

课时10 pandas模块的两种数据结构





P238 答案

1	2	3	4
A	C	D	D

一、Series数据结构

1. 有如下Python程序段：

```
import pandas as pd
s1=pd.Series([162, 175, 183], index=["s01", "s02", "s03"])
for i in s1:
    print(i)
```

执行该程序段后，输出的结果是 (A)

✓ A.162
175
183

B.s01 162
s02 175
s03 183

C.s01
s02
s03

D.162 175 183

该程序段中for循环的功能为查看s1对象中的values属性值，故输出的结果是：

162
175
183

二、DataFrame数据结构

2. 有如下Python程序段：

```
import pandas as pd
data=[[120, 140, 135], [118, 145, 103], [110, 132, 98]]
xm, km=["小明", "小军", "小豪"], ["语文", "数学", "英语"]
df1=pd.DataFrame(data, index=xm, columns=km)
```

下列说法正确的是 (C)

- A. `print(df1. 英语)`显示的第一行数据是：0 135
- B. `print(df1["数学"])`显示的第一行数据是：小豪 132
- ✓ C. `print(df1[df1. 语文<120])`显示的第一行数据是：小军 118 145 103
- D. `print(df1.sort_values("语文"))`显示的第一行数据是：小明 120 140 135

	语文	数学	英语
小明	120	140	135
小军	118	145	103
小豪	110	132	98

`print(df1. 英语)`显示的第一行数据是：小明 135。`print(df1["数学"])`显示的第一行数据是：小明 140。`print(df1.sort_values("语文"))`显示的第一行数据是：小豪 110 132 98。

3. 有如下Python程序段:

```
import pandas as pd
fruit=pd.Series(["watermelon", "peach", "lemon", "apple"], index=["s1", "s2",
"s3", "s4"])
a=fruit["s1"]
b=fruit[2]
res=a+b
print(len(res))
```

执行该程序段后, 输出的结果是 (**D**)

A. 10

B. 11

C. 12

D. 15

根据生成的Series对象fruit可知, a的值是"watermelon", b的值是"lemon", a+b表示将2个字符串拼接, 所以res的值是"watermelonlemon", 故输出的结果是15。

4. 有如下Python程序段:

```
import pandas as pd
data={"姓名":["小明","小夏","小静","小华"],"语文":[90,66,80,78],"英语":
:[98,86,92,75]}
df1=pd.DataFrame(data,columns=["姓名","语文","英语"])
df2=df1[df1.语文>=80]
ans=df2.英语.mean()
print(ans)
```

执行该程序段后,输出的结果是 (**D**)

A. 83

B. 85

C. 87.5

D. 95.0

该程序段的功能为求出DataFrame对象df1中“语文”列的值大于等于80的数据行赋值给df2,最后求出对象df2中“英语”列的平均值,故输出的结果是95.0。

课时11 pandas、matplotlib模块 的基本操作



一、pandas模块

1. 某电商平台2024年7~12月商品销售数据保存在“spxs.xlsx”文件中，部分数据如图所示。编写Python程序，统计该平台12月销售额最高的3件商品。

```
import pandas as pd
df=pd.read_excel("spxs.xlsx")
```

```
df_t=df2.head(3)
```

```
print(df_t)
```

方框中有如下可选代码：

- ①df1=df[df["月份"]==12]
- ②df1=df["月份"]==12]
- ③df2=df1.sort_values("销售额",ascending=False)
- ④df2=df1.sort_values("销售额",ascending=True)

则方框中应填入的正确代码为 (**B**)

- A. ①④
- B. ①③**
- C. ②④
- D. ③②

	A	B	C
1	商品名称	销售额	月份
2	商品A	10000	12
3	商品B	15000	12
4	商品C	8000	12
5	商品D	20000	12
36	商品F	9170	8
37	商品F	2170	7

要求统计12月销售额最高的3件商品，首先筛选12月的销售数据，语句②语法错误。根据代码“df_t=df2.head(3)”可知，选前3行，故数据降序排序，设置ascending=False。

