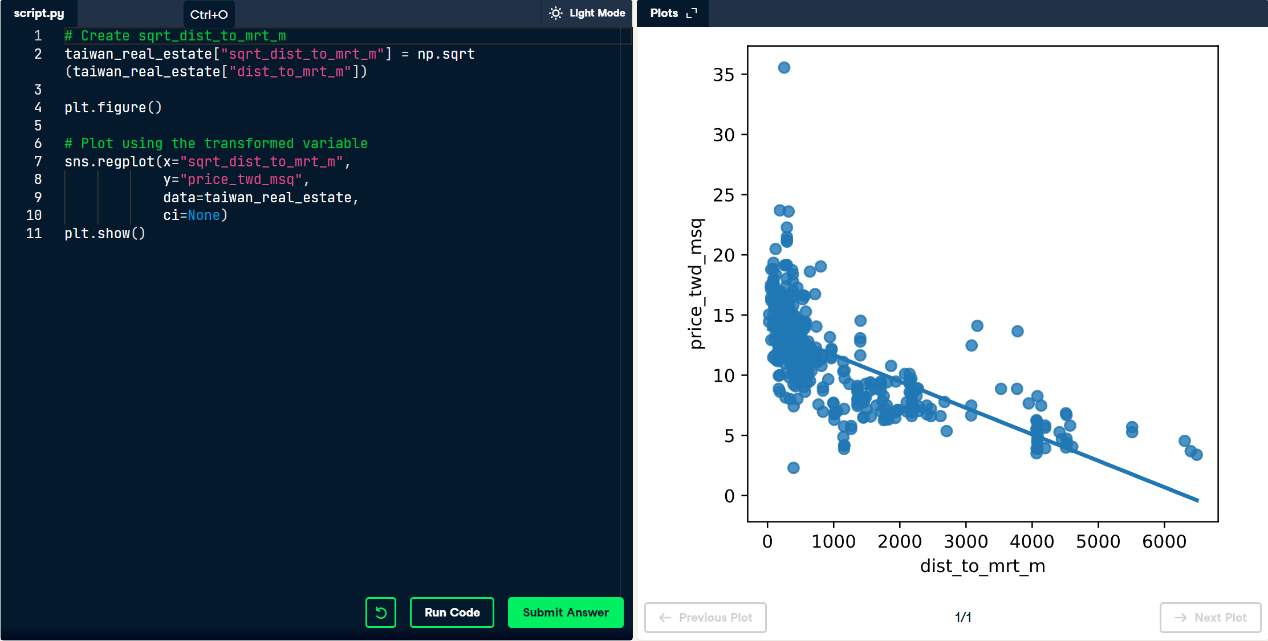
Data Modeling 數據建模

(sns regplot&scatterplot, statemodels.formula.api)

