國立空中大學

「Python 程式設計與實務應用」

期末小組專題

應用程式開發手冊

導師：賴素純

班級：ZZZ002

組別：11組

小組：8089

成員（tronclass平台組員順序）：

112222911詹秉蒼、112224868蔡惠婷、112224871蔡婷羽、113122776張心齡、113122877陳銘泓、113170973余誼姍

1. **專題主題**

房價預測器

1. **專題簡介**

* 本專案開發之軟體，透過爬取內政部實價登錄資料，用最真實的成交資料作為依據，以統計學方法科學地預測房價行情。
* 在現今臺灣房價飛漲，房價所得比不斷上升的年代，購屋已不易，又若購屋前沒有做足功課及進行相關統計分析，不熟悉欲購置區域的行情，就容易被當盤子，以高於行情的價格成交。
* 購屋前的資料蒐集及統計分析費時又費力，而透過本軟體，購屋者可藉由篩選特定條件的房屋交易資料，並輸入欲購置的時間及物件坪數等相關資料，就可以快速得到欲購買標的之預測房價。有了房屋預測價格，購屋時便不易遭屋主或房仲之話術欺騙，導致買貴。
* 本軟體可減輕購屋者購屋前資料收集分析的負擔，增加決策正確性及速度，並提供議價依據（議價籌碼）。

1. **小組分工**

提案發想：蔡婷羽（112224871）

需求規格化：蔡婷羽（112224871）

專案統籌規劃：詹秉蒼（112222911）

需求分析：全體組員

程式撰寫：

* GUI介面、數據驗證：陳銘泓（113122877）
* 資料表設計與優化：詹秉蒼（112222911）
* 交易資料下載：詹秉蒼（112222911）
* 交易資料檔彙整與分類：余誼姍（113170973）、張心齡（113122776）
* 資料庫查詢語句與條件產出：蔡惠婷（112224868）
* 預測模型與演算法：蔡婷羽（112224871）
* 整合程式碼：詹秉蒼（112222911）
* 開發手冊：詹秉蒼（112222911）
* 使用手冊：余誼姍（113170973）
* 執行檔編譯（Windows）：陳銘泓（113122877）
* 執行檔編譯（Mac OS）：代辦
* 整合測試：全體組員
* 專題報告：張心齡（113122776）
* 報告投影片：張心齡（113122776）
* 宣傳影片/海報製作：余誼姍（113170973）