2. [2025绍兴联考] 小张收集了某市高中各班学生7选3选科数据, 数据已经保存在“data.xlsx”文件中, 部分数据如图a所示, 其中学科列的“1”表示相对应行学生选了该学科, “0”表示未选。编写Python程序, 实现输出选考技术人数最多的3个班级的功能, 程序运行界面如图b所示。

```
import pandas as pd
df=pd.read_excel("data.xlsx")
df2=_____①_____
df_sort=_____②_____
print("选考技术最多的3个班级:")
print(df_sort.head(3)["班级"])
```

	A	B	C	D	E	F	G	H	I	J
1	学号	班级	姓名	物理	化学	生物	政治	历史	地理	技术
2	1001	高二(01)班	陈*爽	1	1	0	0	0	0	1
3	1002	高二(03)班	李*军	1	1	0	0	0	0	1
4	1003	高二(09)班	赵*兴	0	0	0	1	1	1	0
5	1004	高二(08)班	严*双	1	1	1	0	0	0	0
6	1005	高二(01)班	孔*达	1	1	0	0	0	0	1
7	1006	高二(01)班	赵*福	0	0	1	1	1	0	0
8	1007	高二(05)班	白*玉	0	0	0	1	1	0	1
9	1008	高二(04)班	王*军	0	0	0	1	1	1	0
10	1009	高二(01)班	郑*芳	0	0	0	1	1	0	1
11	1010	高二(01)班	曹*峰	0	0	0	0	0	0	0
12	1011	高二(01)班	严*旺	1	1	0	0	0	0	1

选考技术最多的3个班级:
0 高二(01)班
1 高二(03)班
2 高二(04)班

划线处应填入的正确代码为 (**B**)

- A. ①df.groupby("班级", as_index=False).sum() ②df.sort_values("技术", ascending=False)
- B. ①df.groupby("班级", as_index=False).sum() ②df2.sort_values("技术", ascending=False)
- C. ①df.sort_values("技术", ascending=True) ②df.groupby("班级", as_index=False).max()
- D. ①df.groupby("技术", as_index=False).max() ②df2.sort_values("班级", ascending=True)

①要实现统计每个班级选技术的人数, 应该按照班级分组, 然后对“技术”列进行求和, 因为每个学生的选科是1或0, 所以求和后的结果就是该班级选技术的人数, 故填入代码为df.groupby("班级", as_index=False).sum()。②根据输出语句可知, 输出前3个班级, 故要找到选考技术人数最多的3个班级, 需要对df2按照“技术”列进行降序排序, 即df2.sort_values("技术", ascending=False)。