The model before tansform 模型轉換前



將x軸用np.sqrt平方根後，數值修正，得到正確的分布

The model after tansform 模型轉換後

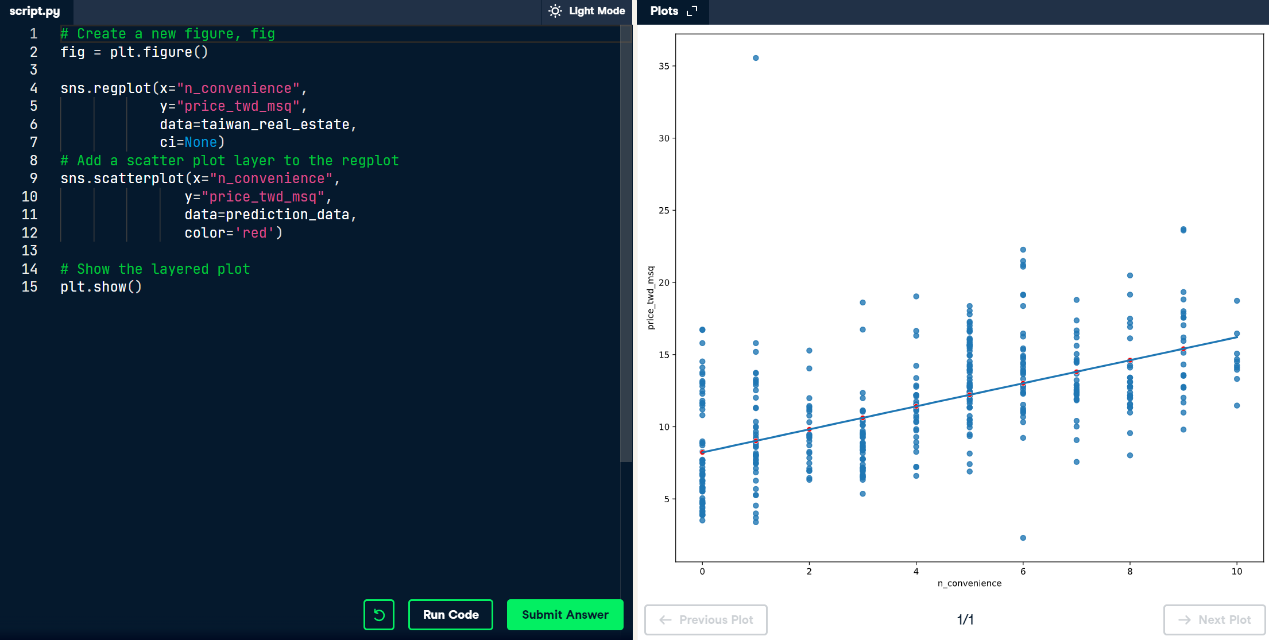


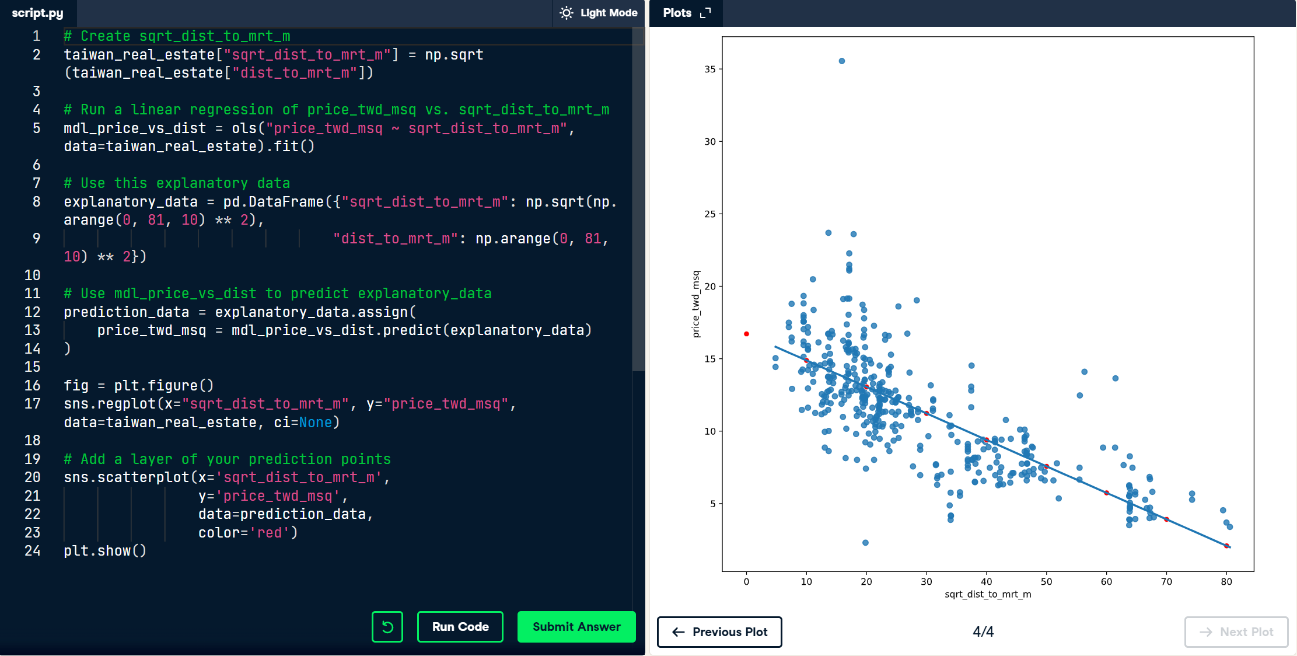
The Comparison of Original and Predict model

