
時間曲線分析

運用 Pandas
Python 程式設計 王穎聰

Pandas

1. Pandas 之於 Python, 就好比 Excel 之於 Windows 一樣
2. 用於機器學習、大數據分析、時間序列分析
3. Python 資料科學必學入門

數據型態

1. Series

一條序列由 index 和數值組成

```
import pandas as pd
```

```
list1 = [2, 5, 4, 8, 10]
```

```
date = pd.date_range( '20181025', periods=5)
```

```
series1 = pd.Series( list1, index=date )
```

Index	Values
2018-10-25	2
2018-10-26	5
2018-10-27	4
2018-10-28	8
2018-10-29	10

Series 資料查找與計算

`series1.loc['20181028']`

`series1.loc['20181026' : '20181028']`

`series1.iloc[1]`

`series1.iloc[1:3]`

請注意查找資料的結果

`series1.sum()`

`series1.max()`

`series1.min()`

`series1.mean()`

`series1.std()`

`series1.cumsum()`

`series1.cumprod()`

Series 資料計算

`series1.rolling(2).sum()`

`series1.rolling(2).max()`

`series1.rolling(2).min()`

`series1.rolling(2).mean()`

`series1.rolling(2).std()`

`series1 + 1`

`series1 - 1`

`series1 * 2`

`series1 / 2`

`series1 > 3`

`series1 < 3`

Series 綜合練習

```
series1.loc[ series1 > 3 ]
```

```
series1.loc[ series1 > 3 ] = series1.loc[ series1 > 3 ] + 1
```

```
( series1.rolling(2).sum().cumsum() + 1 ).max( )
```

```
(( series1 + 1 ).cumprod( ) + 1 ).min( )
```

數據型態

2. DataFrame

```
d1 = [2, 5, 4, 8, 10]
```

```
d2 = [3, 6, 8, 6, 10]
```

```
date_time = pd.date_range('20181025', periods=5)
```

```
df1 = pd.DataFrame({'data1' : d1, 'data2' : d2 }, index=date_time)
```

```
r = [[2, 3], [5, 6], [4, 8], [8, 6], [10, 10]]
```

```
df2 = pd.DataFrame(r, columns=['data1', 'data2'], index=date_time)
```

index	data1	data2
2018-10-25	2	3
2018-11-26	5	6
2018-11-27	4	8
2018-11-28	8	6
2018-11-29	10	10

DataFrame 其他產生法

```
date_time = pd.date_range( '20181025', periods=6)
```

```
s1 = pd.Series( [1, 3, 7, 11, 9, 13], index=date_time)
```

```
s2 = pd.Series( [7, 12, 11, 5, 16, 6], index=date_time)
```

```
s3 = pd.Series( [13, 14, 5, 16, 7, 8], index=date_time)
```

```
df = pd.DataFrame( { 'data1': s1, 'data2': s2, 'data3':s3 } )
```