二、matplotlib模块

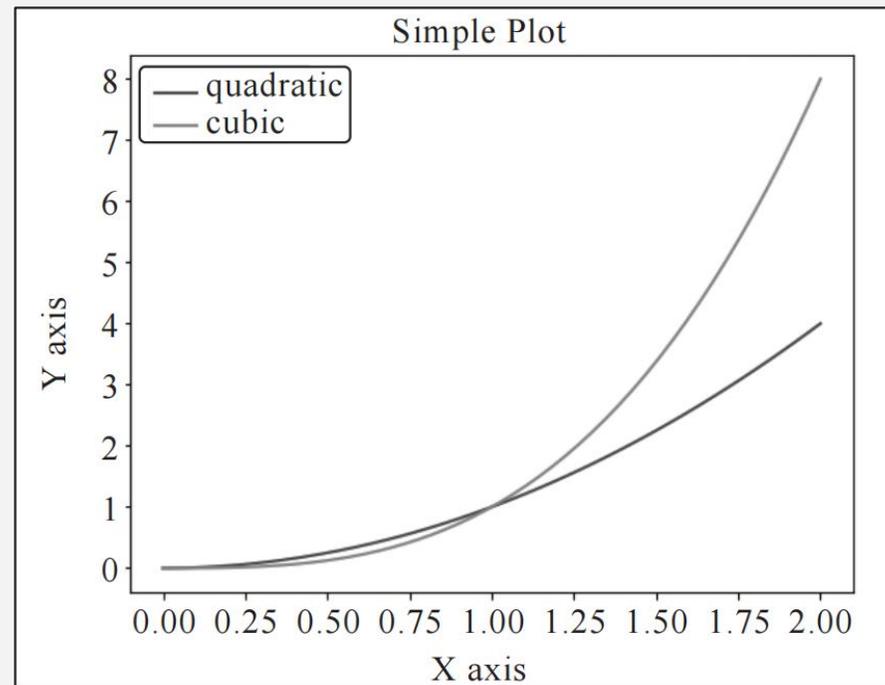
3. 使用matplotlib模块绘制如图所示的图像，代码如下。

```
import matplotlib.pyplot as plt
import numpy as np
x=np.linspace(0, 2, 50)
plt.plot(x, x**2, label="quadratic")
plt.plot(x, x**3, label="cubic")
plt.xlabel("X axis")
```

```
plt.title("Simple Plot")
plt.legend()
plt.show()
```

则划线处应填入的正确代码为 (B)

- A. plt.ylabel()
- B. plt.ylabel("Y axis")
- C. plt.ylim("Y axis")
- D. plt.ylim()



分析图表元素与程序代码，发现划线处缺少设置Y轴标签的代码。

4. 某班级学生的成绩数据存储在“cj.xlsx”文件中，部分数据如图 a 所示。

编写 Python 程序，在删除个别列数据后，根据“性别”进行分组求和，程序运行界面如图 b 所示。则划线处应填入的正确代码为 (A)

```
import pandas as pd
df=pd.read_excel("cj.xlsx")
df=df.drop(["编号","平时成绩","期中成绩","期末成绩"],axis=1)
df1=df. ① .sum()
df2= ②
print(df2)
```

- A. ①groupby("性别",as_index=False)
②df1.sort_values("总评成绩",ascending=True)
- B. ①groupby("性别",as_index=False)
②df1.sort_values("总评成绩",ascending=False)
- C. ①groupby("性别",as_index=True)
②df1.sort_values("总评成绩")
- D. ①groupby("性别",as_index=True)
②df.sort_values("性别",ascending=True)

	A	B	C	D	E	F	G
1	编号	姓名	性别	平时成绩	期中成绩	期末成绩	总评成绩
2	01	李*	女	47	56	45	49
3	02	张*斌	男	87	89	67	81
4	03	王*	男	98	56	84	79
5	04	何*群	女	53	78	69	67
6	05	尹*涛	男	88	89	87	88
26	25	黄*珊	女	80	90	92	87
27	26	郑*辉	男	95	76	85	85
28	27	贺*亮	男	89	86	89	88

图 a

性别	总评成绩
0 女	324.0
1 男	493.0

图 b

①已知根据“性别”进行分组求和，故使用groupby()函数，且不将分组依据行作为结果中的索引，故设置参数as_index为False。②根据图b可知，程序按照df1对象的“总评成绩”升序排序，通过选项ascending=True确定升序，ascending默认值为True。

5. 小王收集了某校高三学生首考4门学科成绩，并存储在“cjfx.xlsx”文件中，其中部分数据如图a所示。编写如下Python程序，对成绩进行分析。

(1) 读取“cjfx.xlsx”文件中的数据，计算4门学科总分，然后求出各班总分的平均分，并按班级总分平均分降序排序，输出前6个班级的信息，如图b所示。

```
import pandas as pd
df=pd.read_excel("cjfx.xlsx")
subjects=list(df.columns[2:])
df["总分"]=df[subjects].sum(axis=1)
df_avg=round(df._____①_____, 2)
df_avg=df_avg.sort_values("总分", ascending=False)
print(_____②_____)
```