利用上述模型，與預測模型(線上紅點)做比較，觀察結果跟預測數值差異

紅點皆在點上，表示預測值與實際結果相符

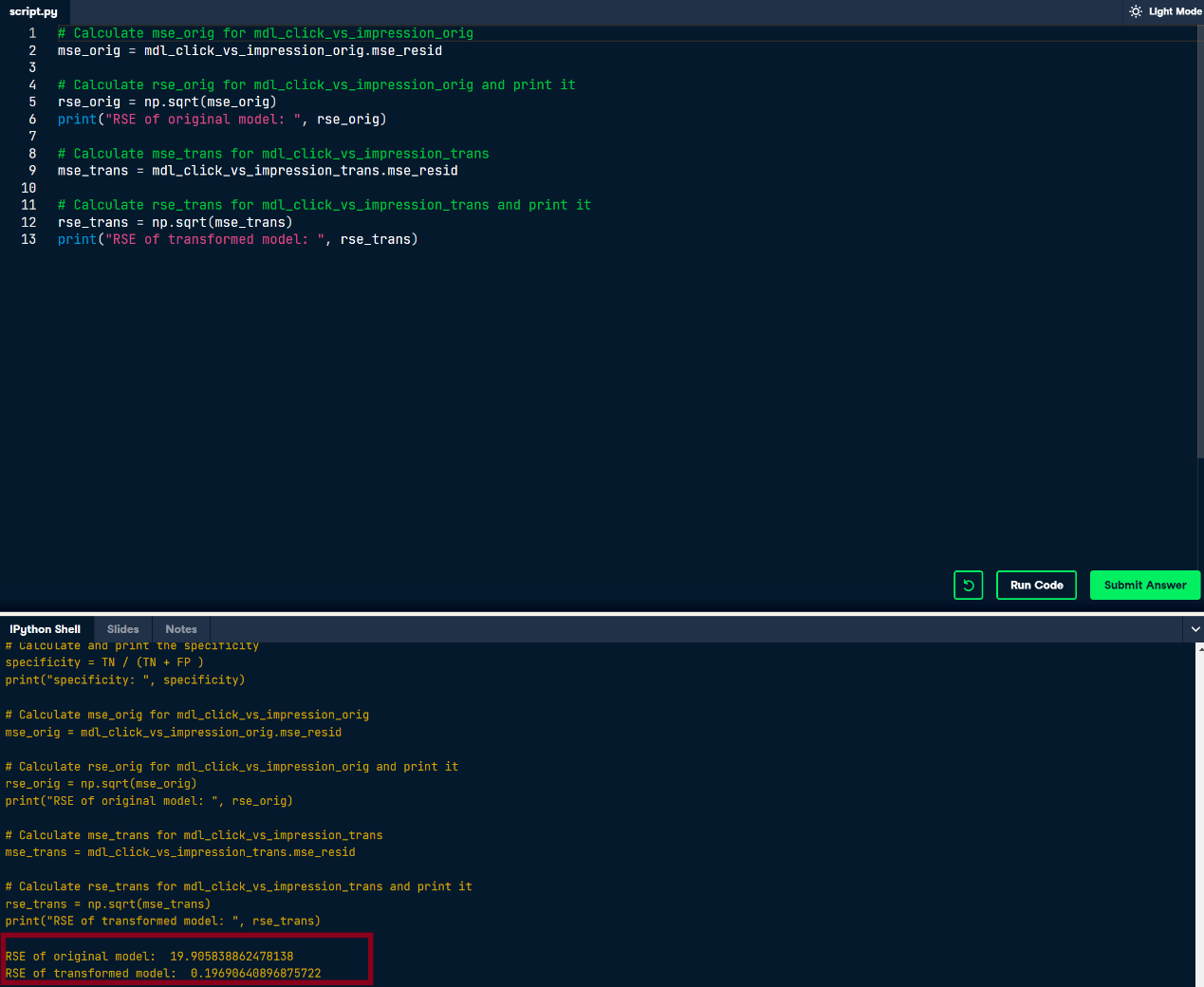
當周圍的便利商店數量多，房地坪數價格會相對高





RSE of original model and transformed model

RSE在模型轉換後，值變小，代表模型準確率提升