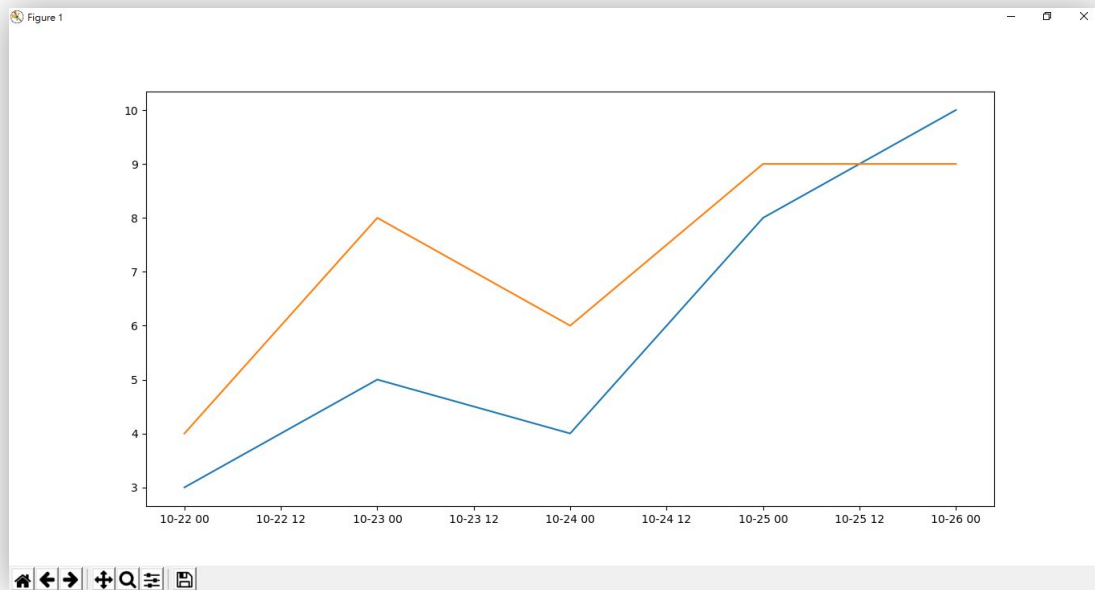

繪圖

使用 matplotlib 套件 (請安裝)

```
import matplotlib.pyplot as plt
```

```
plt.plot(df1)
```

```
plt.show( )
```



DataFrame 取值

df.loc['20181025']

df.iloc[1]

df.loc['20181026' : '20181028' , ['c1' , 'c2']]

df.iloc[1 : 4 , [0 , 1]]

df.cumsum()

df.cumprod()

df.rolling(2).mean()

df['data1']

df.cumsum(axis=1)

df.drop('data1' , axis=1)

Pandas 相關資料

官網: [Pandas Data Analysis Library](#)

十分鐘入門 Pandas: [10 Minutes to Pandas](#)

Panddas 官方文件: [網站版](#)、[PDF 檔](#)

用 Pandas 預測財務曲線 1

```
import pandas as pd
```

```
#每年淨額( 銀行帳戶變化 = 總收入 - 總支出 )
```

```
import matplotlib.pyplot as plt
```

```
每年淨額 = pd.Series( 0, index=預測時段 )
```

```
起始資金 = 30
```

```
每年淨額.iloc[ 0 ] = 起始資金
```

```
每月薪水 = 3
```

```
每年淨額.loc[ : 退休年齡 ] += 每月薪水 * 12
```

```
每月開銷 = 1
```

```
每年淨額 -= ( 每月開銷 + 每月房租 ) * 12
```

```
每月房租 = 1
```

```
plt.plot( 每年淨額 )
```

```
退休年齡 = 65
```

```
plt.show( )
```

```
預測時段 = range( 25, 90, 1 )
```

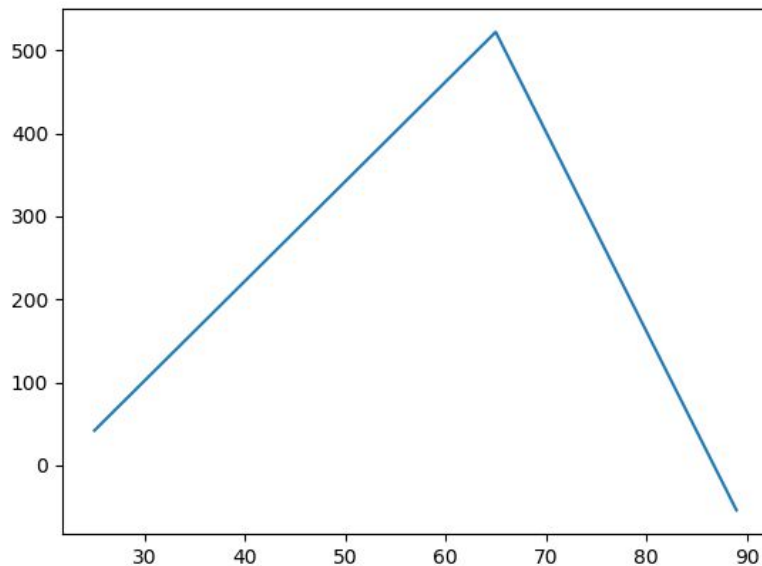
用 Pandas 預測財務曲線 2

#無投資總資產

無投資總資產 = 每年淨額.cumsum()

plt.plot(無投資總資產)

plt.show()



