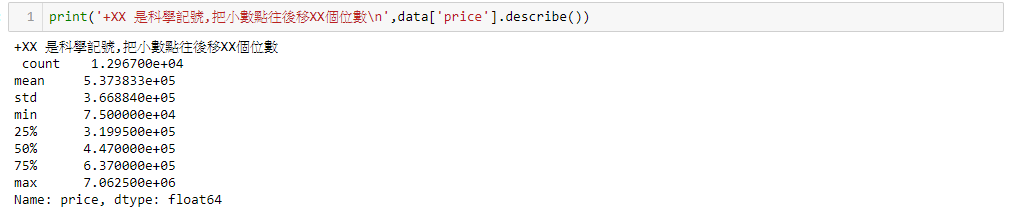
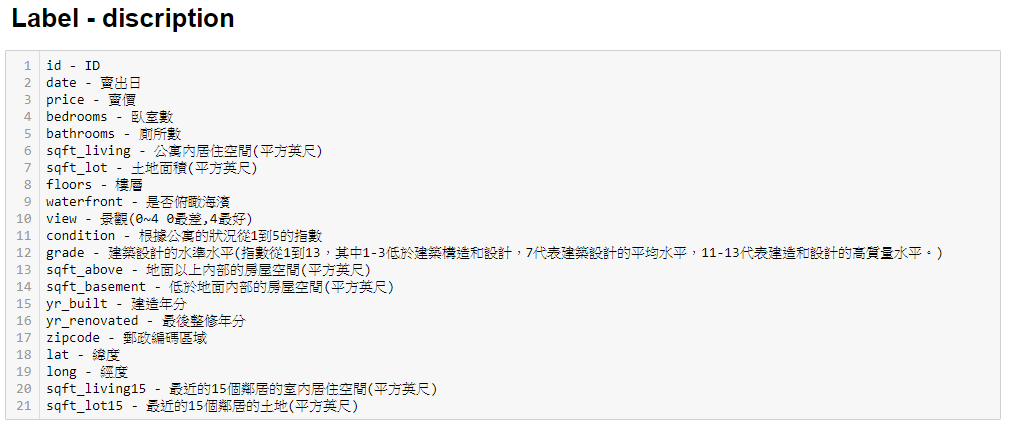
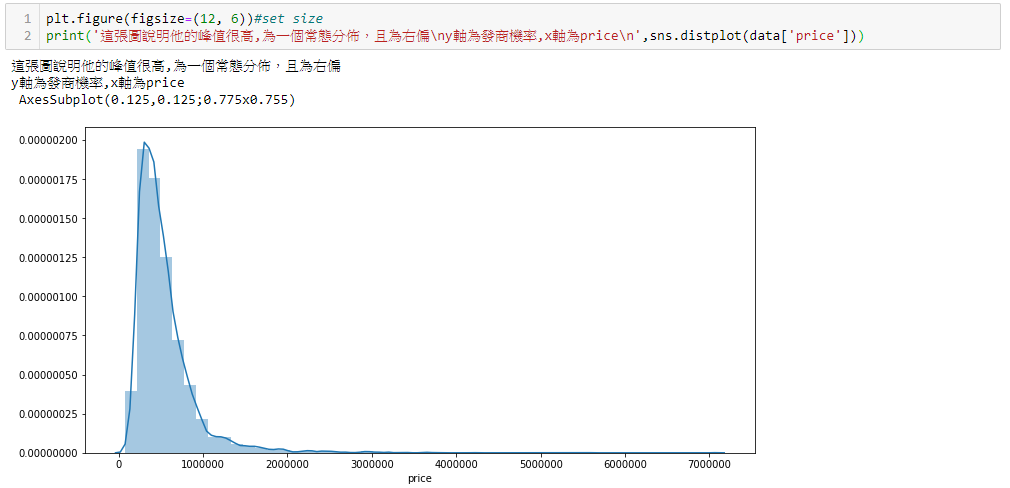
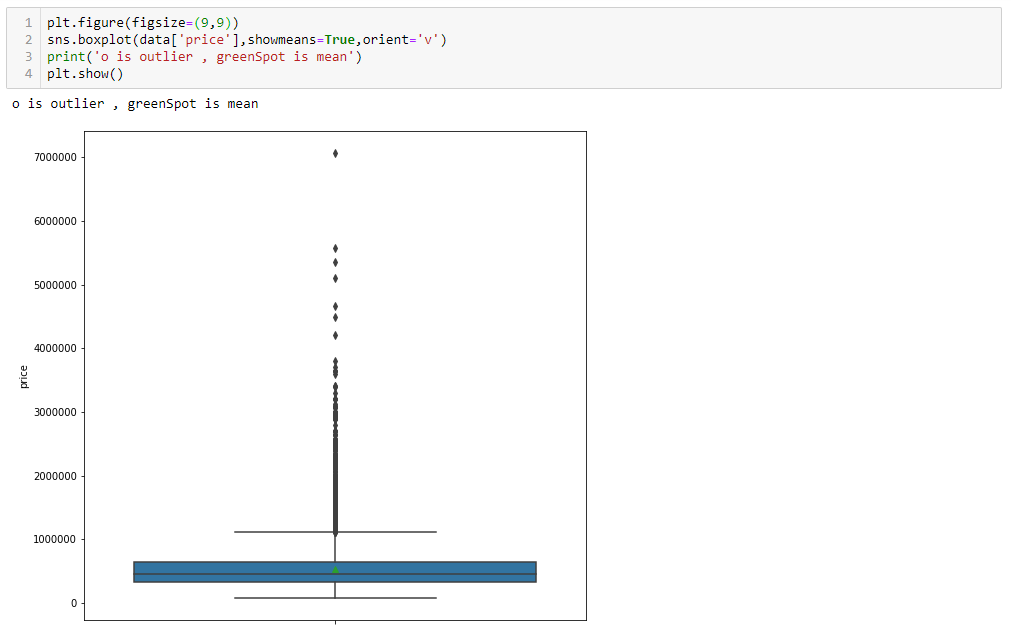
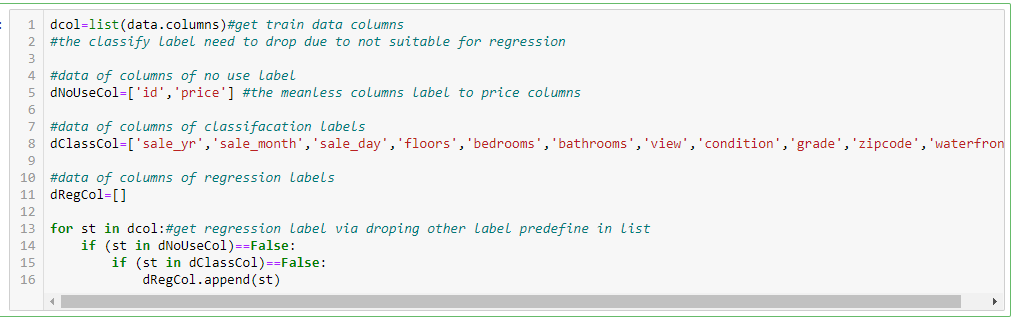
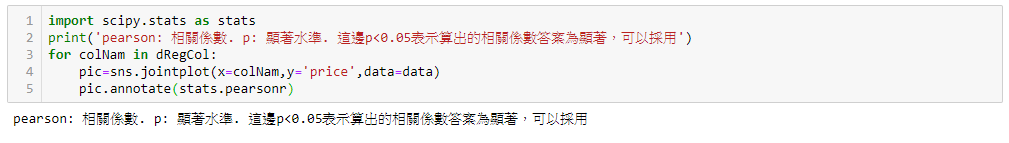
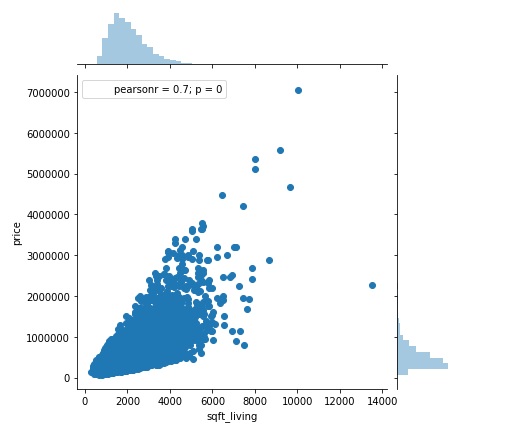
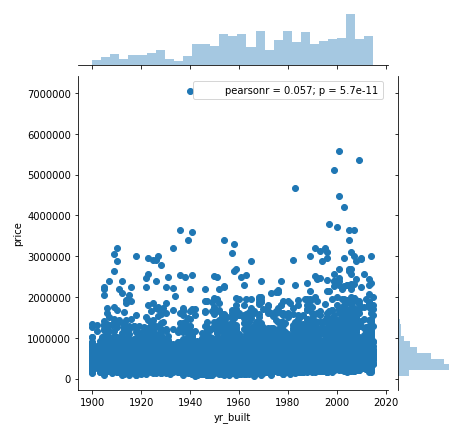