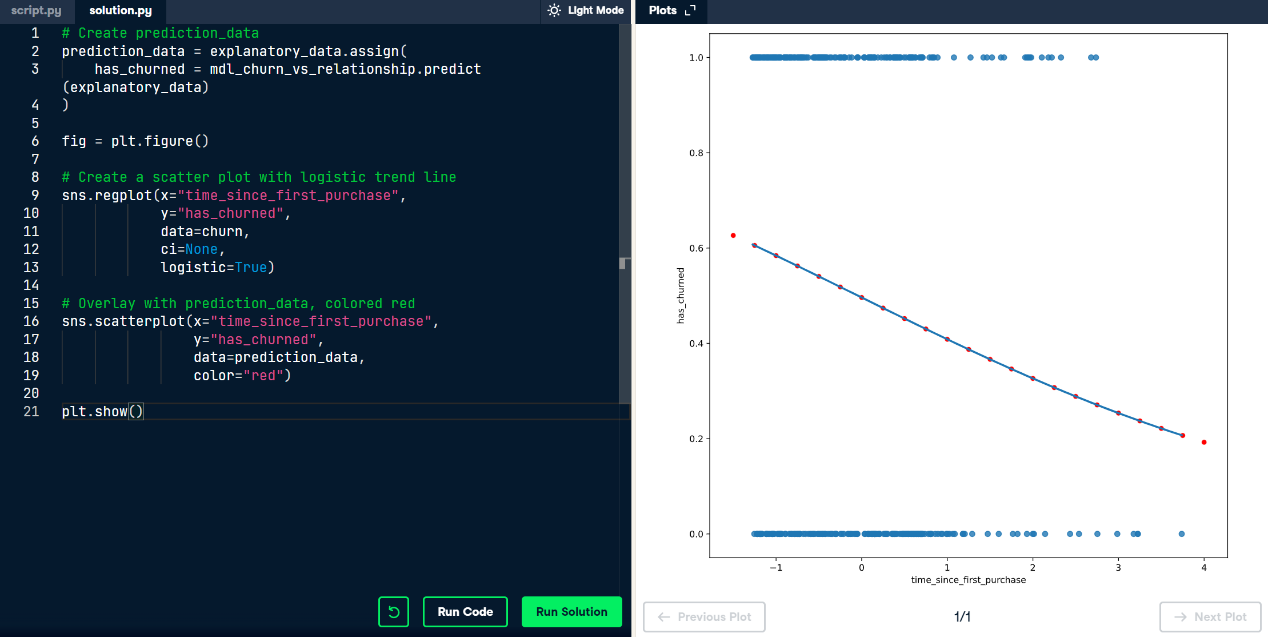


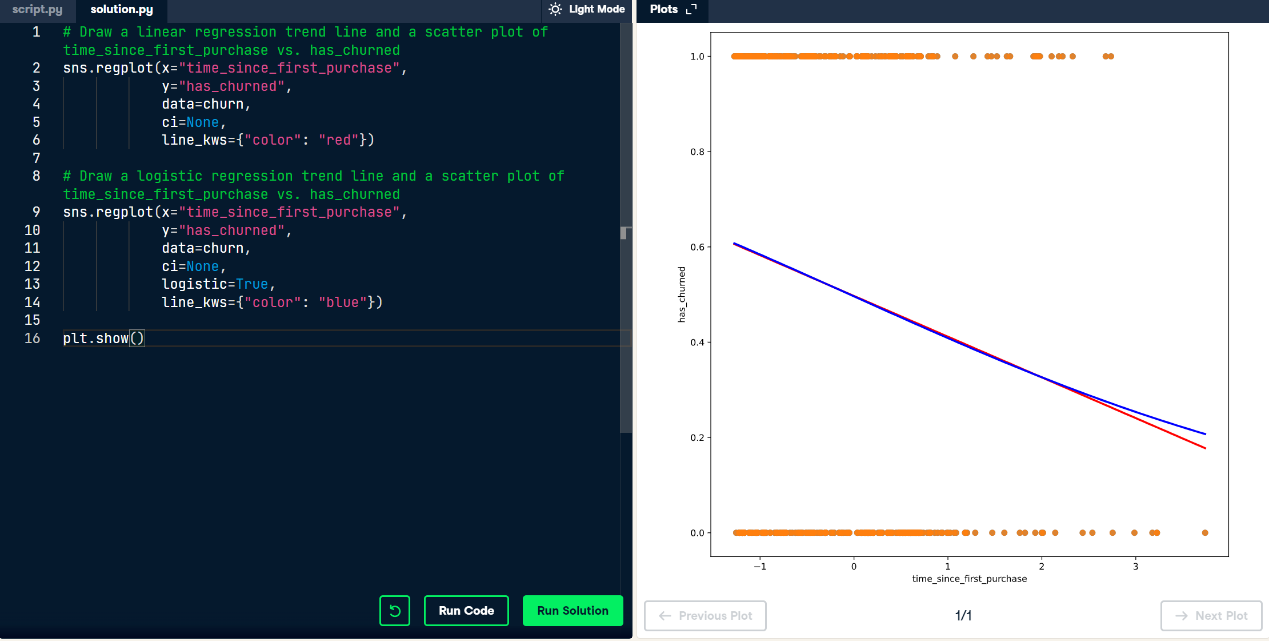
The Comparison of Original and Logistic Predict Model

預測模型(藍色)在logistic後變成曲線，尾端拉升

代表客戶距離第一次購買產品(time\_since\_first\_purchase)，時間隔得越久

客戶越容易流失 (has\_churned = 0~1 未流失~流失)





將預測模型轉為scatterplot

將模型改為另一種簡單預測(紅點)

Most Likely Outcome

(has\_churned, 即<0.5客戶不會流失，>0.5客戶更有可能會流失)