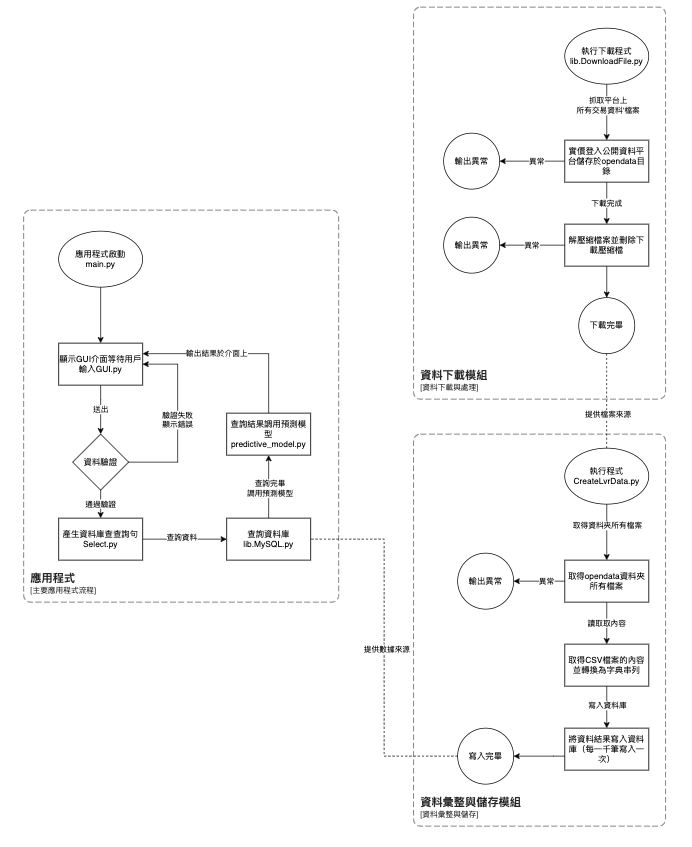
1. **開發環境**

* 語言：Python 3.12.7
* 依賴套件：
  + 檔案下載：requests
  + 檔案解壓縮：zipfile
  + CSV檔案處理：os、sys、csv、datetime、collections、json
  + 資料庫操作：mysql.connector
  + GUI介面與資料驗證：tkinter、customtkinter、functools. Partial
  + 預測模型：pandas、numpy
* 開發系統：Windows、MacOS
* 流程圖：drawio
* 資料庫：MySQL 8.1.30
* 版本控制：Git、GitHub
* IDE：Microsoft Visual Studio Code