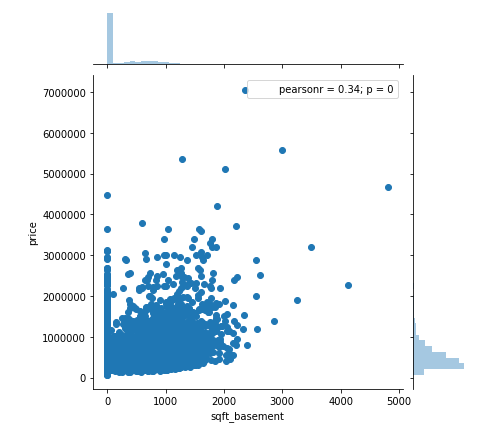
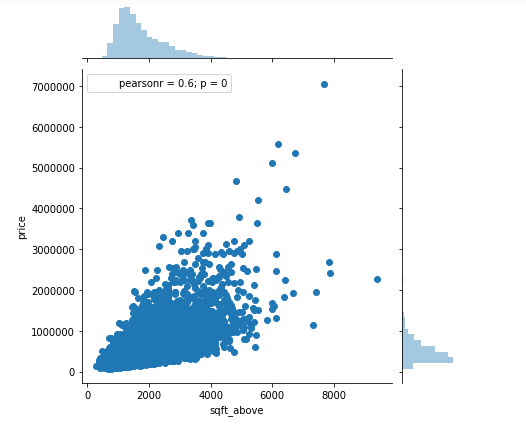
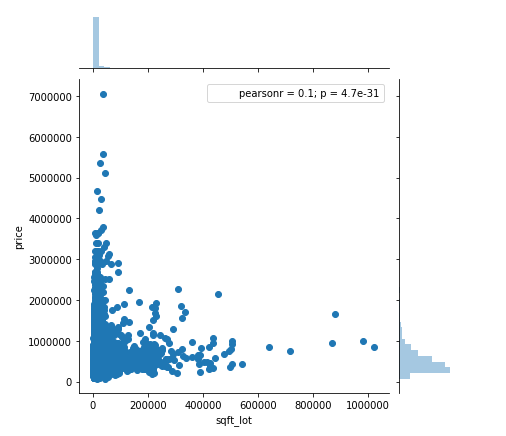
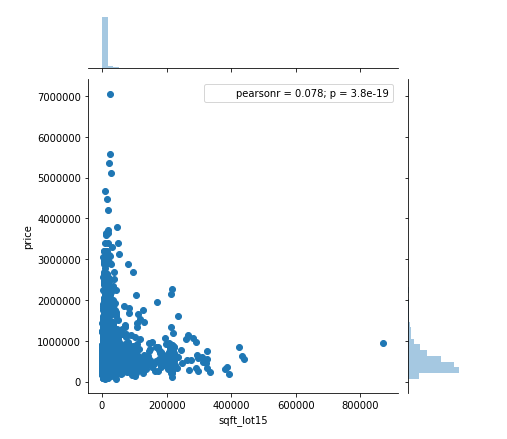
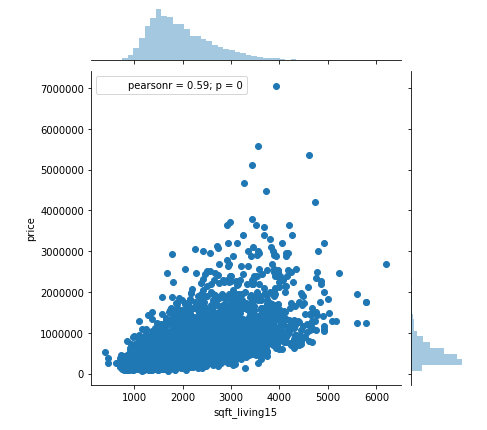
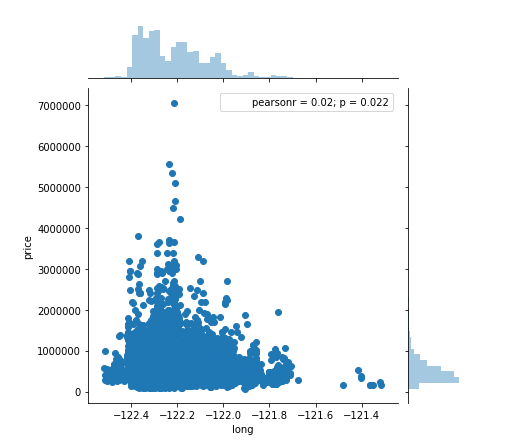
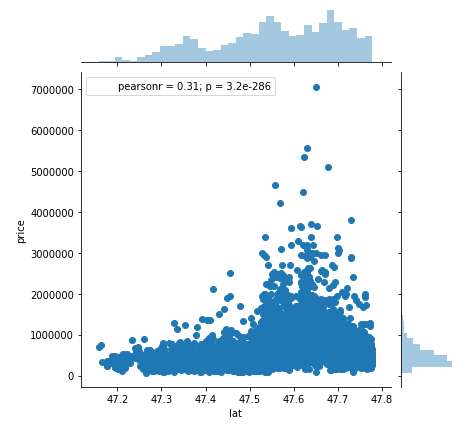
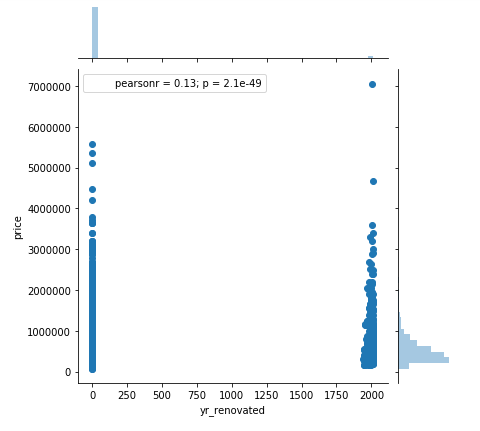
資料類別與資料型態­­­­­­­­

