课程专业实践报告

学号					
实验题目	pandas 模块使用				

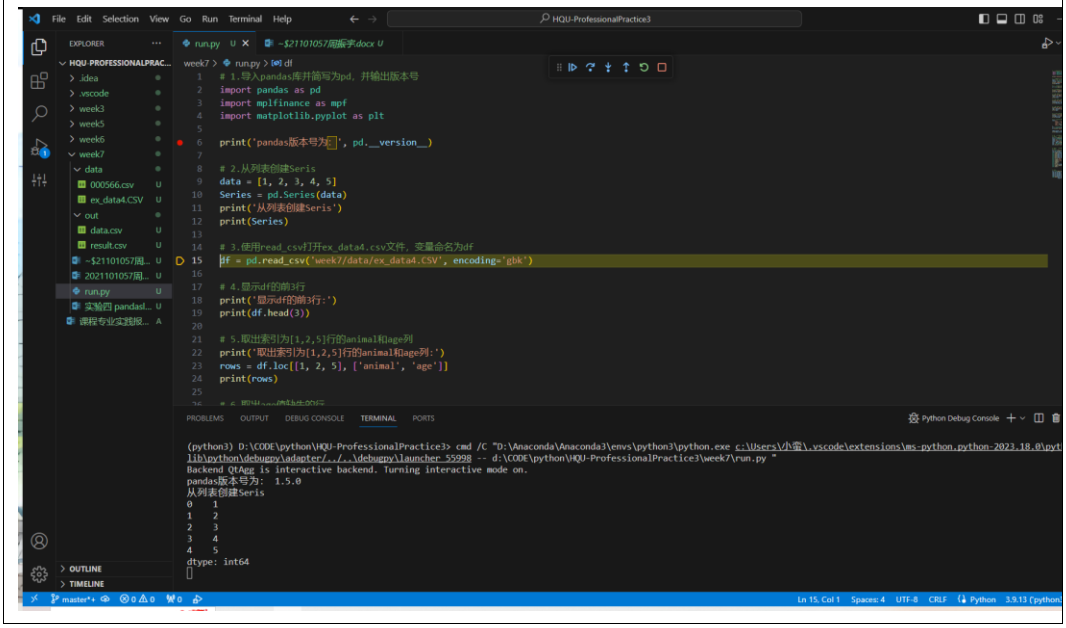
一．程序代码

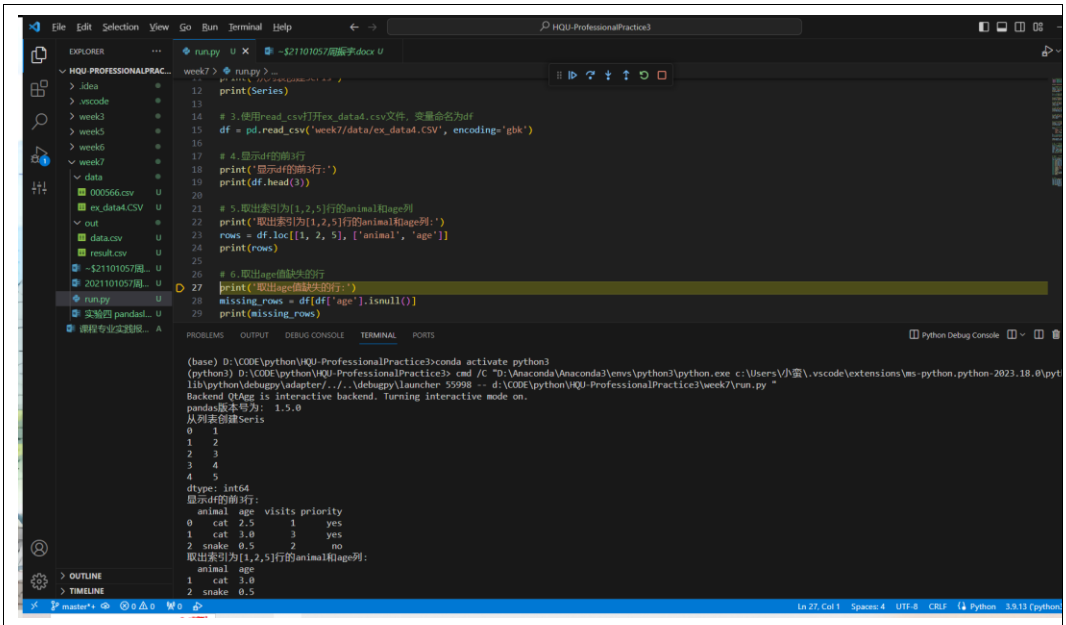
```
# 1.导入 pandas 库并简写为 pd，并输出版本号
import pandas as pd
import mplfinance as mpf
import matplotlib.pyplot as plt

print('pandas 版本号为：', pd.__version__)

# 2.从列表创建 Seris
data = [1, 2, 3, 4, 5]
Series = pd.Series(data)
print('从列表创建 Seris')
print(Series)

# 3.使用 read_csv 打开 ex_data4.csv 文件，变量命名为 df
df = pd.read_csv('week7/data/ex_data4.CSV', encoding='gbk')

# 4.显示 df 的前 3 行
print('显示 df 的前 3 行:')
print(df.head(3))

# 5.取出索引为[1,2,5]行的 animal 和 age 列
print('取出索引为[1,2,5]行的 animal 和 age 列:')
rows = df.loc[[1, 2, 5], ['animal', 'age']]
print(rows)

# 6.取出 age 值缺失的行
print('取出 age 值缺失的行:')
missing_rows = df[df['age'].isnull()]
print(missing_rows)


# 7.取出 age 值在 2，4 间的行
print('取出 age 值在 2，4 间的行:')
age = df[(df['age'] >= 2) & (df['age'] <= 4)]
print(age)


# 8.计算每个不同种类 animal 的 age 的平均数
print('计算每个不同种类 animal 的 age 的平均数:')
```

```
    average = df.groupby('animal')['age'].mean()
    print(average)