①划线①处应填入的代码为_____ **B** _____ (单选，填字母)。

- A. groupby("班级").总分.mean()
- B. groupby("班级", as_index=False).总分.mean()
- C. groupby("班级").mean()
- D. groupby("班级", as_index=False).mean()

②划线②处应填入的代码为 **df_avg.head(6)**

	A	B	C	D	E	F
1	班级	姓名	英语	选科1	选科2	选科3
2	高三(01)班	张*惟	140	99	100	83
3	高三(01)班	荣*康	138	99	87	81
4	高三(01)班	蔡*婷	101	83	82	86
5	高三(01)班	张*好	143	83	93	98
6	高三(01)班	吴*伯	149	87	84	82
7	高三(01)班	沈*君	150	84	80	100

图 a

	班级	总分
7	高三(08)班	397.16
15	高三(16)班	397.14
11	高三(12)班	396.89
6	高三(07)班	396.66
13	高三(14)班	396.62
2	高三(03)班	395.21

图 b

	班级	姓名	总分	校名	班名
6	高三(01)班	林*瑜	429	14	1
20	高三(01)班	连*思	423	37	2
0	高三(01)班	张*惟	422	42	3
19	高三(01)班	张*雯	422	42	3
3	高三(01)班	张*好	417	73	5
5	高三(01)班	沈*君	414	93	6

图 c

(2) 根据每位学生的总分计算校名次和班名次,并按“班级”和“班名”升序排序后输出,部分数据如图c所示。代码如下,请在划线处填入合适的代码。

```
df=df.drop(columns=subjects)
df["校名"]=0
df["班名"]=0
for index in df.index:
    score=_____① df.at[index,"总分"]
    num_school=len(df[df["总分"]>score])+1
    df.at[index,"校名"]=num_school
    df_class=df[_____② df["班级"]==df.at[index,"班级"]]
    num_class=len(df_class[df_class["总分"]>score])+1
    df.at[index,"班名"]=num_class
df=df.sort_values(["班级","班名"],ascending=[True,True])
print(df)
```

(1) ①根据题目可知，对“班级”进行分组，并计算每组的“总分”列的平均值，使用`groupby()`函数和`mean()`函数，保留原始的“班级”列作为`DataFrame`的一部分，故设置`as_index=False`。②输出前6个班级的信息，使用`head()`函数。 (2) ①处`score`为当前学生的总分，故填入代码为`df.at[index, "总分"]`。②处筛选当前学生所在班级的`DataFrame`，故填入代码为`df["班级"]==df.at[index, "班级"]`。

6. [2025绍兴模拟] 2025年1月的部分学生订餐数据如图a所示。现要统计1月学生最喜欢的3个套餐（即订餐数量最多的3个套餐），并绘制如图b所示的柱形图。实现上述功能的部分Python程序如下，请在划线处填入合适的代码。

POS 机 ID	学生 ID	消费金额	消费时间	套餐
1	83124151	8	2025-01-01 17:05:10	A
2	83124150	8	2025-01-01 17:05:13	A
2	83124164	8	2025-01-01 17:06:12	A
13	83124163	10	2025-01-01 17:06:13	C
16	83124146	11	2025-01-01 17:06:14	D
8	83124156	9	2025-01-01 17:07:15	B
4	83124148	8	2025-01-01 17:07:25	A
15	83124170	10	2025-01-01 17:08:17	C