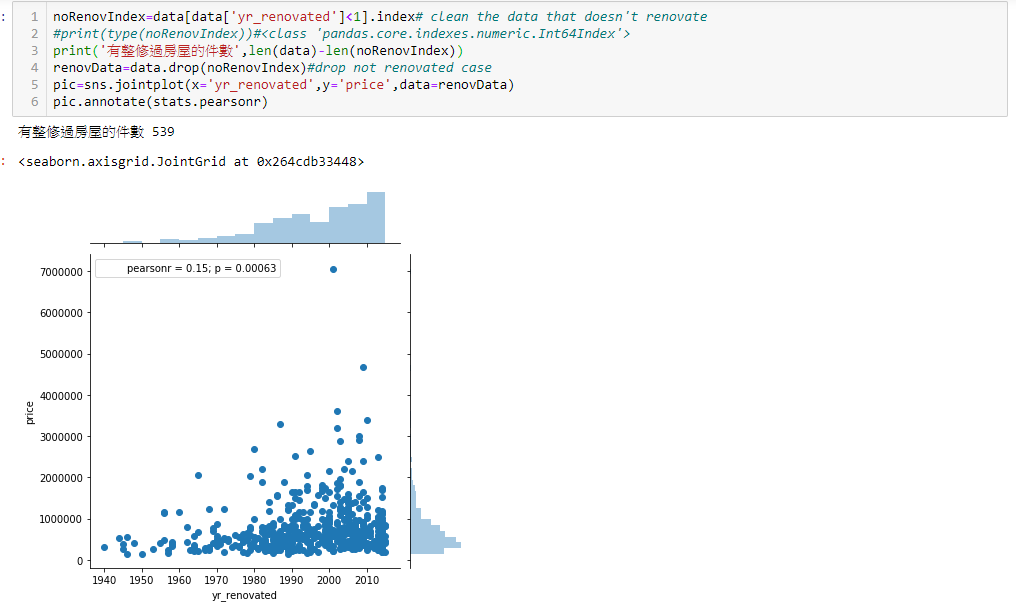
房價的分布

長條圖分布圖可以看到我們的資料是有集中性的。從我們的盒型圖也可以剛好驗證我們對於房價分布的推論，離群值除外。

分類屬性列出數值屬性與房價關係 列出可與房價做線性回歸評估類別的散佈圖，與Pearson相關係數。

下圖為yr\_renovated和price的線性分析圖，有些可能沒有裝修房子，所以後面會在分析有裝修的房子案件。u

分析有yr\_renovated裝修的房子案件。移除沒有裝修房子的紀錄。

分析數值屬性與房價的相關性

一般研究認為，相關係數0.3以下為低相關，0.3~0.7為中等相關，0.7以上為高度相關。以下的散佈圖皆採此標準為評估，並無條件捨去到小數第一位。

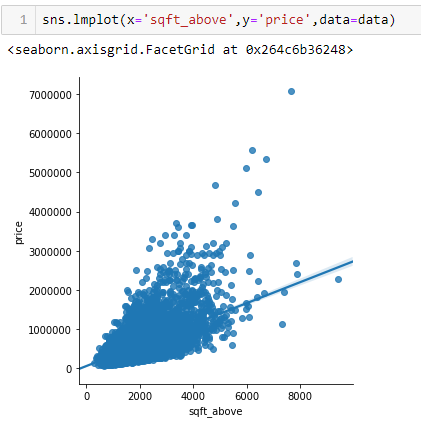
列出有超過Pearson相關係數0.7的屬性，表示與房價具有高度相關信。

公寓內居住空間對房價有高度相關信，事實上也是如此，買方對於房子居住空間的需求是明顯的，因為他關係到買方以後的生活模式是一個很重要的特徵。



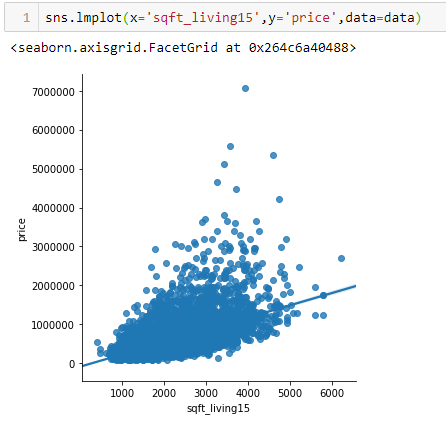
列出Pearson相關係數界在0.3~0.7的屬性，表示與房價具有中度相關信。

地面以上內部的房屋空間也具中度相關信，因為本身屬性也涵蓋一定的室內居住空間，所以與Sqft\_living也會有相關性。



最近的15個鄰居的室內居住空間有中度相關信，推論依據為同樣地段的房子，房屋大小也會近似，代表同一性質的人會住在這，附近鄰居居住空間通常為一致，不會有大好宅和小坪數混搭的比例出現，這顯得建商沒有做好評估社區與環境。