    # 9.使用字典创建如下信息所示的 dataframe,并保存至 data.csv，使用 groupby 分别计算 Male 和 Female 的分数 Score 的均值。
    data = {
        'Name': ['Alen', 'Bob', 'Cidy', 'Daniel', 'Ellen', 'Frankie', 'Gate', 'Hebe'],
        'Gender': ['Male', 'Male', 'Female', 'Male', 'Female', 'Male', 'Male', 'Female'],
        'Age': [18, 19, 18, 20, 17, 21, 20, 22],
        'Score': [80, 90, 93, 87, 96, 100, 88, 98],
    }
    df = pd.DataFrame(data)
    df.to_csv('week7/out/data.csv', index=False)

    mean_scores = df.groupby('Gender')['Score'].mean()
    print(mean_scores)


    # 10. 使用 read_csv 打开 000566.csv 文件，要求：
    # （1） 计算每月份平均换手率，并保存在文件 result.csv。
    df = pd.read_csv('week7/data/000566.csv', encoding='gbk')
    data = pd.DataFrame(df)
    data['日期'] = pd.to_datetime(data['日期']) # 将时间作为索引
    data = data.set_index('日期')
    df1 = data.resample('m', closed='left')['换手率'].mean()
    df1 = pd.DataFrame(df1)
    df1.to_csv('week7/out/result.csv')


    # （2） 绘制 5 月-8 月期间的周 k 线图。(上网查阅股票周 k 线图概念)
    # 方法一
    df = pd.read_csv('week7/data/000566.csv', encoding='gbk')