1. **系統開發流程**

* 需求分析：

由全體組員進行討論，對於提案內容進行、流程圖繪製、功能分解、使用套件、分配開發人員、制定函數、類別接口與參數相關工作分派．

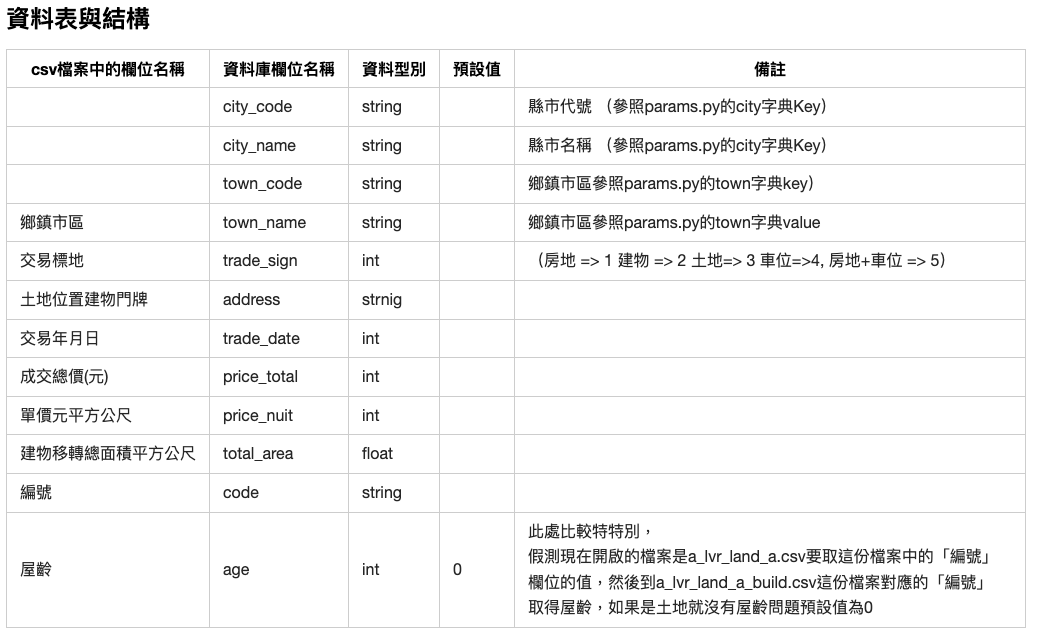


* 定義功能的輸出、輸入與邏輯規則

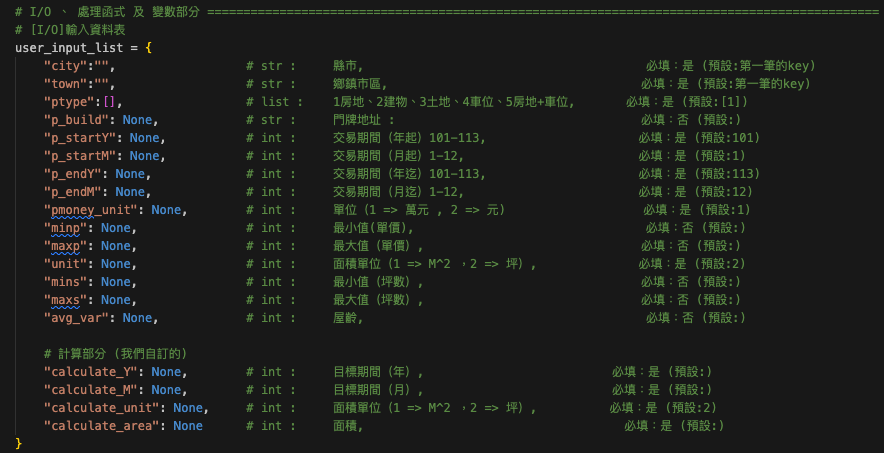
資料下載邏輯



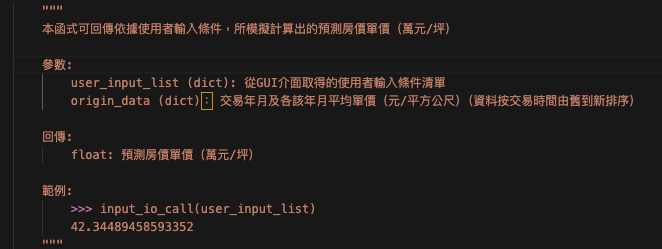
CSV資料轉換為資料庫格式與注意事項



GUI介面提供的輸出內容



預測模型的輸入與輸出



* 小組成員分派開發項目與制定完成日期



* 開發挑戰問題解決

問題：實價登入網站的版本更新，無法獲取資料來源

解決方式：使用爬蟲抓取「[不動產成交案件實際資訊資料供應系統](https://plvr.land.moi.gov.tw/Index" \t "_blank)」中的公開資料，來解決無法即時獲取交易資料問題，作為解決方案