图 a

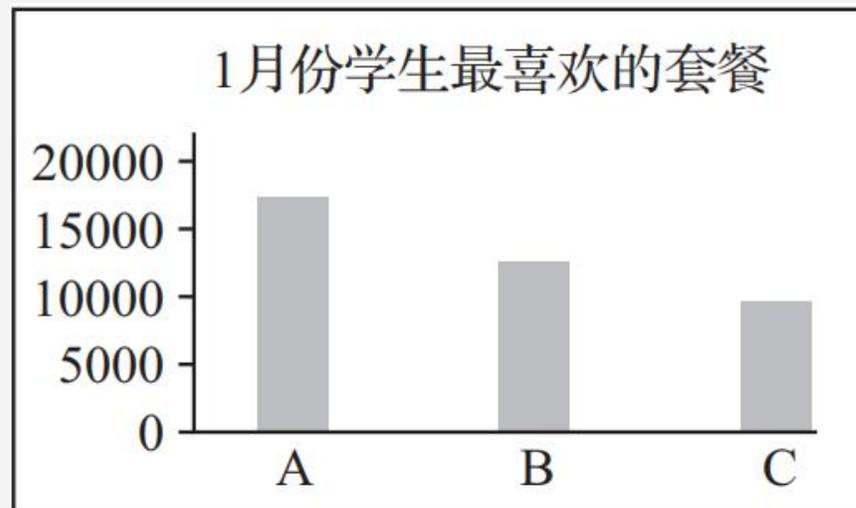


图 b

```
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_excel("食堂刷卡数据.xlsx")
df["月份"]=1 #插入“月份”列
for i in df.index:
    df.at[i,"月份"]=int(df.at[i,"消费时间"][5:7])
df1=_____① df[df.月份==1] 或 df[df["月份"]==1]
df1=df1.groupby("套餐",as_index=False).count()
#把列标题“学生ID”改为“数量”
df1=df1.rename(columns={"学生ID":"数量"})
df1=df1.sort_values("数量",ascending=False)
df1=df1.head(3)
plt.title("1月份学生最喜欢的套餐")
plt.bar(_____②, df1.数量)
plt.show() df1.套餐 或 df1["套餐"]
```

7. [2023.6浙江选考T14, (5)题] 小华将系统中某天 24 小时的湿度数据导出, 部分数据如图 a 所示(时间格式为“时:分:秒”), 分析每小时的最高湿度值, 线形图如图 b 所示, 部分 Python 程序如下:

时间,类型,监测值
00:00:00,湿度,75
00:01:00,湿度,75
00:02:00,湿度,75
00:03:00,湿度,75
00:04:00,湿度,75
00:05:00,湿度,74
00:06:00,湿度,75
00:07:00,湿度,75
00:08:00,湿度,75