    may_to_aug_data = df[(df['日期'] >= '2021-05-01') & (df['日期'] <= '2021-08-31')]
    plt.figure(figsize=(12, 6))
    plt.plot(may_to_aug_data['日期'], may_to_aug_data['收盘价'], label='收盘价', marker='o')
    plt.plot(may_to_aug_data['日期'], may_to_aug_data['最高价'], label='最高价', linestyle='--')
    plt.plot(may_to_aug_data['日期'], may_to_aug_data['最低价'], label='最低价', linestyle='--')
    plt.plot(may_to_aug_data['日期'], may_to_aug_data['开盘价'], label='开盘价', linestyle='--')
    plt.title('周 K 线图 (5 月-8 月)')
    plt.xlabel('日期')
```

```
    plt.ylabel('价格')
    plt.legend()
    plt.grid()
    plt.xticks(rotation=45)
    plt.tight_layout()
    plt.show()



    # 方法二
    df = pd.read_csv("week7/data/000566.csv", encoding='gbk')
    may_to_aug_data = df[(df['日期'] >= '2021-05-01') & (df['日期'] <= '2021-08-31')]
    may_to_aug_data = may_to_aug_data.rename(columns={'开盘价': 'Open', '最高价': 'High', '最低价': 'Low', '收盘价': 'Close', '成交量': 'Volume'})
    may_to_aug_data['日期'] = pd.to_datetime(may_to_aug_data['日期'])
    may_to_aug_data.set_index('日期', inplace=True)