問題：下載的數據量約有460多萬筆資料，使用資料庫儲存時，遭遇到效能問題

解決方式：透過資料庫的分析語句工具對相關的查詢語句，進行索引的優化與調教

* 整合測試

由各成員在本機上進行完整的測試

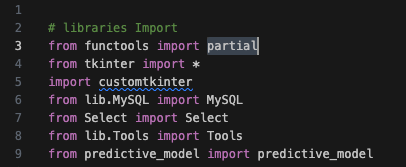
* 問題修正

測試階段遇到問題反映給原作者，由原作完成修正後再進行測試

1. **程式設計主要技巧**

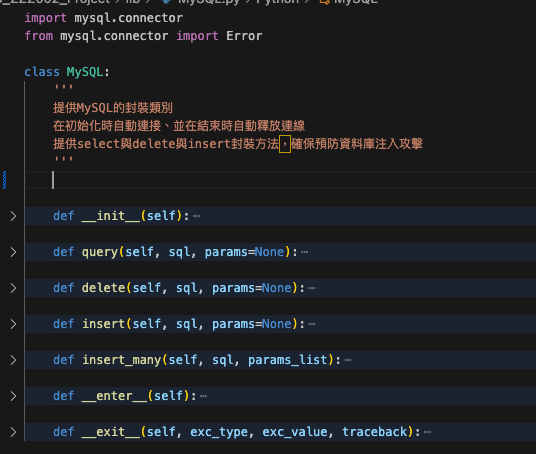
* 導入模組

導入Python原生、第三方模組加速建構，以及成員開發之各實作程式，使開發工作可以分工又不受干擾



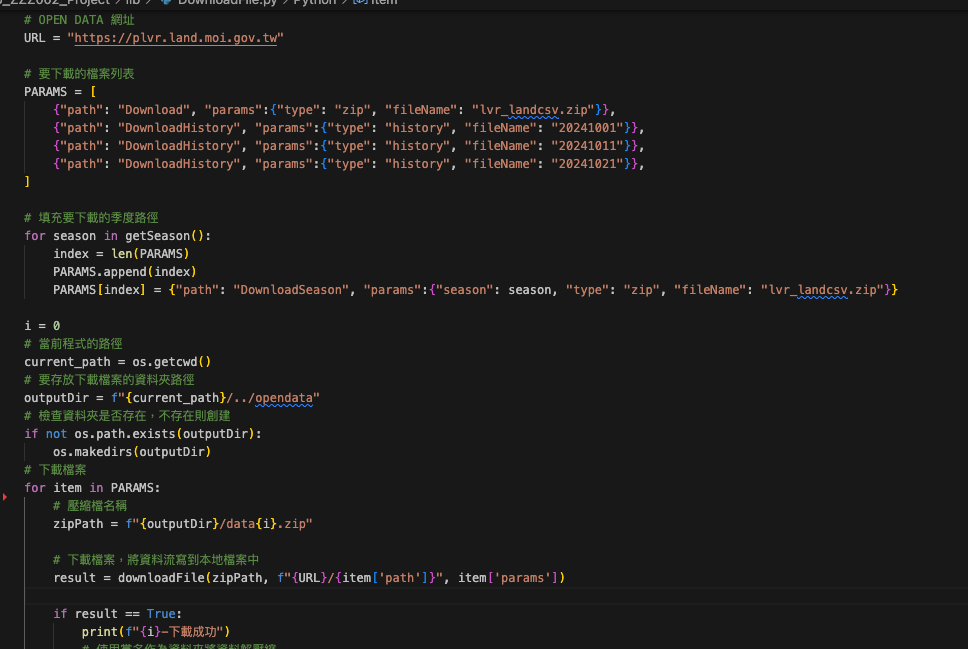
* 類別方法封裝

圖為資料庫的新增、查詢、刪除、更新個方法封裝為類別



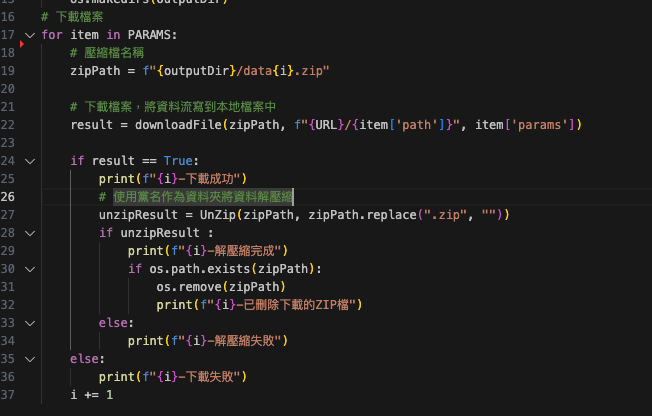
* HTTP訪問

使用requests套件，下載「[不動產成交案件實際資訊資料供應系統](https://plvr.land.moi.gov.tw/Index" \t "_blank)」中的交易資料，圖為先產生所有要下載的連結，透過http get方法將所有公開檔案下載到本地端

