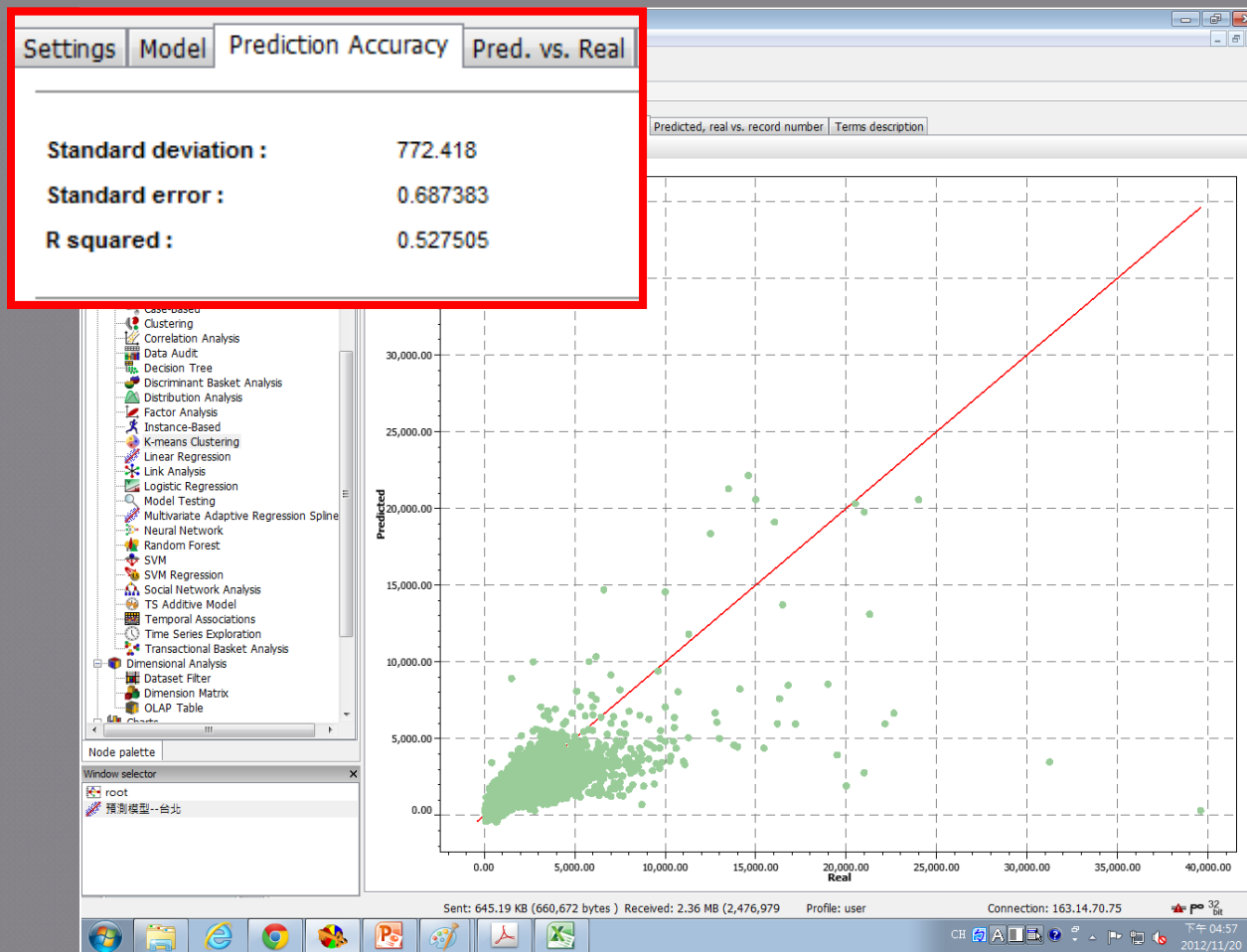
◎ 台北市

Settings	Model	Prediction Accuracy	Pred. vs. Real	Predicted, real vs. record number	Terms description
Prediction model: 總價 = -794.770 +5.83316*路寬 +3.09451e-008*移轉層數 +34.2970*移轉房屋面積 +1.49599e-008*房貸利率 -0.0495255*發行量加權 +4.92759*營造建材類 -3.50697*屋齡					