    mpf.plot(may_to_aug_data, type='candle', title='Weekly K-chart for May-August', ylabel='价格',
             style='yahoo', volume=True, figratio=(12, 6), tight_layout=True)
```

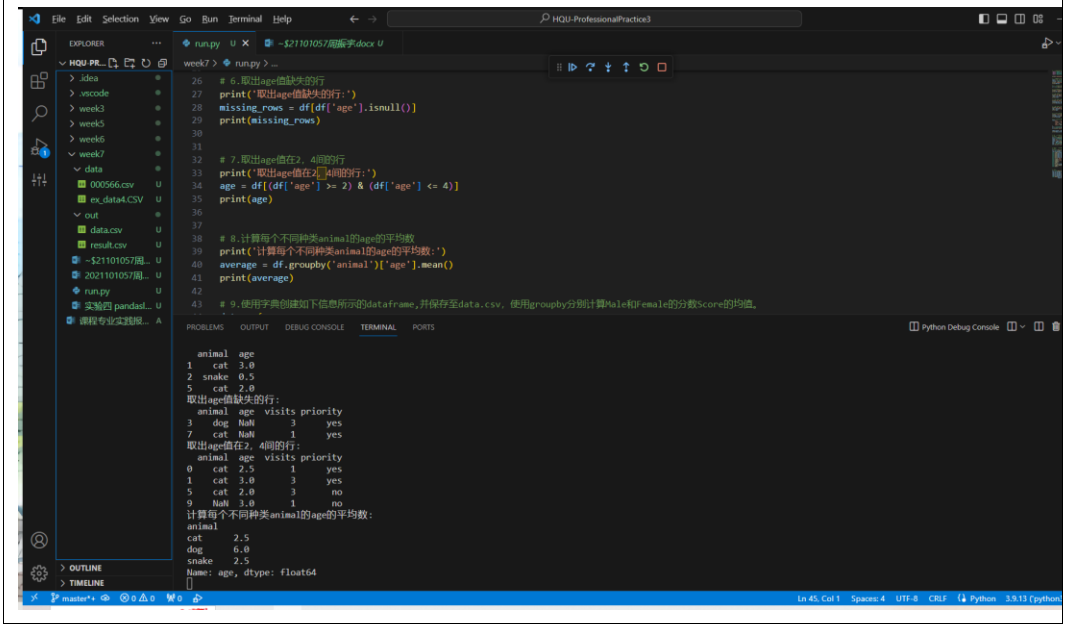
二．运行结果





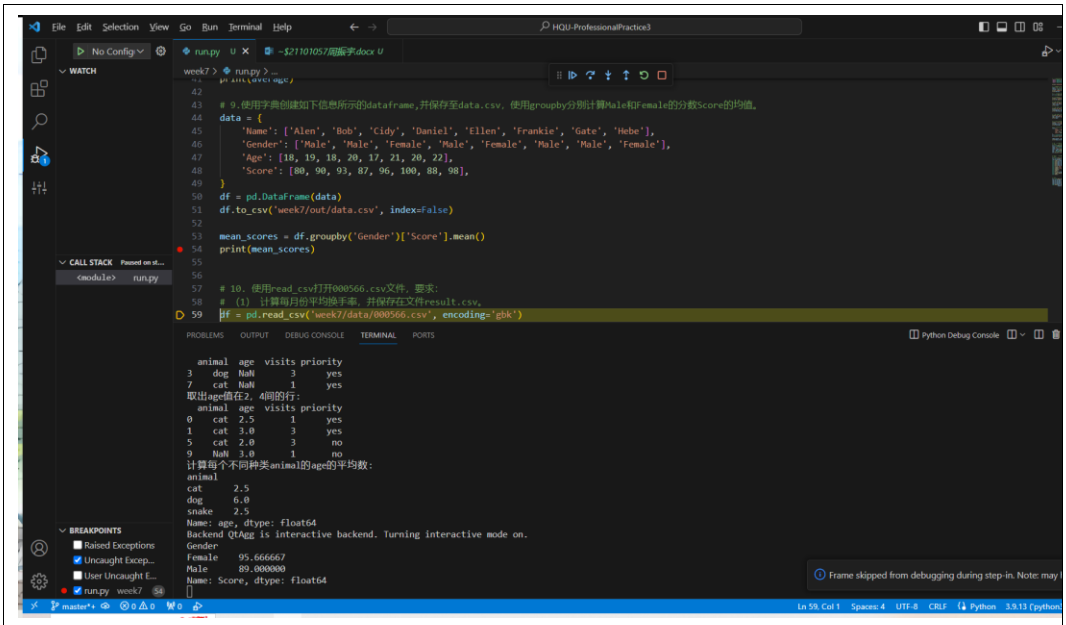
```
print(Series)

# 3.使用read_csv打开ex_data4.csv文件，变量命名为df
df = pd.read_csv('week7/data/ex_data4.CSV', encoding='gbk')

# 4.显示df的前3行
print('显示df的前3行:')
print(df.head(3))

# 5.取出索引为[1,2,5]行的animal和age列
print('取出索引为[1,2,5]行的animal和age列:')
rows = df.loc[[1, 2, 5], ['animal', 'age']]
print(rows)

# 6.取出age值缺失的行
print('取出age值缺失的行:')
missing_rows = df[df['age'].isnull()]
print(missing_rows)
```

```
(base) D:\CODE\python\HQU-ProfessionalPractice3>conda activate python3
(python3) D:\CODE\python\HQU-ProfessionalPractice3> cmd /C "D:\Anaconda\Anaconda3\envs\python3\python.exe c:\Users\小蜜\.vscode\extensions\ms-python.python-2023.18.0\pyt
lib\python\debugpy\adapter/../..\debugpy\launcher 55998 -- d:\CODE\python\HQU-ProfessionalPractice3\week7\run.py "
Backend QtAgg is interactive backend. Turning interactive mode on.