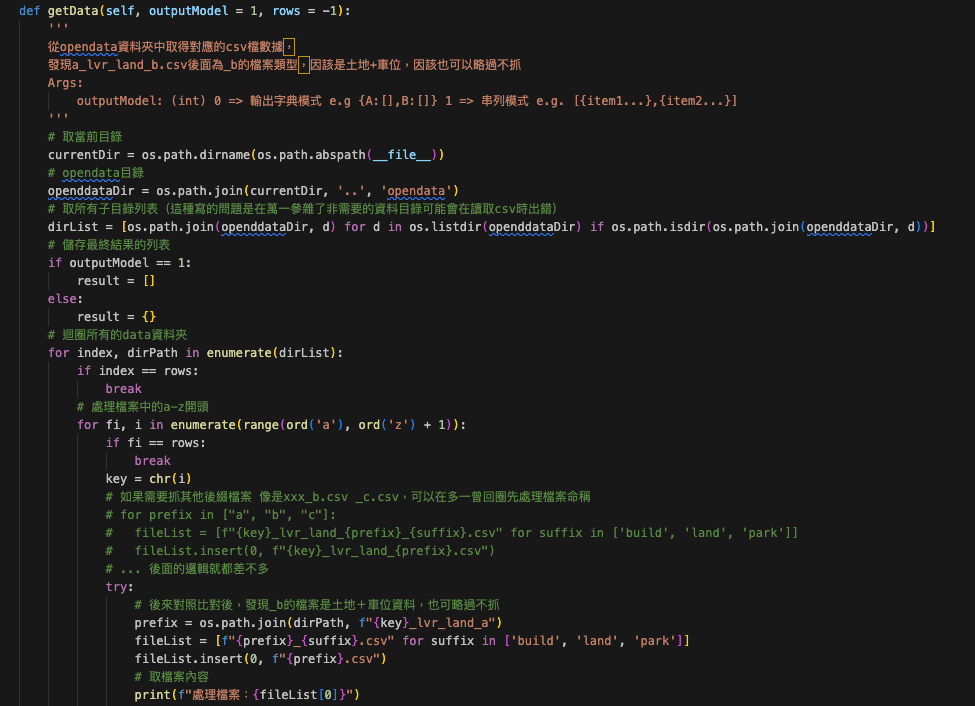


* 檔案解壓縮與解析

將下載的檔案(壓縮檔進行解壓縮)完成後刪除避免佔用過多空間



使用CSV套件來取得檔案內容，並彙整成資料庫所需格式



* 資料庫批次寫入

因搜集到的數據量超過百萬筆（約460多萬筆），若一次性寫入資料庫，會造成大量數據庫開銷，採用每一千筆批次寫入，讓資料庫的IO達到較良好的效能



* 圖形界面

透過圖形界面來搜集用戶需要過濾的數據，以及提供預測結果



* 數據驗證

提醒用戶必要輸入的數據，以及進行數據型態的控制與驗證



1. **原始碼重點解說**

* 資料下載與解壓縮（lib.DownloadFile.py）：

主要目的用於下載由101年開始到當下的實價登入資料，並將其解壓縮後儲存於本地端，主要邏輯片段

|  |
| --- |
| 函數主要產生101年到當年的年+季以[明國年] S [第幾季]為返回格式  透過迴圈產生出對應公開資料平台的下載點URL  以字典串列形式返回，字典中包含檔名、路徑、類型    將資料解壓縮後，在刪除壓縮檔節省儲存空間 |
|  |

* 資料庫連接與CRUD操作（lib.MySQL.py）

主要用於封裝mysql.connector套件中的資料操作方法，方便重複調用，以及在每次操作資料庫完畢後，自動釋放連線資源

|  |
| --- |
| 封裝CRUD Create 、Read、Update、Delete資料操做方法，以及多提供一個針對批次寫入的方法    透過\_\_enter\_\_、\_\_exit\_\_在結束時自動釋放資料庫連線 |
|  |

* 寫入資料庫（CreateLvrData.py）

主要用於提取所有下載的CSV檔案中的內容，並整理成可以寫入資料庫的數據型態，將資料儲存於MySQL資料庫中

|  |
| --- |
| CreateLvrData .getData()方法  取得lib.DownloadFile.py下載的CSV檔案並從中提取數據，進行數據的資料轉換    CreateLvrData．getCsv()方法  對CSV中的數據進行預處理與轉換    CreateLvrData．insertSQL ()方法  將取得的數據以每批一千筆的的速率寫入資料庫 |

* GUI介面（資料搜集）

|  |
| --- |
| 使用tkinter及customtkinter套件製作GUI介面，允許用戶通過GUI介面圖像化輸入資料，介面中物件程式碼範本如下:  使用元件包括: customtkinter的CTkScrollableFrame、CTkFrame、CTkButton、CTkLabel、CTkOptionMenu、CTkCheckBox、CTkRadioButton及CTkEntry。  使用者填寫資料時，會通過回調函數將資料寫入user\_input\_list: |

* 數據驗證

|  |
| --- |
| 當使用者按下[生成資料]按鈕時，會開始檢查輸入資料是否正確、完整:    檢查函數分別驗證每項資料是否合規，若不符合規則，則判斷是否有預設值。若有，則填入預設值；若無，則對使用者進行文字警告:  如果資料填寫完整且合規，則呼叫資料庫及預測模型: |

* 資料庫語句生成（Select.py）

主要用於接收GUI介面所搜集到的用戶輸入資料，並產生對應的資料庫查詢語句

|  |
| --- |
| Select.createQuery()方法  處理傳入的數據，產生對應的資料庫查詢語句，若數據沒有值或者為None時，將會不產生對應的欄位查詢語句 |

* 預測模型

|  |
| --- |
|  |

* 結果輸出

|  |
| --- |
| output\_show()函數取得回傳值後，判斷回傳值是計算結果(正數)或錯誤代碼(負數)，並以此輸出結果至GUI:    一個錯誤範例: |