图 a

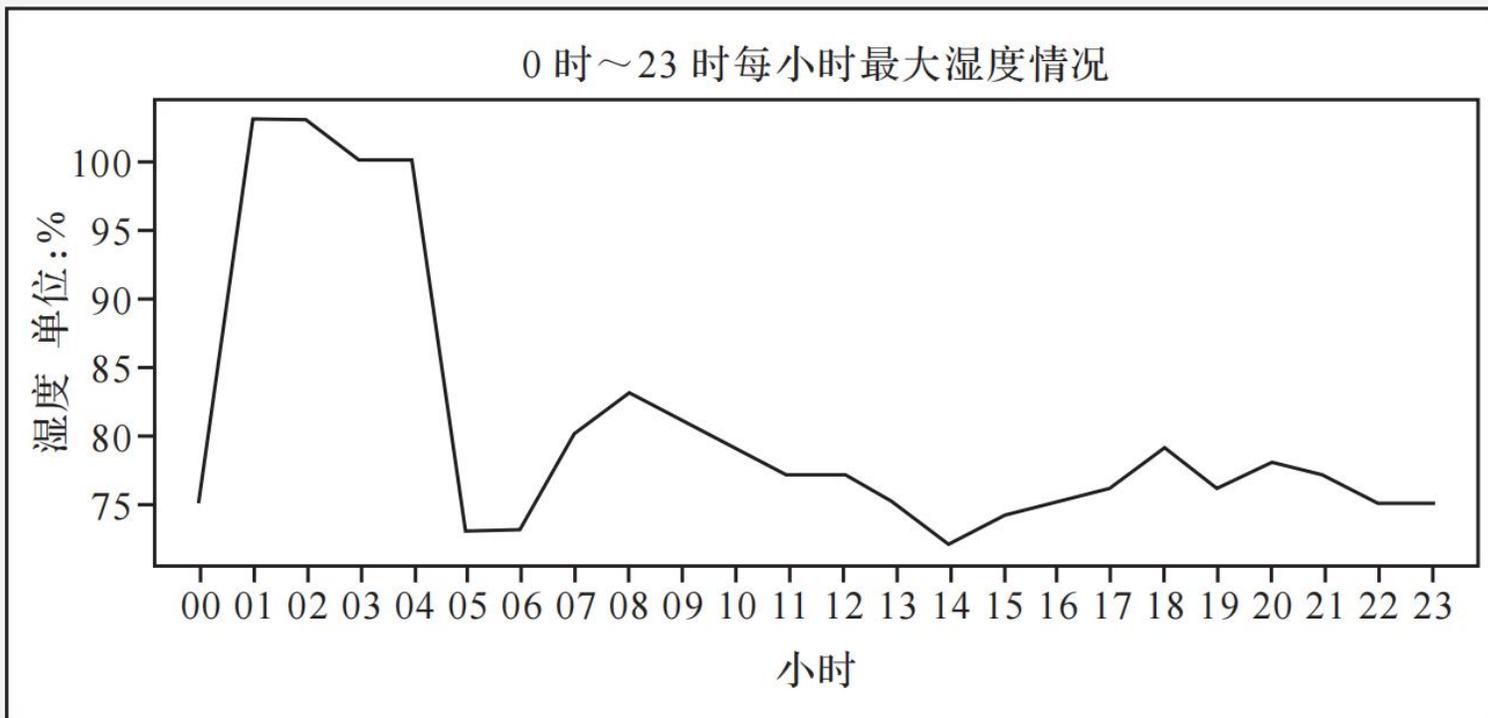


图 b

```
import pandas as pd
import matplotlib.pyplot as plt
#读取文件data.csv中的数据
dft=pd.read_csv('data.csv')
dft.insert(0,'小时','') #插入列
for i in dft.index:
    #通过行标签和列标签选取单个值
    t=dft.at[i,'时间']
    dft.at[i,'小时']=t[0:2]
#分组求最大值
dfh=dft.groupby('小时', as_index=False).max()
plt.plot(dfh['小时'],dfh['监测值']) #绘制线形图
#设置绘图参数,显示如图b所示线形图,代码略
```

(1) 请在程序中划线处填入合适的代码。
(2) 小华分析线形图发现存在湿度值大于等于100的噪声数据,要删除dft对象中的噪声数据,下列代码段中,能正确实现的有 AD (多选,填字母)。

- A. dft=dft[dft['监测值']<100]
- B. dft=dft['监测值']<100
- C. n=len(dft[dft['监测值']>=100])
dft=dft.sort_values('监测值') #升序排序
dft=dft.tail(n) #获取尾部数据行
- D. for i in dft.index:
if dft.at[i,'监测值']>=100:
dft=dft.drop(i) #删除行

8. [2025 浙江名校协作体] 某城市 2024 年前 13 周新能源汽车销量表部分数据如图 a 所示, 表中共有四个汽车品牌, 记录着各品牌每日的销量数据, 请回答下列问题:

日期	销量	厂商属性	品牌
2024/01/07	50	自主	C
2024/01/22	47	自主	C
2024/01/28	38	合资	B
.....			
2024/02/08	11	自主	B
2024/02/09	9	自主	A
2024/02/09	0	合资	D
2024/03/31	0	合资	B