用 Pandas 預測財務曲線 3

#有投資總資產

投資部位 = 0.7

投資年利率 = 1.05

#投資金額 = 總帳戶金額 * 投資部位

#存在戶頭 = 總帳戶金額 * (1 - 投資部位)

#帳戶餘額 = 投資金額 * 投資年利率 + 存在戶頭

用 Pandas 預測財務曲線 4

```
def compound(arr, ratio, return_rate):  
    ret = [ arr.iloc[0] ]  
    for v in arr[1:]:  
        ret.append(ret[-1] * ratio * return_rate + ret[-1] * (1 - ratio) + v)  
    return pd.Series(ret, 預測時段)
```

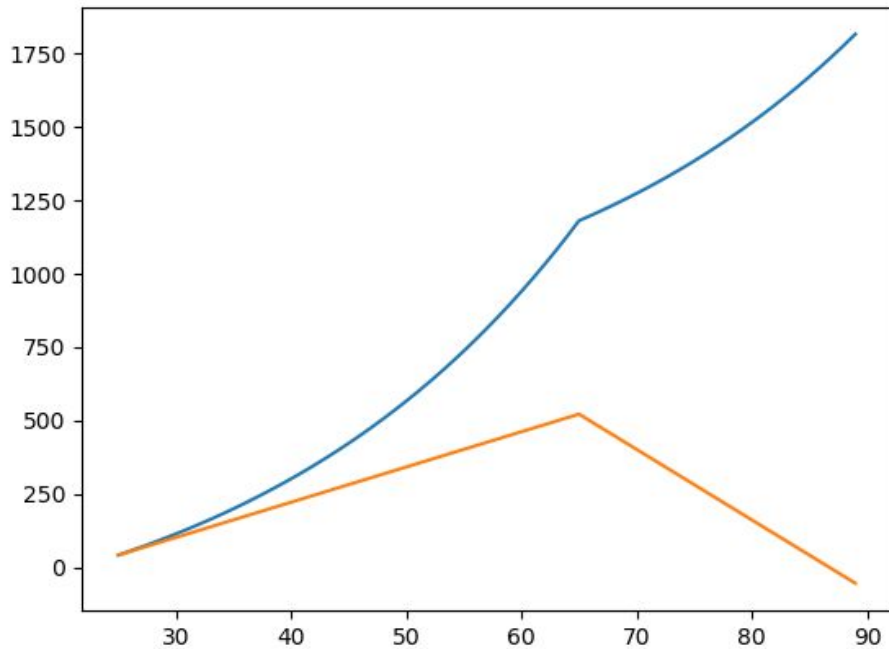
投資總資產 = compound(每年淨額, 投資部位, 投資年利率)

plt.plot(投資總資產)

plt.plot(無投資總資產)

plt.show()

用 Pandas 預測財務曲線 5



#考慮買房

買房價格 = 300

買房頭期款 = 100

買房年紀 = 35

房貸利率 = 3

貸款年數 = 20

用 Pandas 預測財務曲線 6

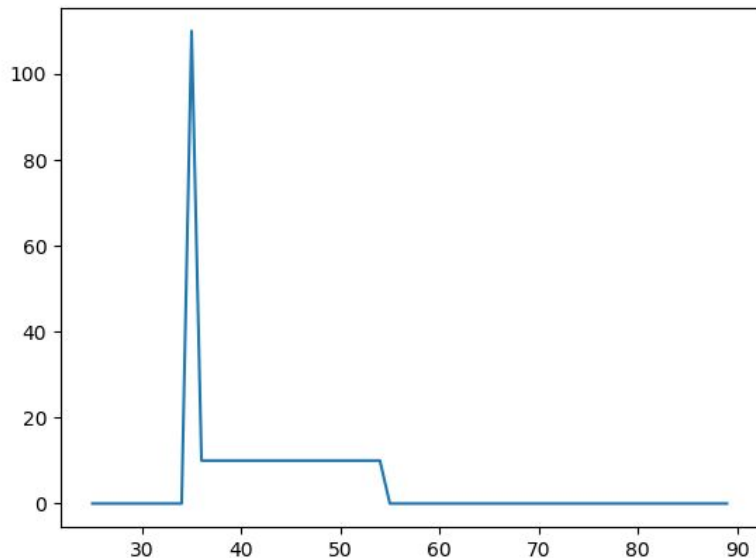
買房花費 = pd.Series(0, index=預測時段)