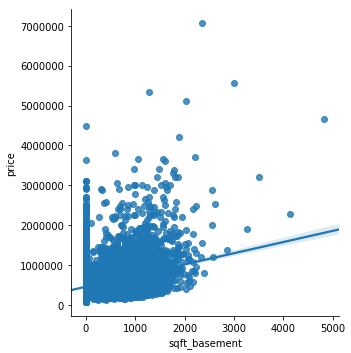




低於地面的內部房屋空間(平方英尺) 有中度相關信。

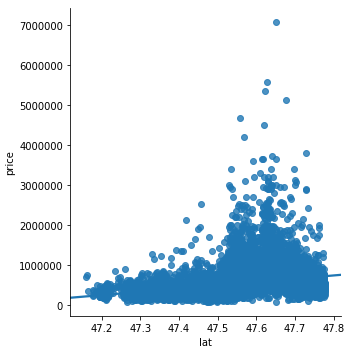
地下室也具有一定的引響庫戶的購買意願





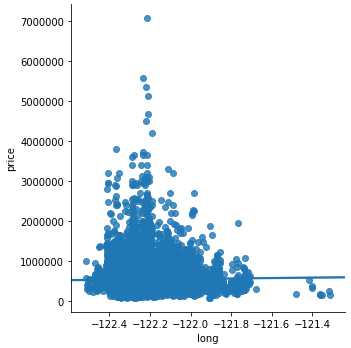
緯度比較特別，與地下室有著差不多的引響力，都為中度相關信。但緯度具有區間性，不會至太高，本次結果範圍是落在47.2~47.8，但經度又顯示很低的相關結果，可能為巧合或是有待查明。



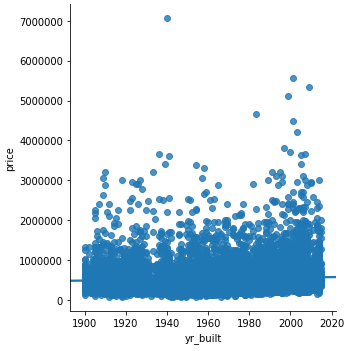


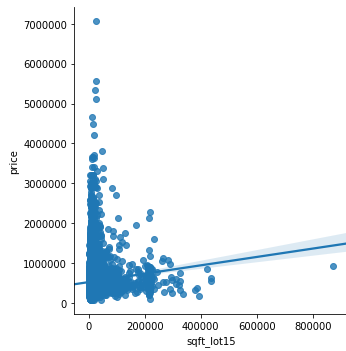
列出Pearson相關係數在0.3以下的屬性，表示與房價具有低度相關信。

經度相關係數顯示近乎為0，表示與房價沒有任何關係。從圖中也可以看到點散佈的趨勢區間與線並沒有明顯關係。

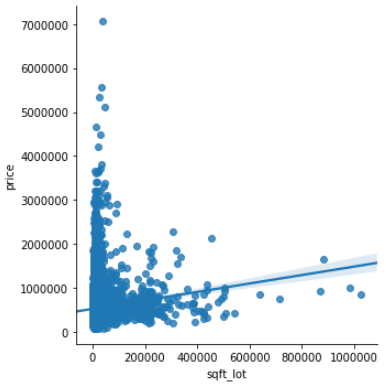
 

建造年分相關係數顯示近乎為0，表示與房價沒有任何關係。從圖中也可以看到點散佈的趨勢平均在年分上，但房價卻沒什麼改動。

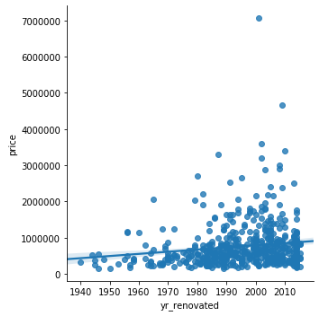
最近的15個鄰居的土地相關係數顯示近乎為0，表示與房價沒有任何關係。從圖中也可以看到多數點的散佈趨勢與線沒什麼關係。 

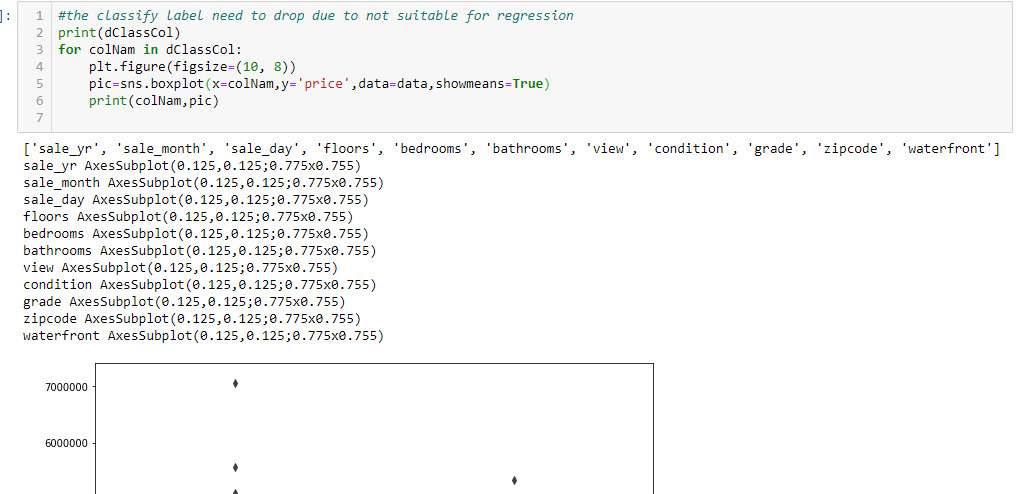
土地面積相關係數顯示近乎為0.1，表示與房價有低度相關性。從圖中也可以看到多數點的散佈趨勢與線沒什麼關係。