图 a

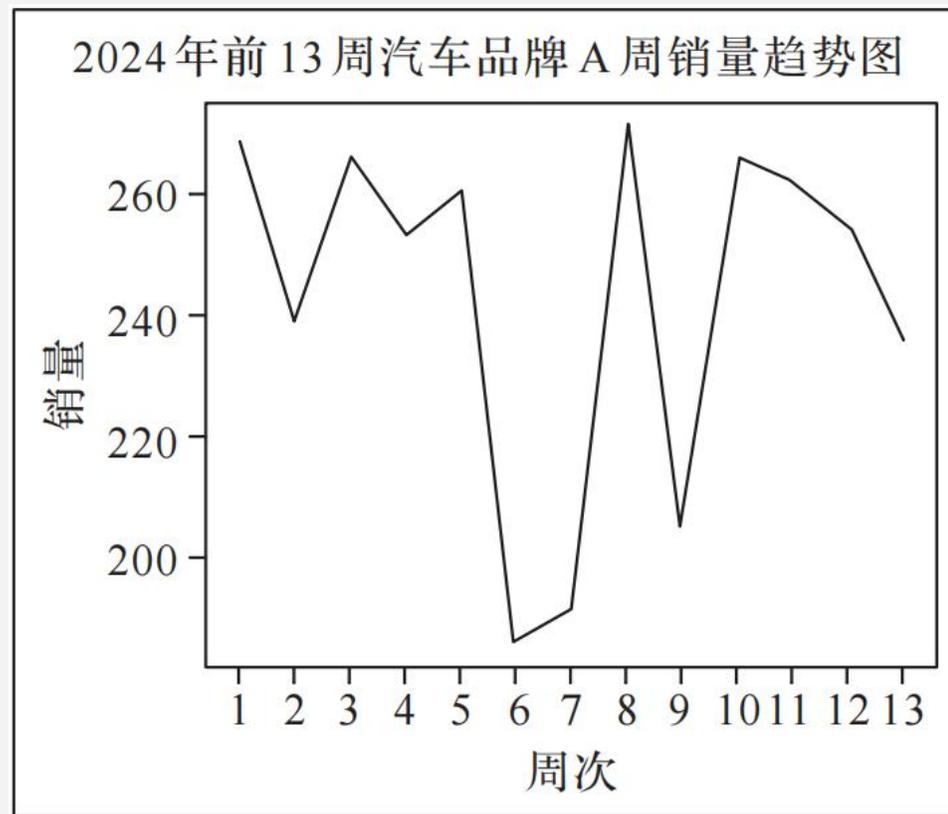


图 b

(1) 定义week_sale函数，功能为读取该城市前13周各品牌新能源汽车每日销量表，返回某品牌新能源汽车的每周销量。函数代码如下，请在划线处填入合适的代码。

```
def week_sale(brand):
    import pandas as pd
    df=pd.read_excel("sale.xlsx")
    df1=df[df["品牌"]==brand]
    df1=df1.sort_values("_____①_____")
    df1=df1.reset_index(drop=True) #索引重建
    sale=[0]*13
    s=0
    for i in df1.index:
        day=df1.at[i,"销量"]
        _____② s+=day 或 s=s+day
        if i%7==6:
            sale[i//7]=s
            _____③ s=0
    return sale
```

(2) 输入某新能源汽车品牌（例如品牌A），绘制前13周该品牌周销量线形图如图b所示。部分Python代码如下，加框处应填入的正确代码为

C (单选，填字母)。

```
import matplotlib.pyplot as plt
brand=input("请输入品牌：")
sales=week_sale(brand)
weeks=[i for i in range(1,14)]
plt.plot(_____ )
'''显示如图b所示的线形图，设置绘图参数，代码略'''
```

A. df["周次"], df["销量"]
 B. df1["weeks"], df1["sales"]
✓ C. weeks, sales

9. [2025 湖州联考] 某校德育处每周进行寝室内务检查, 从“床铺打理”等 5 个方面进行评分, 每项 20 分, 满分 100 分。高一年级内务检查扣分情况部分界面如图 a 所示, 每项扣分范围为 0~9 分。