pandas版本号为:  1.5.0
从列表创建Seris
0    1
1    2
2    3
3    4
4    5
dtype: int64
显示df的前3行:
  animal  age  visits priority
0    cat  2.5       1      yes
1    cat  3.0       3      yes
2  snake  0.5       2       no
取出索引为[1,2,5]行的animal和age列:
  animal  age
1    cat  3.0
2  snake  0.5
```



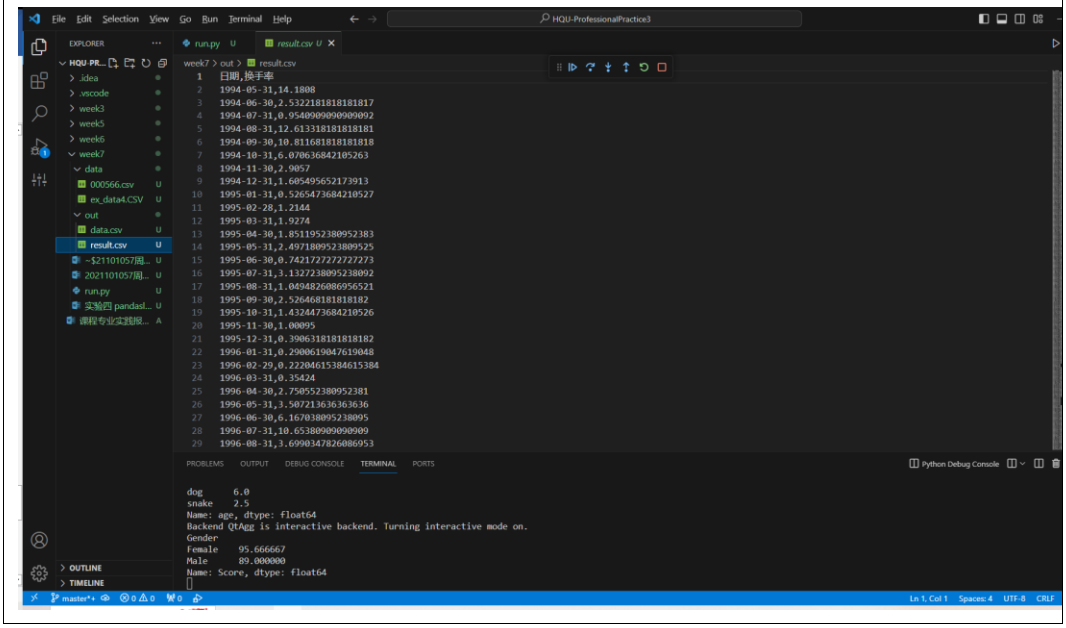
```
# 6.取出age值缺失的行
print('取出age值缺失的行:')
missing_rows = df[df['age'].isnull()]
print(missing_rows)


# 7.取出age值在2, 4间的行
print('取出age值在2, 4间的行:')
age = df[(df['age'] >= 2) & (df['age'] <= 4)]
print(age)


# 8.计算每个不同种类animal的age的平均数
print('计算每个不同种类animal的age的平均数:')
average = df.groupby('animal')['age'].mean()
print(average)

# 9.使用字典创建如下信息所示的dataframe,并保存至data.csv，使用groupby分别计算Male和Female的分数Score的均值。
```

```
  animal  age
1    cat  3.0
2  snake  0.5
5    cat  2.0
取出age值缺失的行:
  animal  age  visits priority
3    dog  NaN       3      yes
7    cat  NaN       1      yes
取出age值在2, 4间的行:
  animal  age  visits priority
0    cat  2.5       1      yes
1    cat  3.0       3      yes
5    cat  2.0       3       no
9    NaN  3.0       1       no
计算每个不同种类animal的age的平均数:
animal
cat      2.5
dog      6.0
snake    2.5
Name: age, dtype: float64