買房花費[買房年紀] = 買房頭期款

買房花費.loc[買房年紀:買房年紀+貸款年數-1] += (買房價格 - 買房頭期款) / 貸款年數

plt.plot(買房花費)

plt.show()



用 Pandas 預測財務曲線 7

#貸款利息

欠款 = pd.Series(0, index=預測時段)

欠款[買房年紀] = 買房價格

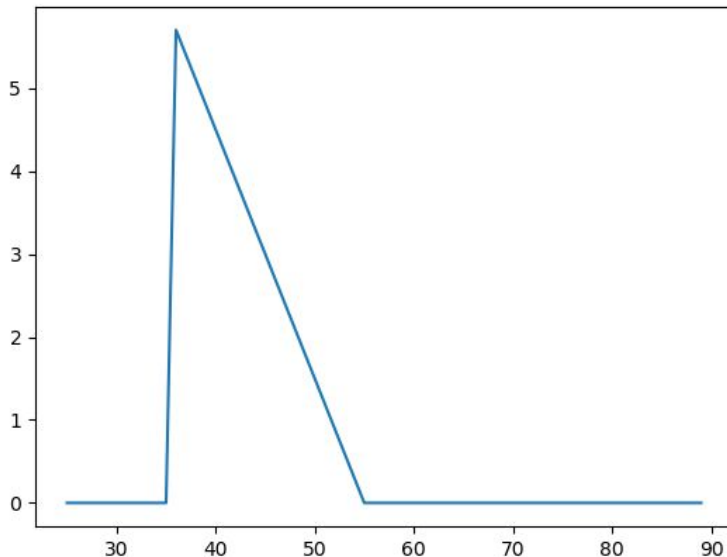
欠款 = 欠款.cumsum()

欠款 = 欠款 - 買房花費.cumsum()

利息 = 欠款.shift().fillna(0) * 房貸利率 / 100

plt.plot(利息)

plt.show()



用 Pandas 預測財務曲線 8

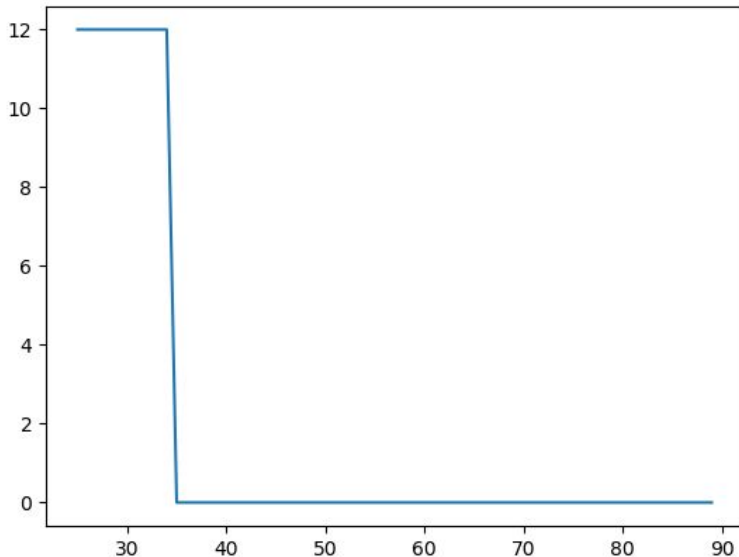
#計算繳房租

房租年繳 = pd.Series(每月房租*12, index=預測時段)

房租年繳.loc[買房年紀:] = 0

plt.plot(房租年繳)

plt.show()



用 Pandas 預測財務曲線 9

每年淨額_買房 = pd.Series(0, index=預測時段)