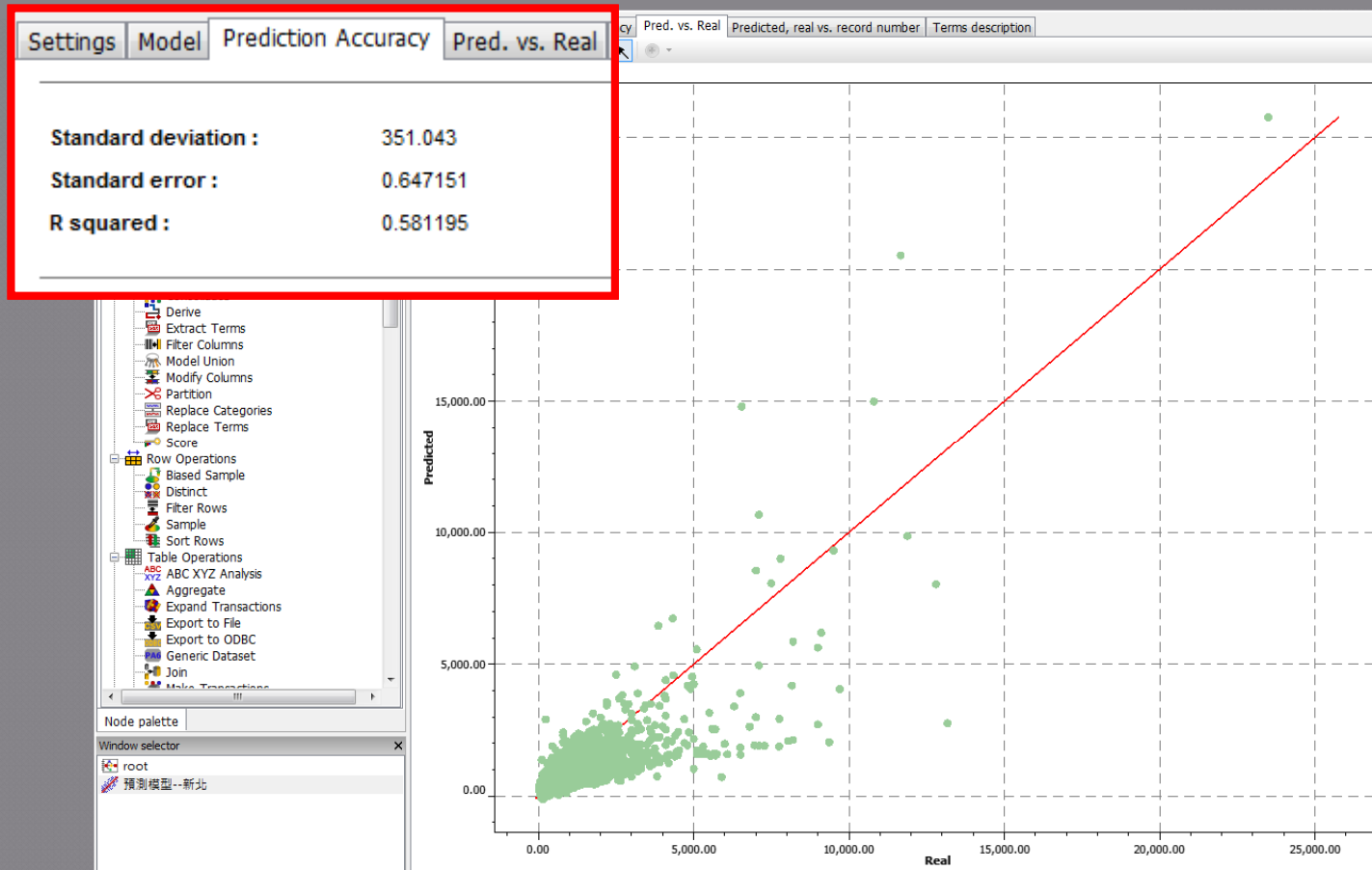
◎ 新北市

Settings	Model	Prediction Accuracy	Pred. vs. Real	Predicted, real vs. record number	Terms description
Prediction model: 總價 = -191.155 +5.72677*路寬 +3.24085e-009*移轉層數 +13.8758*移轉房屋面積 -93.2609*房貸利率 +1.69448*營造建材類 -2.07819*屋齡					