```



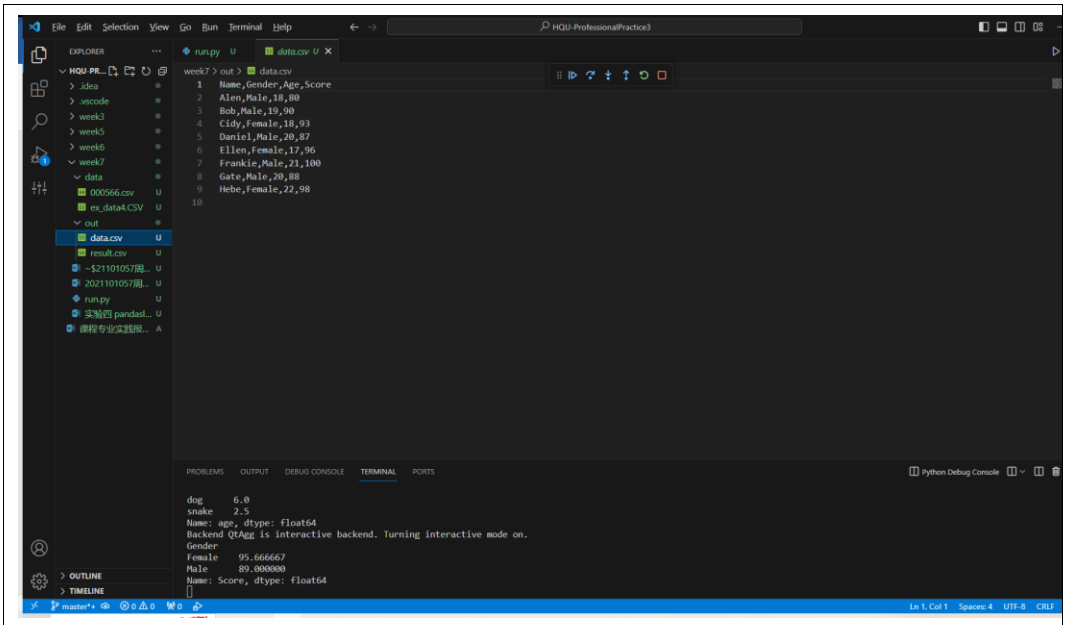
```
# 9.使用字典创建如下信息所示的dataframe,并保存至data.csv，使用groupby分别计算Male和Female的分数Score的均值。
data = {
    'Name': ['Alen', 'Bob', 'Cidy', 'Daniel', 'Ellen', 'Frankie', 'Gate', 'Hebe'],
    'Gender': ['Male', 'Male', 'Female', 'Male', 'Female', 'Male', 'Male', 'Female'],
    'Age': [18, 19, 18, 20, 17, 21, 20, 22],
    'Score': [80, 90, 93, 87, 96, 100, 88, 98],
}
df = pd.DataFrame(data)
df.to_csv('week7/out/data.csv', index=False)

mean_scores = df.groupby('Gender')['Score'].mean()
print(mean_scores)


# 10. 使用read_csv打开000566.csv文件，要求：
# (1) 计算每月份平均换手率，并保存在文件result.csv。
df = pd.read_csv('week7/data/000566.csv', encoding='gbk')
```

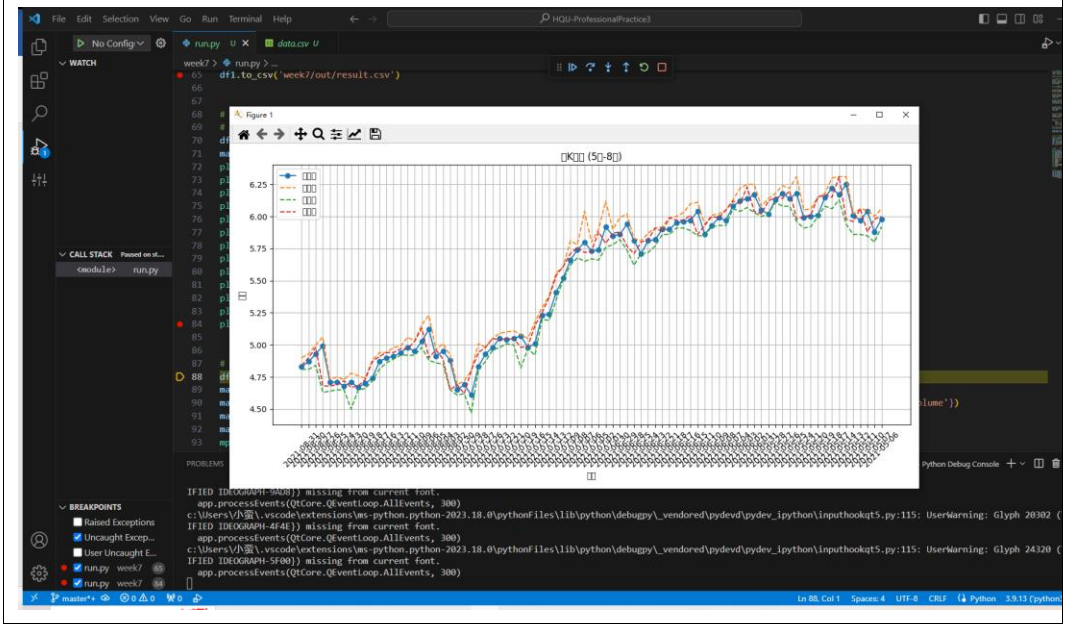
```
  animal  age  visits priority
3    dog  NaN       3      yes
7    cat  NaN       1      yes
取出age值在2, 4间的行：
  animal  age  visits priority
0    cat  2.5       1      yes
1    cat  3.0       3      yes
5    cat  2.0       3       no
9    NaN  3.0       1       no
计算每个不同种类animal的age的平均数：
animal
cat      2.5
dog      6.0
snake    2.5
Name: age, dtype: float64
Backend QtAgg is interactive backend. Turning interactive mode on.
