

類神經網路簡介 HW3

Q36111281 張偉治

1. 原始資料處理

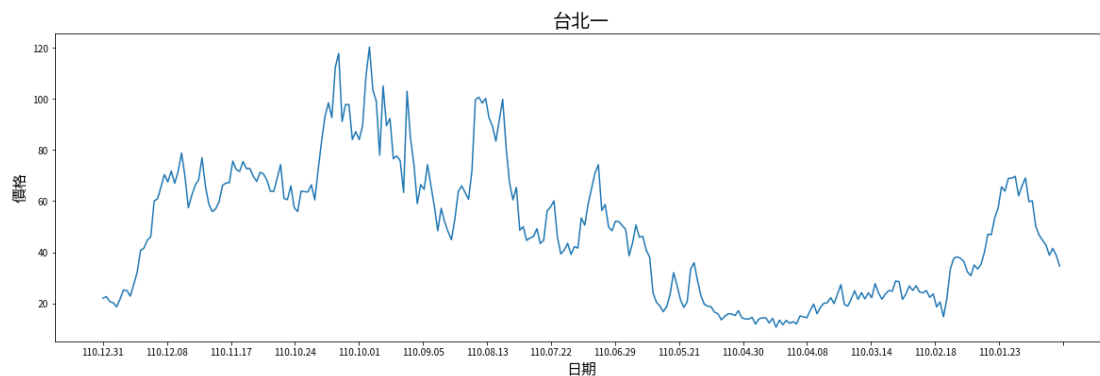
Pandas 在 class 的封裝上下了很多心思，封裝了許多功能，只需要短短幾行即可完成十分複雜的功能。同時相對的，使用上也受到很多的限制，因為相同的功能已經被封裝進 class 內，若用其他寫法去完成可能會使程式碼變得十分冗長且不易閱讀，有時更甚至是無法完成，再加上 pandas 的網站上並沒有對所有情況做說明，使得 pandas 在使用上十分仰賴對每種使用方法的了解。

其中，第三小題「呈現高雄市一年當中每個月的交易量」，若使用 pandas 預想外的方法，需要先取出「高雄市」的資料，再將取出的資料建立一個隨日期判斷的 flag，紀錄某一個月那些資料需要加總，這是因為日期的資料型態是字串，因此若直接對矩陣中的字串再做矩陣運算會出現不合法的指令。而 pandas 在 class 的設計上，若要對矩陣中的字串進行操作只需要在後面加上.str 即可進行傳統字串操作，使得整個程式在撰寫上變得簡潔並且易於維護。

```
p3_tomato[ p3_tomato['TransDate'].str[4:-3]==month ]['Trans_Quantity'].sum()
```