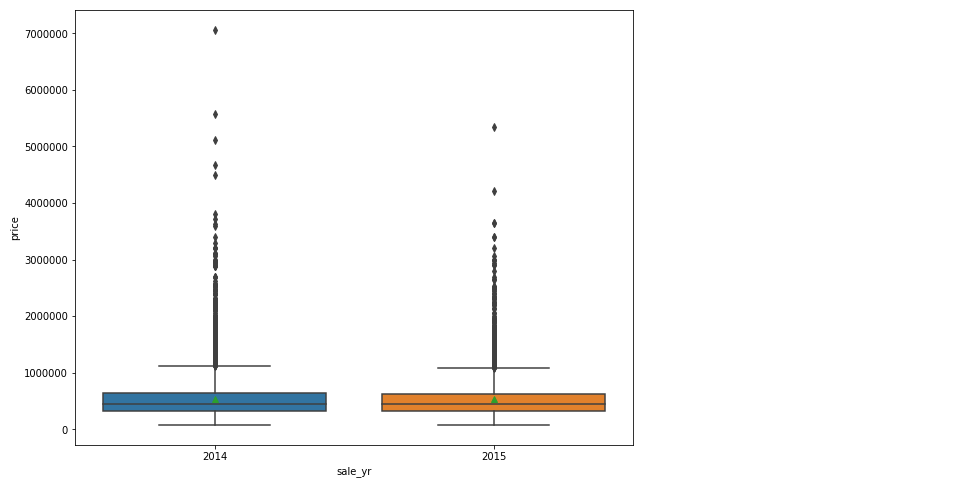


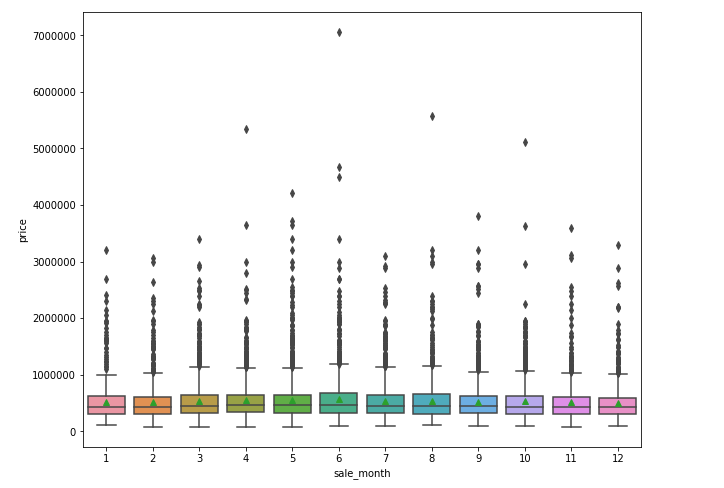
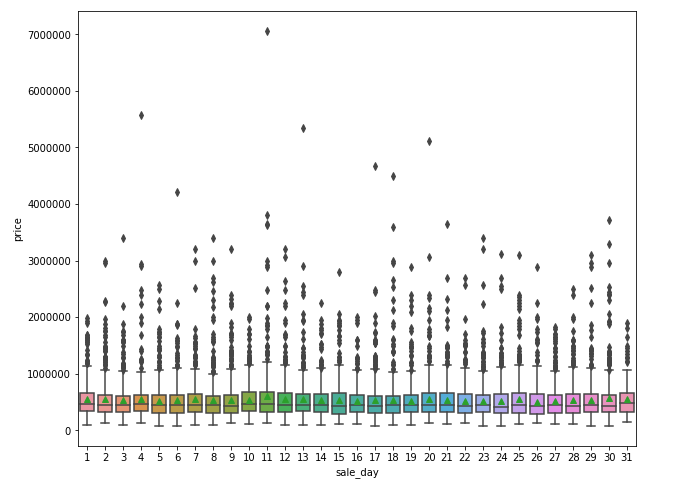
分析有裝修的房子案件。移除沒有裝修房子的紀錄。但最後還是從相關係數中可以看出，有翻修過的房子與房價沒什麼實質上的差別，相關係數很低0.15，表示與房價有低度相關性。從圖中也可以看到多數點的散佈趨勢與線沒什麼關係。

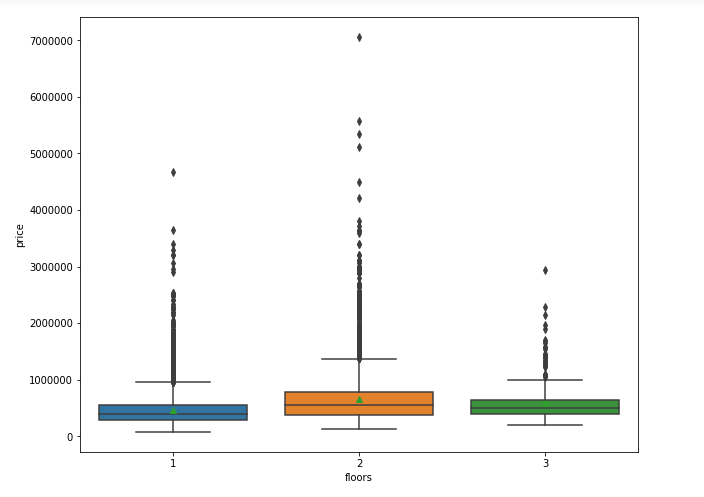
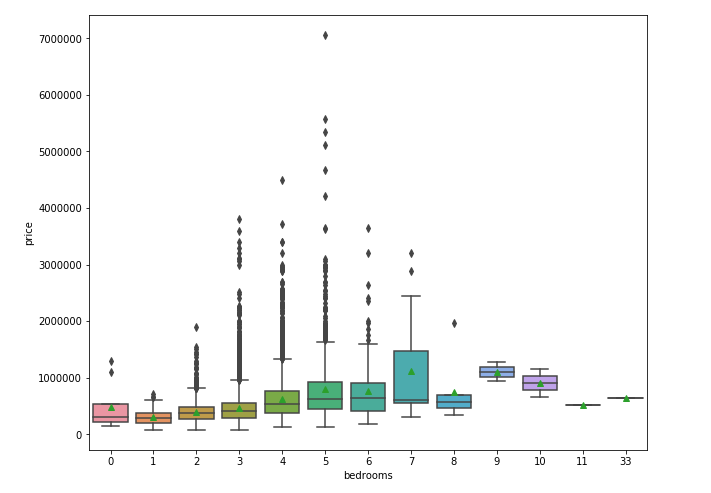