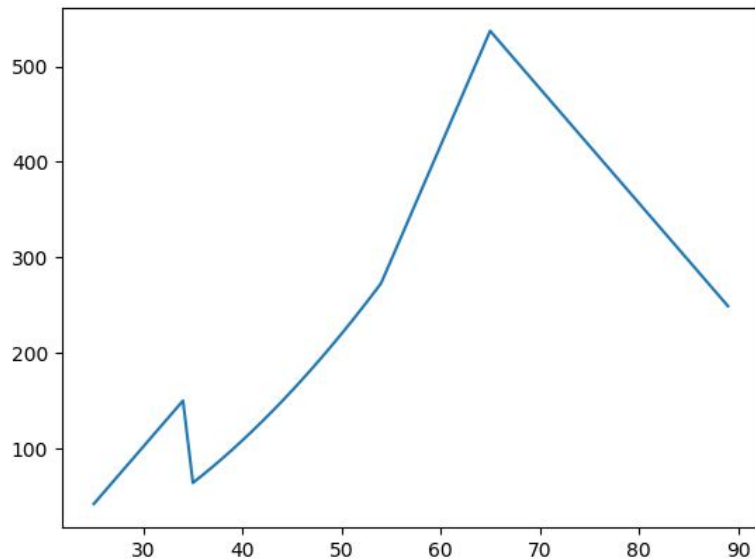
每年淨額_買房.iloc[0] = 起始資金

每年淨額_買房.loc[:退休年齡] += 每月薪水 * 12

每年淨額_買房 -= (每月開銷*12 + 房租年繳 + 利息 + 買房花費)

plt.plot(每年淨額_買房.cumsum())

plt.show()



用 Pandas 預測財務曲線 10

#投資且買房

投資_買房_總資產 = compound(每年淨額_買房, 投資部位, 投資年利率)

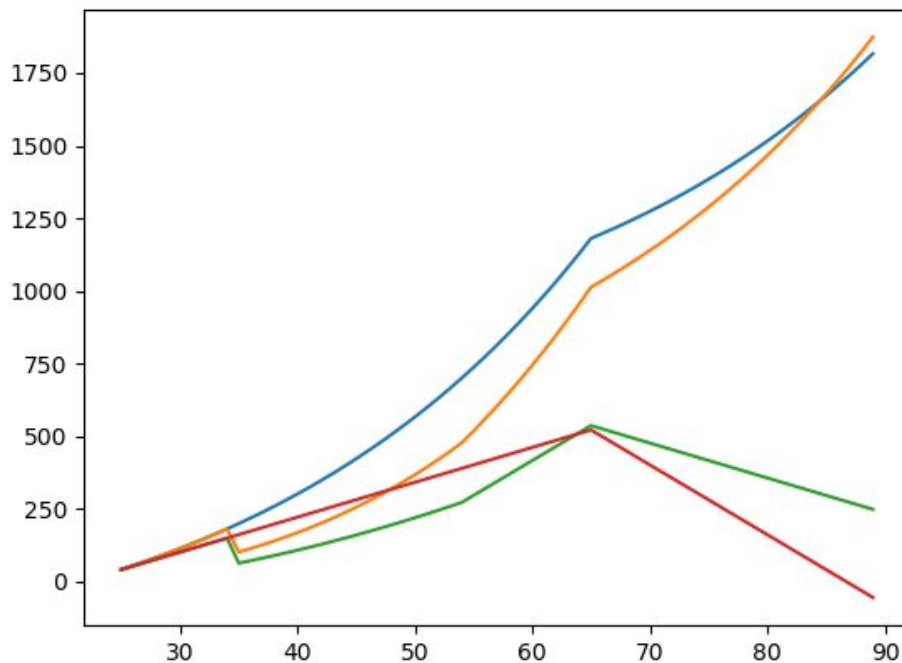
plt.plot(投資總資產)

plt.plot(投資_買房_總資產)

plt.plot(每年淨額_買房.cumsum())

plt.plot(無投資總資產)

plt.show()



matplotlib 顯示中文的方式

#以 Windows 的內建標楷體為例

```
from matplotlib.font_manager import FontProperties
```

```
myFont = FontProperties( fname=r"C:\Windows\Fonts\kaiu.ttf" )
```

```
plt.title("預測人生財務曲線", fontproperties=myFont)
```

```
plt.xlabel("年齡", fontproperties=myFont)
```

```
plt.ylabel("金額", fontproperties=myFont)
```