可得出距離首次購買時間間隔越久，流失機率越高



The Comparison of Odds Ration and the Time First Purchase

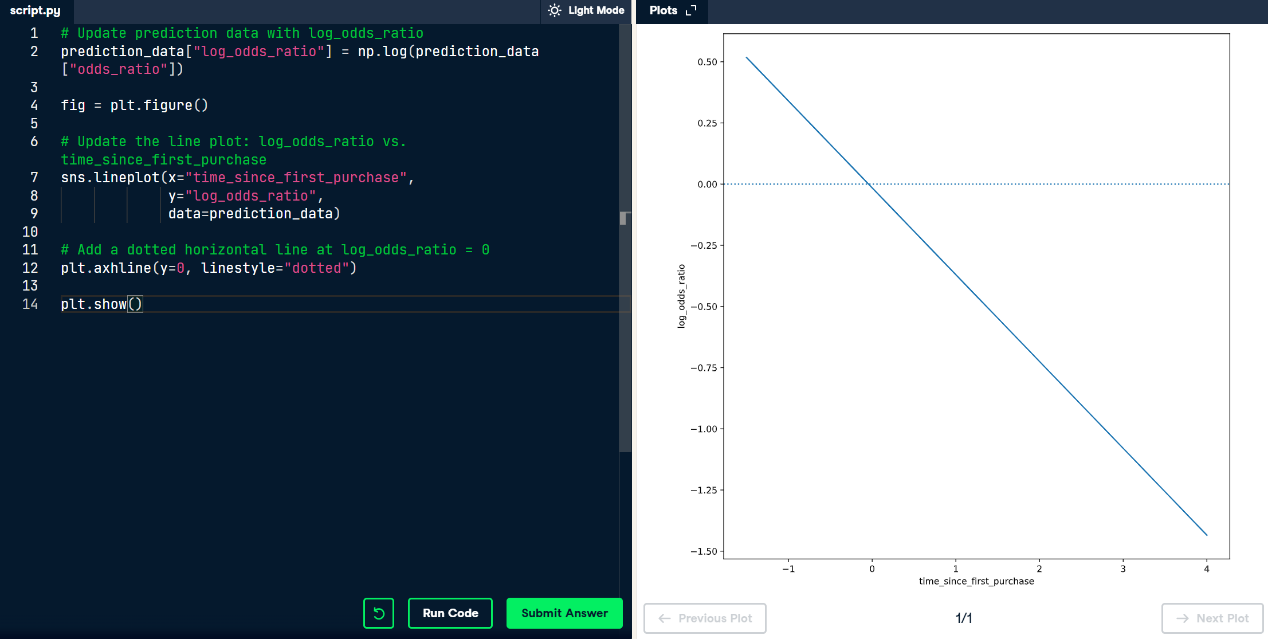
利用Odds Ratio賠率比與首次購買時間做比較

距離首購時間越久，賠率比越接近1，流失率越高