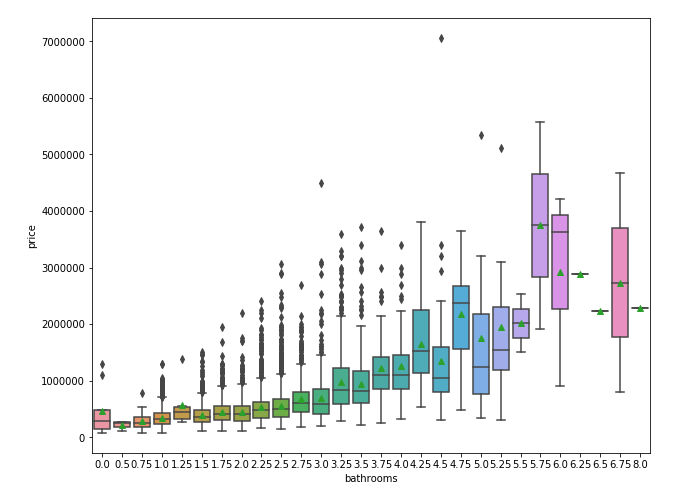
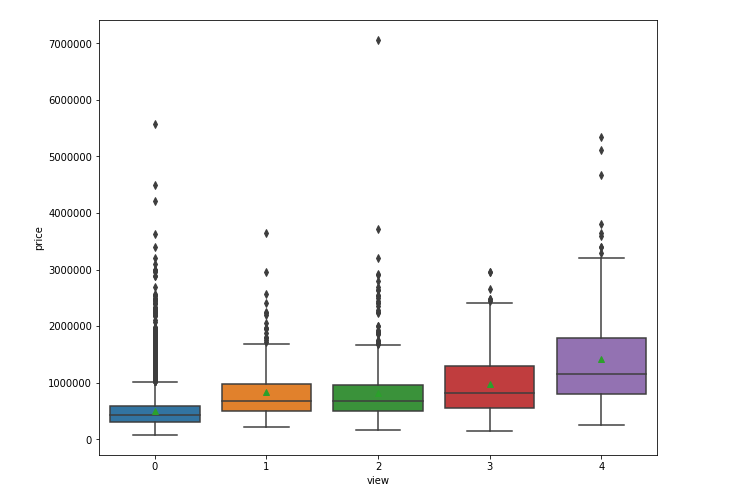


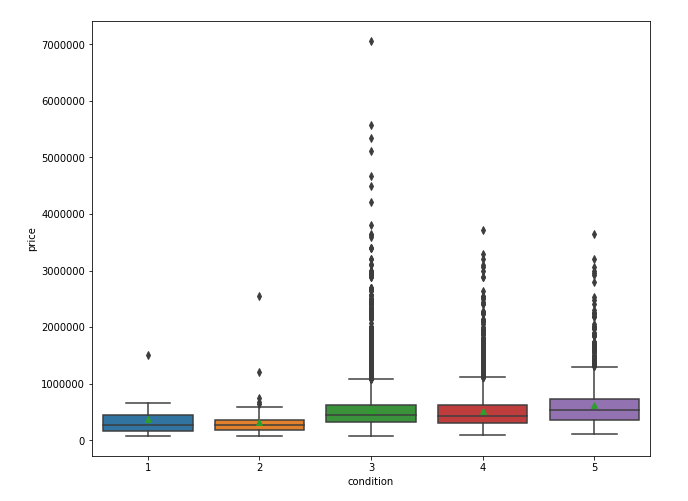
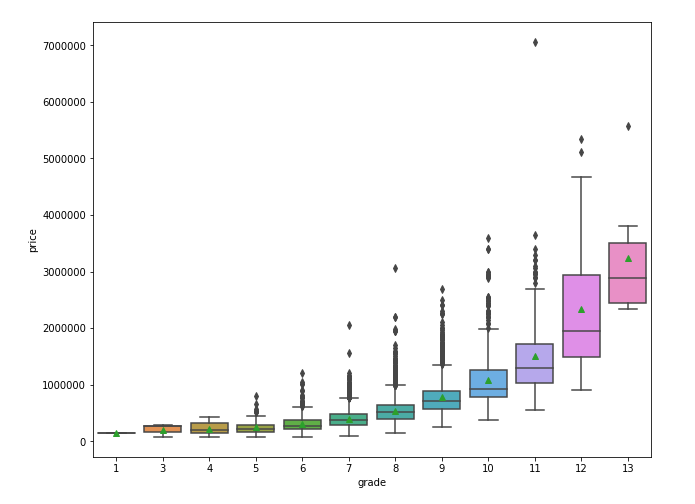
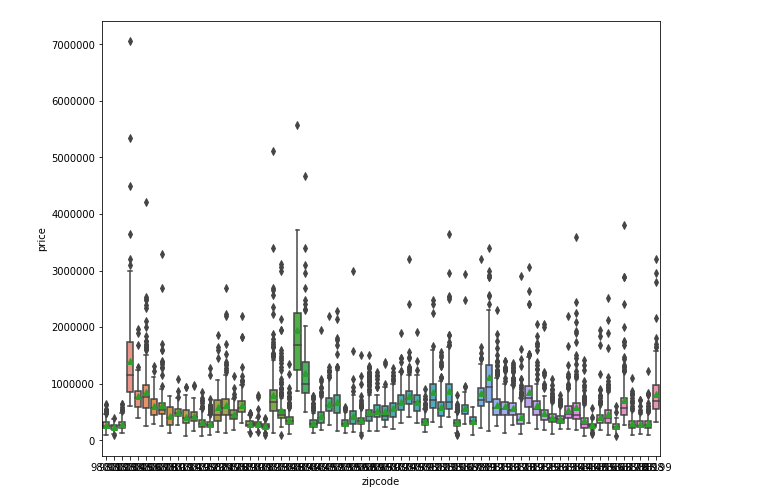
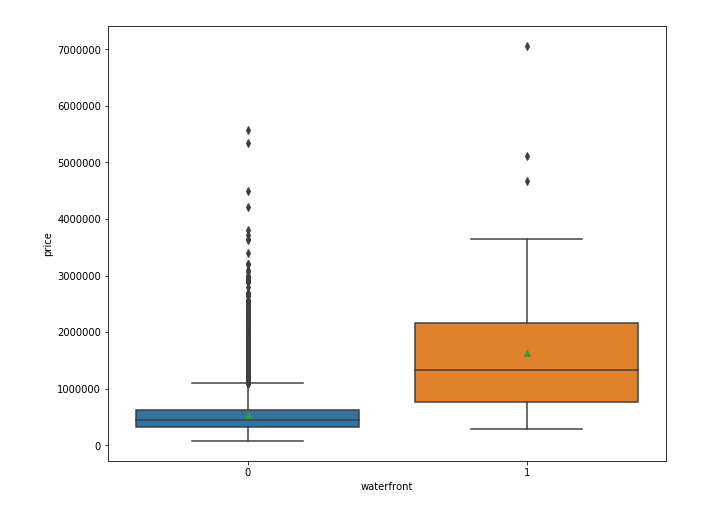
列出類別屬性與房價關係