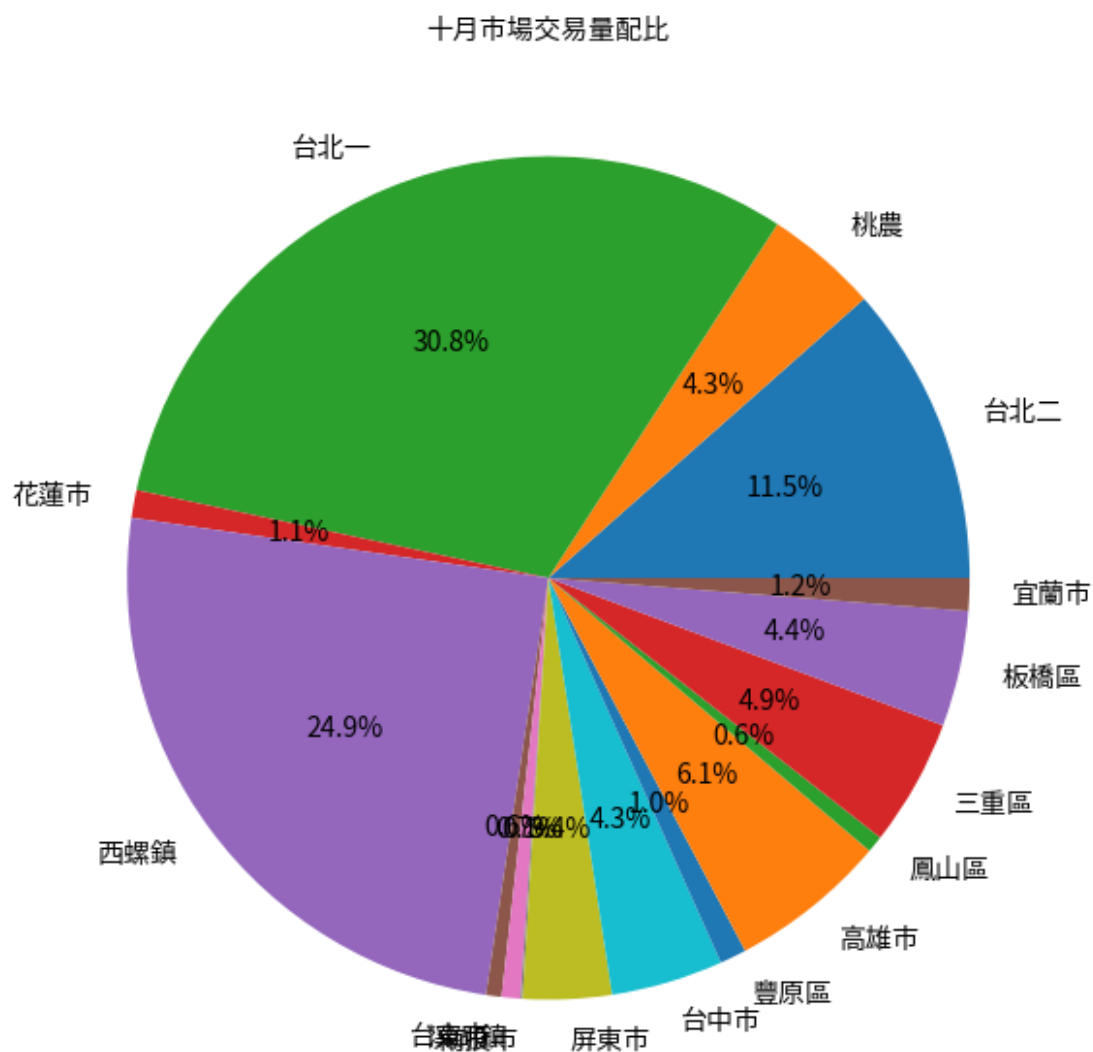
2. 資料觀察

2.1 台北一價格走勢



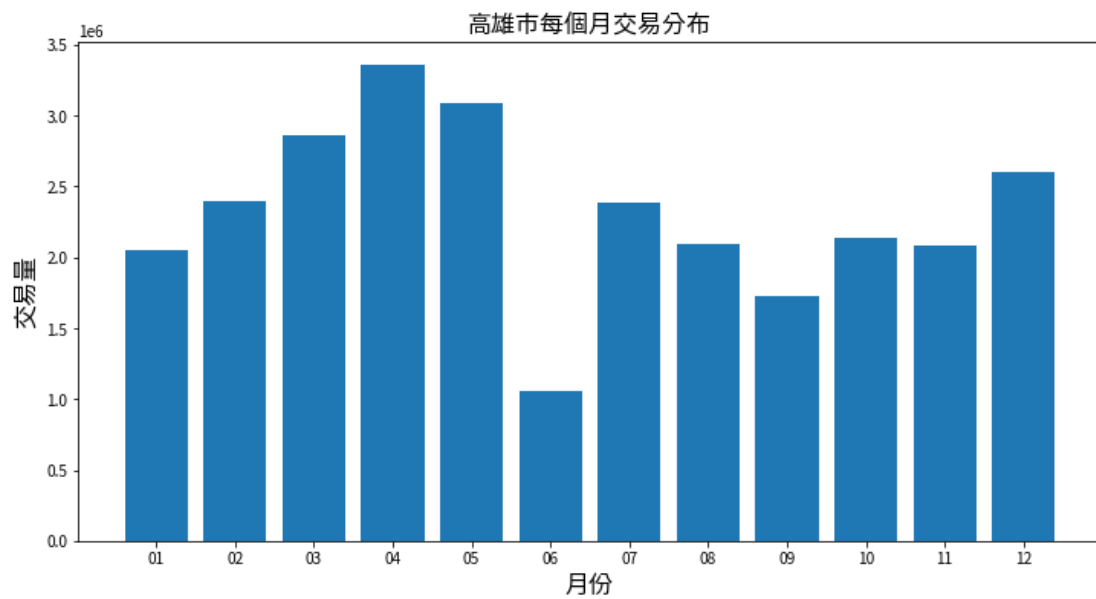
在台北一的牛番茄價格在四月份有明顯的低谷，可能是四月是牛番茄的盛產期因此供過於求，根據供需法則造成價格明顯下跌。

2.2 現各個市場在十月份交易量占比



根據十月份的交易量占比可以得知，牛番茄在台北與西螺交易量最大，再根據相關經驗知識，可以推論台北是牛番茄最大宗的交易地點，而西螺鎮則是牛番茄最大宗的產地，農會大量向農民收購牛番茄。

2.3 110 年高雄市每個月的交易量



四月和六月分別有最高的交易量和最少的交易量，可能是四月份是產季，高雄市農會大量收購牛番茄，六月則因為收購得差不多了，交易量則為整年最少。

3. pandas 使用心得

Pandas 在副程式上有相當豐富的設計，可以幫助在面對未知資料時，透過平均、偏度或是四分位數，對資料進行分析或是離群值的調整，使資料在使用上會比原始未處理的資料更為精確。