Gender
Female    95.666667
Male      89.000000
Name: Score, dtype: float64
```



```
日期,换手率
1994-05-31,14.1808
1994-06-30,2.5322181818181817
1994-07-31,0.9540909090909092
1994-08-31,12.613318181818181
1994-09-30,10.811681818181818
1994-10-31,6.070636842105263
1994-11-30,2.9057
1994-12-31,1.605495652173913
1995-01-31,0.5265473684210527
1995-02-28,1.2144
1995-03-31,1.9274
1995-04-30,1.8511952380952383
1995-05-31,2.4971809523809525
1995-06-30,0.7421727272727273
1995-07-31,3.1327238095238092
1995-08-31,1.0494826086956521
1995-09-30,2.526468181818182
1995-10-31,1.4324473684210526
1995-11-30,1.00095
1995-12-31,0.3906318181818182
1996-01-31,0.2900619047619048
1996-02-29,0.22204615384615384
1996-03-31,0.35424
1996-04-30,2.75055238095238l
1996-05-31,3.507213636363636
1996-06-30,6.167038095238095
1996-07-31,10.65380909090909
1996-08-31,3.6990347826086953
```

```
dog      6.0
snake    2.5
Name: age, dtype: float64
Backend QtAgg is interactive backend. Turning interactive mode on.
Gender
female    95.666667
Male      89.000000
Name: Score, dtype: float64
```


File Edit Selection View Go Run Terminal Help
HQU-ProfessionalPractice3
EXPLORER
HQU-PR...
.idea
.vscode
week3
week5
week6
week7
data
000566.csv
ex_data4.CSV
out
data.csv
result.csv
~$21101057周...
2021101057周...
run.py
实验四 pandasl...
课程专业实践报...
run.py
data.csv
week7 > out > data.csv
1 Name,Gender,Age,Score
2 Alen,Male,18,80
3 Bob,Male,19,90
4 Cidy,Female,18,93
5 Daniel,Male,20,87
6 Ellen,Female,17,96
7 Frankie,Male,21,100
8 Gate,Male,20,88
9 Hebe,Female,22,98
10
PROBLEMS
OUTPUT
DEBUG CONSOLE
TERMINAL
PORTS
Python Debug Console
dog 6.0
snake 2.5
Name: age, dtype: float64
Backend QtAgg is interactive backend. Turning interactive mode on.
Gender
Female 95.666667
Male 89.000000
Name: Score, dtype: float64
OUTLINE
TIMELINE
master*+
Ln 1, Col 1
Spaces: 4
UTF-8
CRLF


File Edit Selection View Go Run Terminal Help
HQU-ProfessionalPractice3
No Config
WATCH
CALL STACK
<module> run.py
BREAKPOINTS
Raised Exceptions
Uncaught Excep...
User Uncaught E...
run.py week7
run.py week7
run.py
data.csv
week7 > run.py > ...