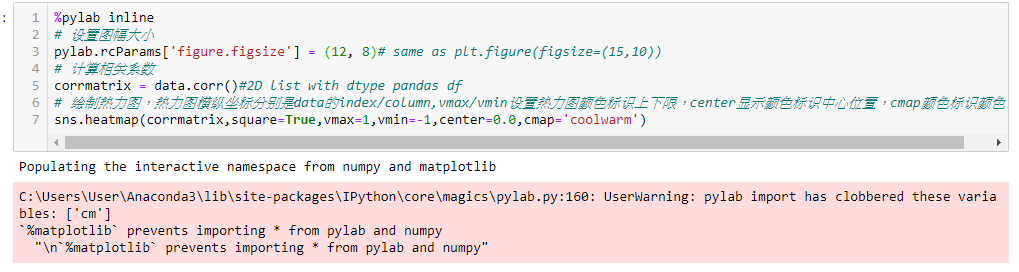
說明兩年分四分位距、中位數、mean，都差不多，房價與年份無關。離群值ˋ列入考慮

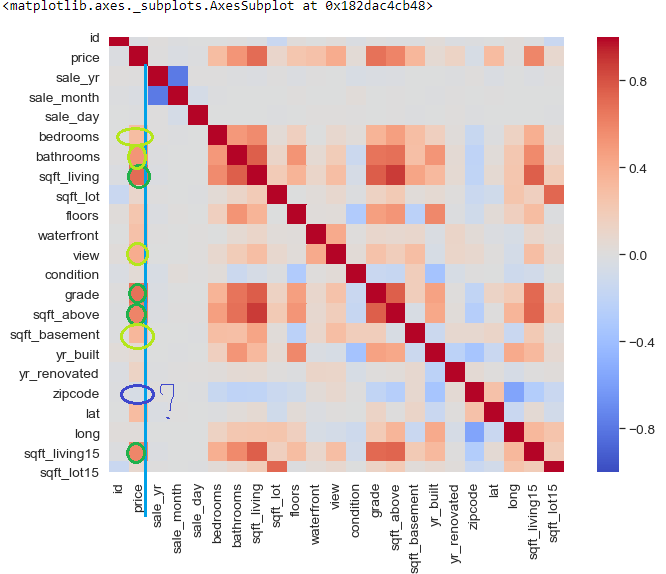
說明四分位距、中位數、mean，都差不多，房價與月份無關。離群值ˋ列入考慮  說明四分位距、中位數、mean，都差不多，房價與月份無關。離群值ˋ列入考慮

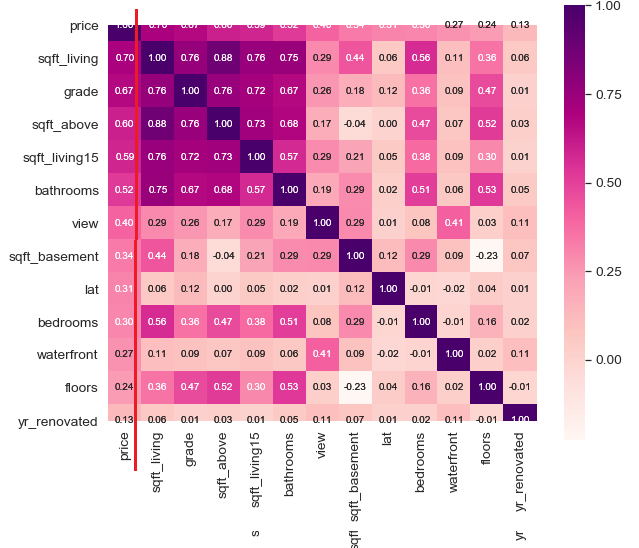
2F的和型圖較長，且max值長出一節，表示大家偏愛2F，而3F又降會與1F差不多的座落位置，可推出數據不具有大的相關性，但可以知道有坐落性。 臥房越多，房均價越高，有一定的相關性。 11、33臥房圍籬群值可以不參考，因為圖可以看出IQR、MEAN都在一點上。

洗手間越多，房均價越高，有一定的相關性。 景觀越好明顯房價就越貴，且具有高度相關性，和型圖也展現遞增房價上移的現象。 

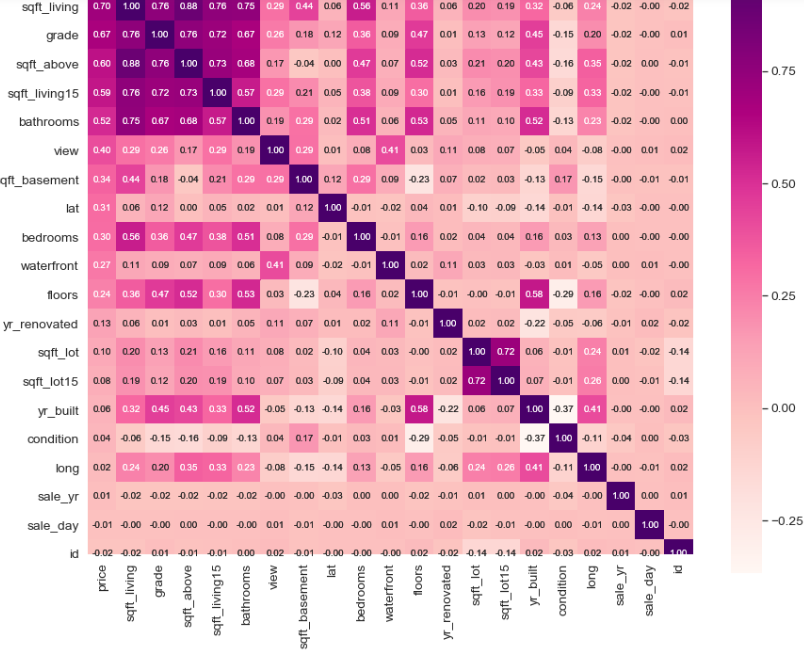
公寓指數越好房價就越貴，且具有一定程度相關性。  建築設計水準水平明顯具有高度相關性，甚至有接近指數成長的現象。 郵遞區號編碼圖形散佈不平均，因為郵遞區號是依地段一一編下去，所以可以從圖中看到的峰值看出，峰值通常為較貴，而越靠近峰值的地段都有明顯往上攀升的現象，所以可以推斷地段的好壞從這張圖可以明確地具有重大影響，甚制可以因為地段而拉高房家的平均一節。 可以看的到海拚的地方也明顯較貴，所以也對價錢相當的引響力。