台北市



新北市



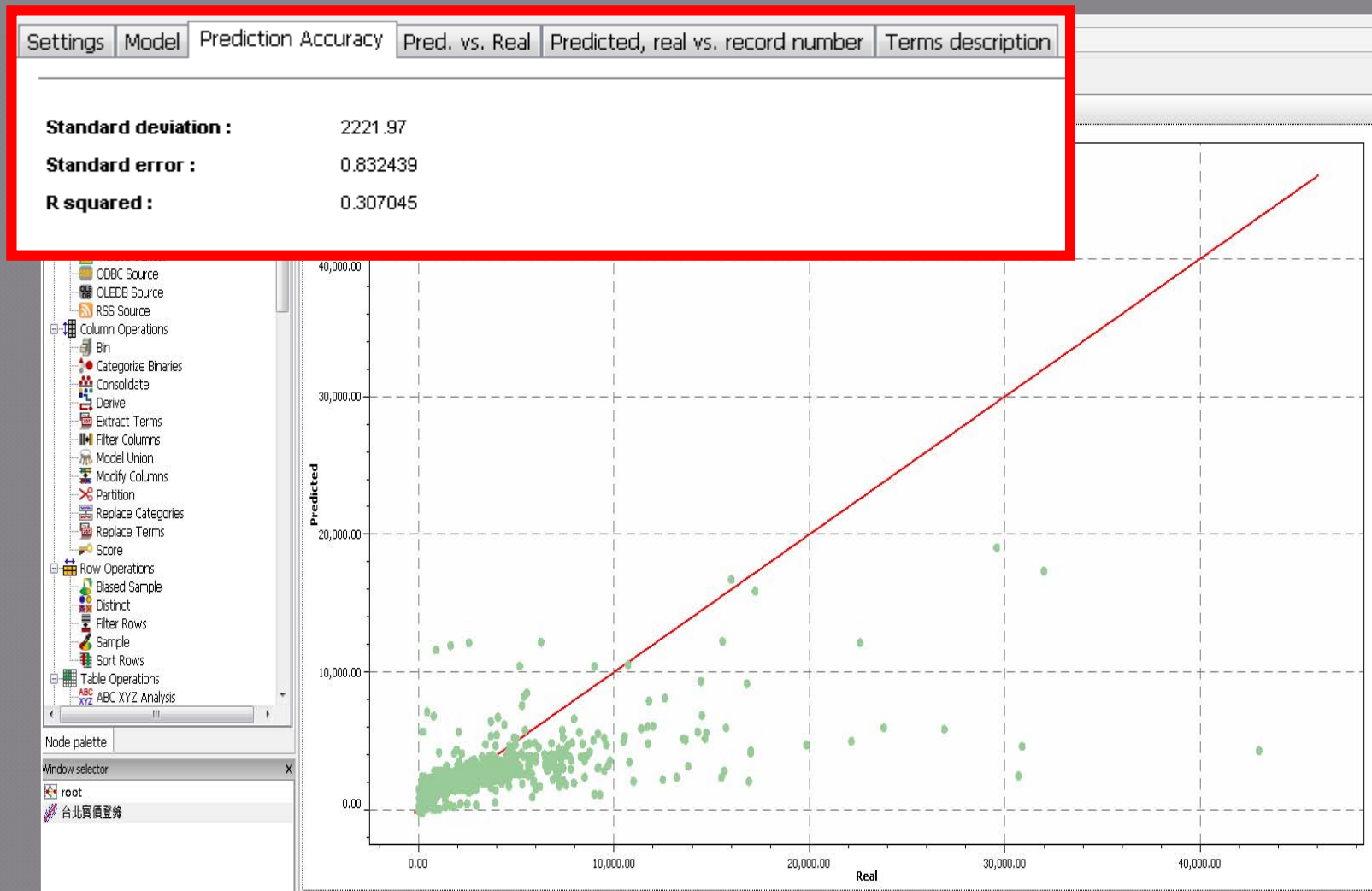
預測模型

◎ 相關係數 r ：

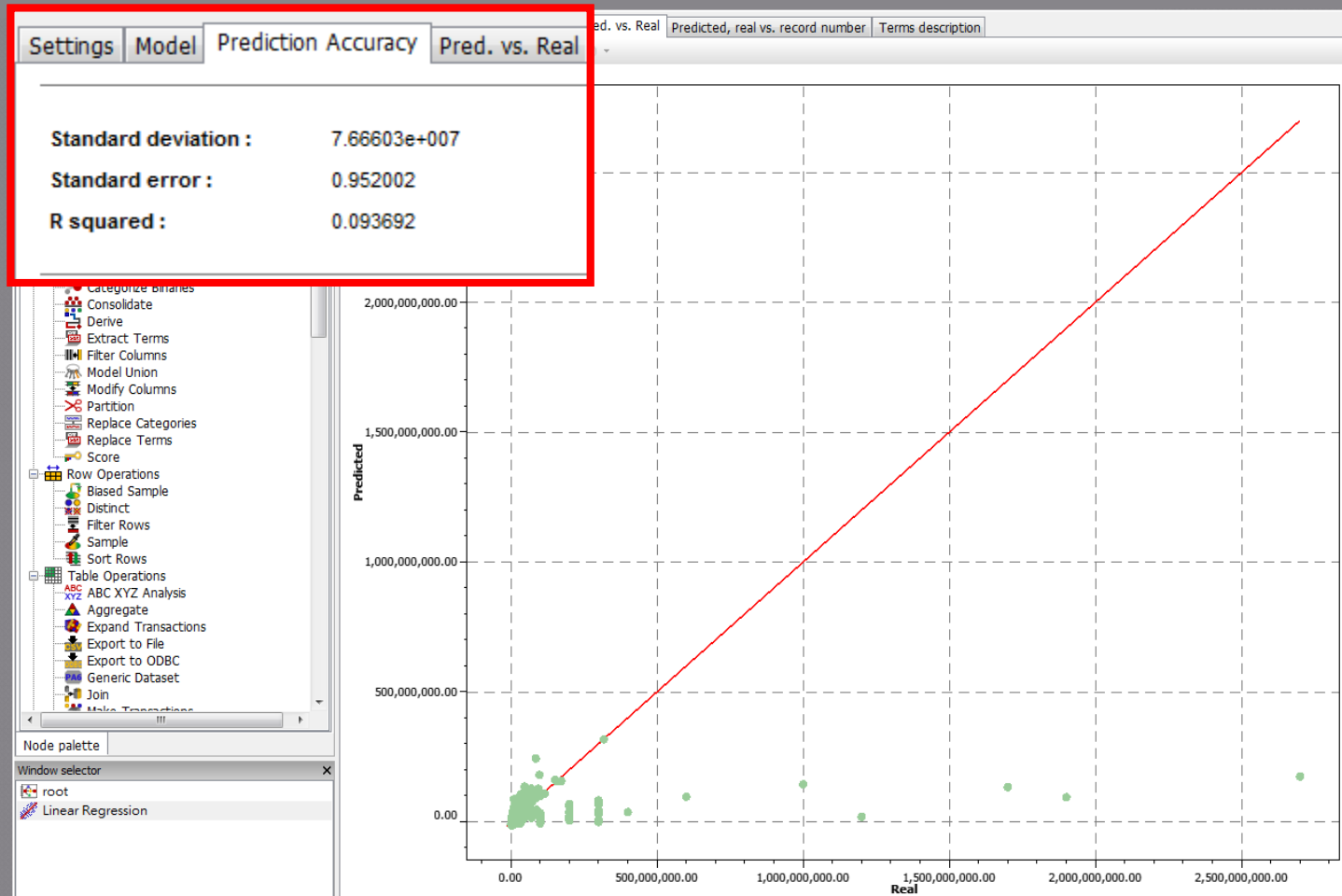
- $|r| > 0.6$ ，為高度線性相關
- $0.3 < |r| \leq 0.6$ ，為中度線性相關
- $|r| \leq 0.3$ ，為低度線性相關

◎ 證明預測模型的效力，可以提供資訊來幫助決策

預測模型—實價登錄



預測模型—捷運站附近



預測分析—結論

我們的預測模型具有相當效力做為決策之用

◎ 實價登錄:

- 數量較少
- 對於房地產市場有相當的影響

◎ 捷運站附近:

- 房屋附近有重大的影響設施也會影響預測效力

誌謝

感謝

趙景明教授的指導以及技師的協助

特別感謝