65 df1.to_csv('week7/out/result.csv')
Figure 1
□K□□ (5□-8□)
□□
□□
□□
□□
□□
6.25
6.00
5.75
5.50
5.25
5.00
4.75
4.50
PROBLEMS
Python Debug Console
IFIED IDEOGRAPH-9AD8}) missing from current font.
app.processEvents(QtCore.QEventLoop.AllEvents, 300)
c:\Users\小蜜\.vscode\extensions\ms-python.python-2023.18.0\pythonFiles\lib\python\debugpy_vendored\pydevd\pydev_ipython\inputhookqt5.py:115: UserWarning: Glyph 20302 (
IFIED IDEOGRAPH-4F4E}) missing from current font.
app.processEvents(QtCore.QEventLoop.AllEvents, 300)
c:\Users\小蜜\.vscode\extensions\ms-python.python-2023.18.0\pythonFiles\lib\python\debugpy_vendored\pydevd\pydev_ipython\inputhookqt5.py:115: UserWarning: Glyph 24320 (
IFIED IDEOGRAPH-5F00}) missing from current font.
app.processEvents(QtCore.QEventLoop.AllEvents, 300)
master*+
Ln 88, Col 1
Spaces: 4
UTF-8
CRLF
Python

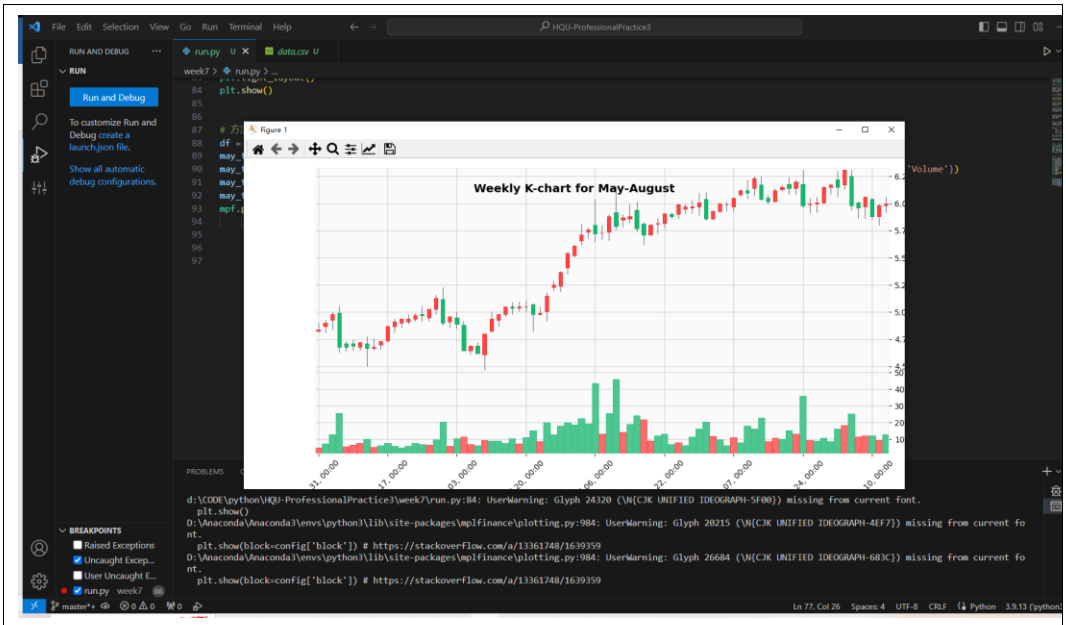


Figure 1
Weekly K-chart for May-August
d:\CODE\python\HQU-ProfessionalPractice3\week7\run.py:84: UserWarning: Glyph 24320 (\N{CJK UNIFIED IDEOGRAPH-5F00}) missing from current font.
plt.show()
D:\Anaconda\Anaconda3\envs\python3\lib\site-packages\mplfinance\plotting.py:984: UserWarning: Glyph 20215 (\N{CJK UNIFIED IDEOGRAPH-4EF7}) missing from current font.
plt.show(block=config['block']) # https://stackoverflow.com/a/13361748/1639359
D:\Anaconda\Anaconda3\envs\python3\lib\site-packages\mplfinance\plotting.py:984: UserWarning: Glyph 26684 (\N{CJK UNIFIED IDEOGRAPH-683C}) missing from current font.
plt.show(block=config['block']) # https://stackoverflow.com/a/13361748/1639359