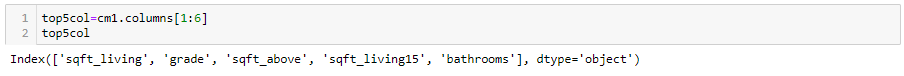
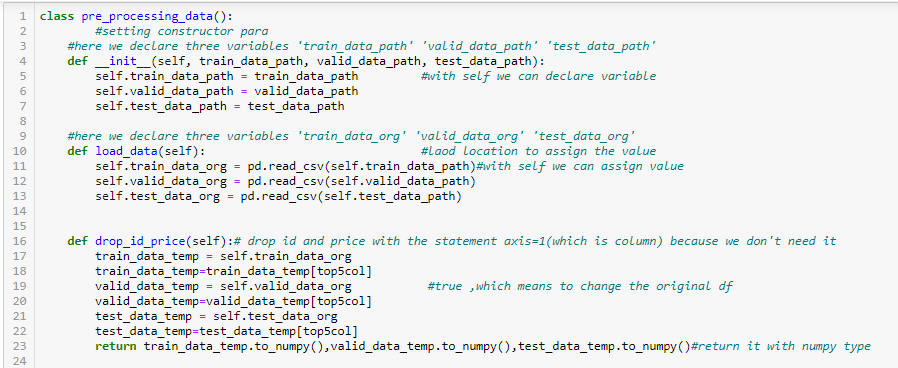
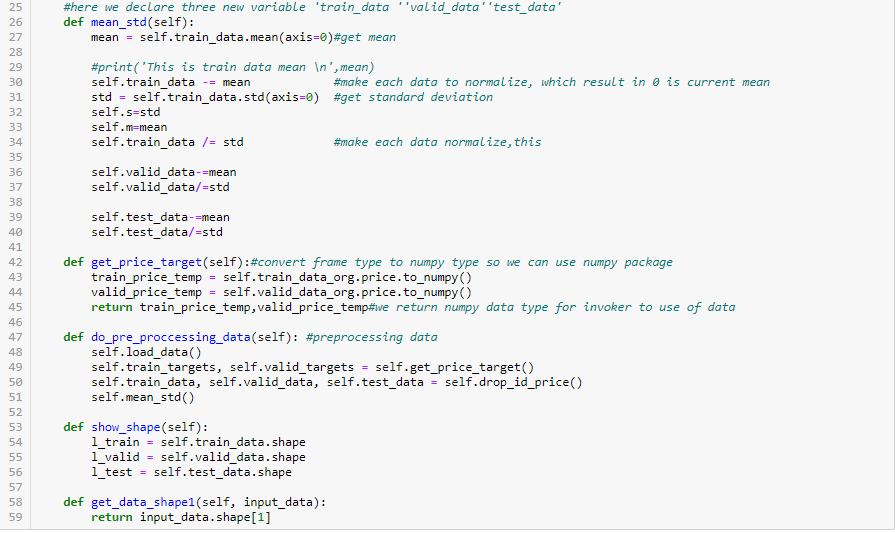
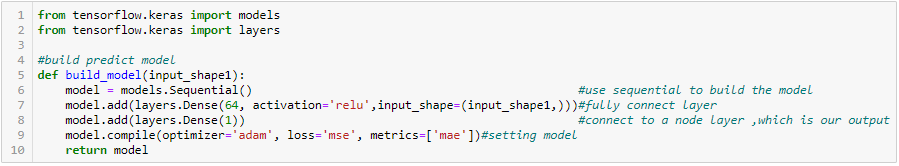
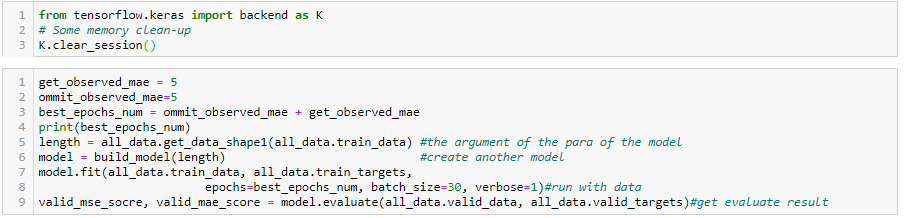
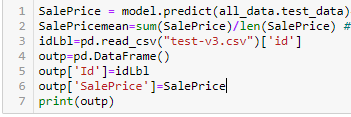


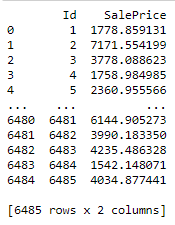
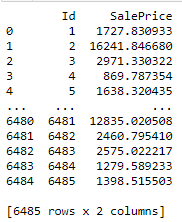
深綠色畫圈圍相關度較高的影響因子，其中sqft\_linving、grade、sqft\_above、sqft\_living15，我們都有在前面提到，淺綠色畫圈bedroom、bathrooms、sqft\_basement、view，依樣有在前面提及，深藍色畫圈，zipcode是因為有多峰值的原因，多峰值也代表特定區域，所以zipcode的編號是特定區間，而非隨著zipcode大小與房價成正比，所以在此當然不會有相關。可以看到相關性前幾項依序為sqft\_living、grade、sqft\_above、sqft\_living15、bathrooms、view、sqft\_basement。圖中，從三角形範圍，可以明顯地得知， 們彼此因子相互也是有一定的關係，且最低的相關係是0.57，代表bathrooms與sqft\_living，相互關係為(彼此影響對方數值成正比的性質)0.57，再看到他們彼此對房價的影響，0.52和0.59，也都對房價有著不小的影響力；在看最高的相關係數是0.88(sqft\_living與sqft\_above)，他們與price的相關係數，分別是0.7和0.60，由此可以推論結果，當房價越貴實，相關係數與房價越高的欄位數值也會越高，且欄位數值互相關聯的話也會使彼此呈現正相關的性質。EX 1.sqft\_living與sqft\_above，公寓內居住空間與地面以上內部的房屋空間當然會呈正相關，他們也會與房價呈正相關。

2.lat與price，雖然他們相關係數為0.31，從我們前面推論的結果，他也必須要與price相關的欄位有一定的潛在相關性，但圖表顯示她與任何欄位屬性皆沒有一定的相關性，所以我們可以推論，經度屬性會與房價有這麼高的相關性，很可能是純粹巧合。

下面畫線的也可以當作相關的依據(0.5)。

使用相關係數最高的前五項欄位，來當作我們的訓練資料、驗證資料與測試資料。

   初始化資料，與前處理 建模 訓練模型 訓練結果

左圖為TOP 5 COLs，右圖為只有drop ID與salePrice兩個欄位，也就是用21個欄位去預測的結果。