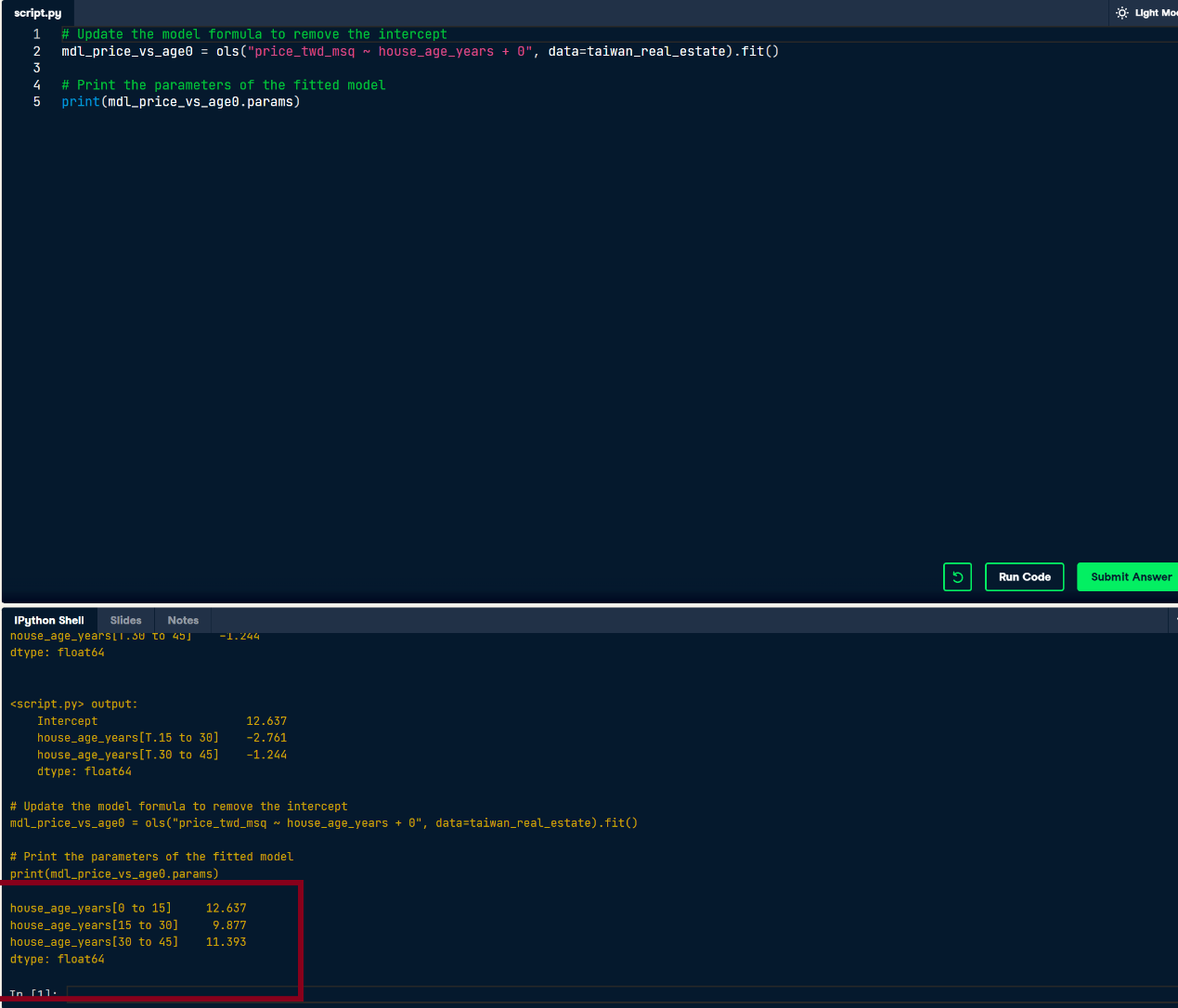


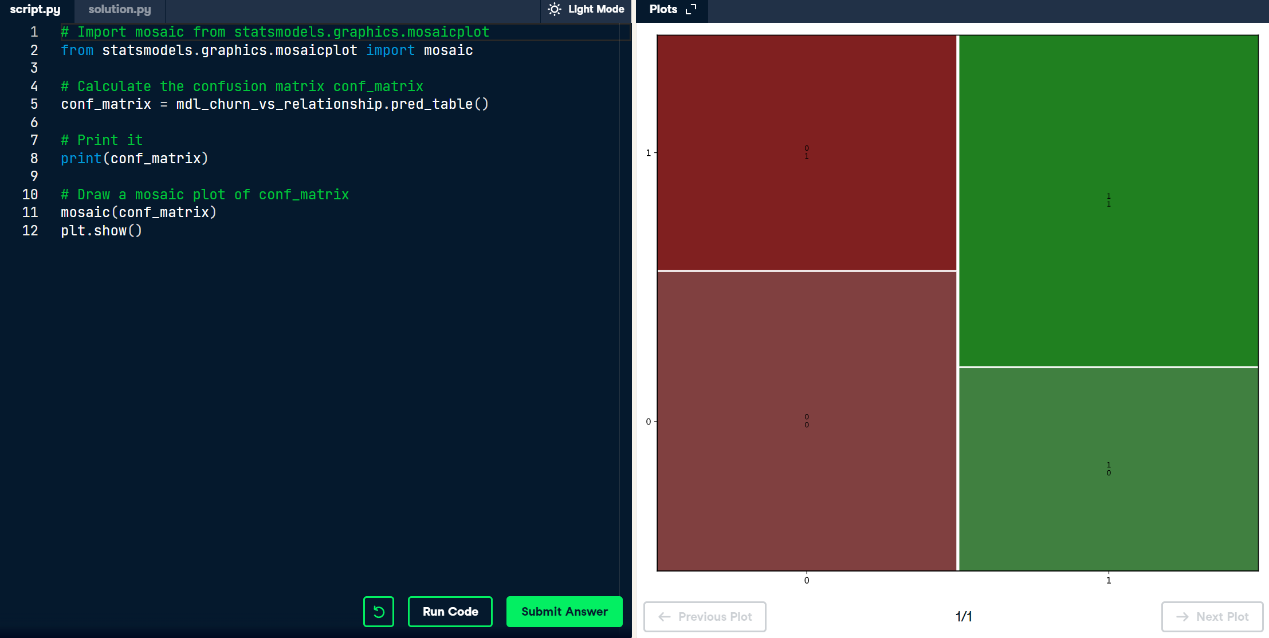
將賠率比log，使模型穩定

沒有log的賠率比(即曲線)，模型波動較大，不易穩定



Linear Regression 線性回歸分類比較



Confusion Matrix 模糊矩陣