寝室号	班级	床铺打理	物品整理	环境清扫	垃圾分类	寝室美化	总分
1101	高一(1)班	-6	-1	-6		-3-2	
1102	高一(1)班	-5	-2	-1		-6-5	
1103	高一(1)班	-0	-1	-5		-2-1	
2101	高一(1)班	-1	-4	-3		-6-0	
.....							
1610	高一(10)班	-0	-3	-5		-2-0	
2608	高一(10)班	-0	-6	-6		-5-3	
2609	高一(10)班	-6	-1	-1		-6-5	
2610	高一(10)班	-1	-5	-2		-5-6	

图 a

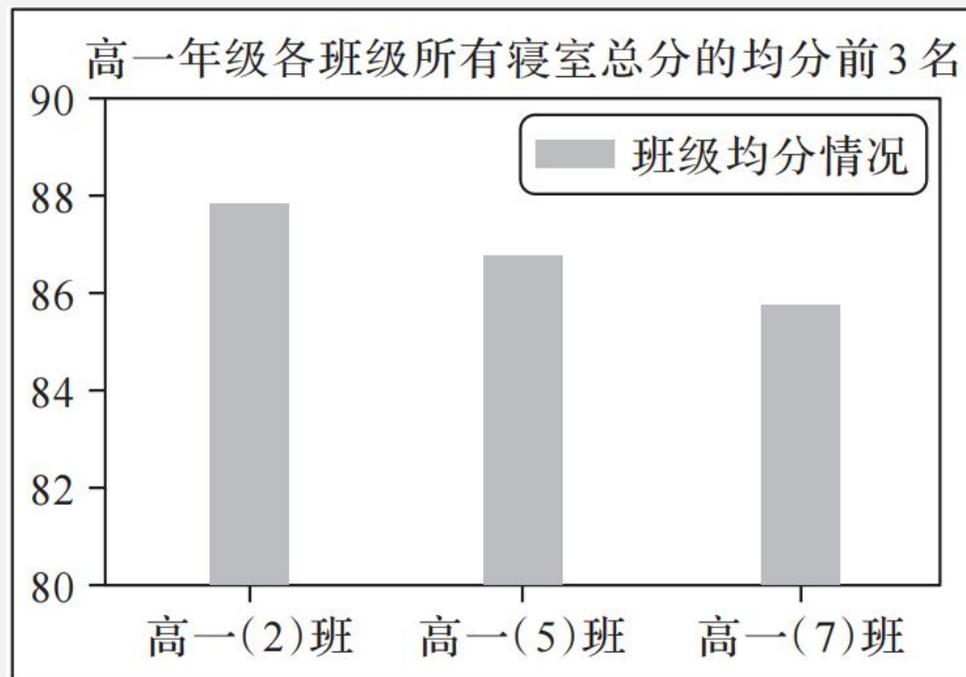


图 b

请回答下列问题：

(1) 整理数据时，误将“垃圾分类”和“寝室美化”两列的分数录在同一列，为纠正该错误，并计算各寝室的总分，编写如下 Python 程序，请在划线处填入合适的代码。

```
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_excel("data.xlsx")
for i in df.index: #DataFrame 的行索引
    _____ ① k=df.at[i,"寝室美化"]
    df.at[i,"垃圾分类"]=k[0:2]
    df.at[i,"寝室美化"]=k[2:]
    s=100
    #存放各列的列标题
    for j in df.columns[2:7]:
        s=s+int(df.at[i,j])
    df.at[i,"总分"]=_____ ② S
```

(2) 统计各班级所有寝室总分的均分情况，找到年级前3名，并绘制如图b所示的柱形图。

```
x=df2. 班级
```

```
y=df2. 总分
```

```
plt.bar(x, y, label="班级均分情况")
```

''' 设置绘图参数，显示如图b所示的柱形图，代码略'''

方框中应填入的语句依次为 ③②④ (选3项，填数字序列)。

①#升序排序

```
df1=df.sort_values("总分", ascending=True)
```

②#降序排序

```
df1=df1.sort_values("总分", ascending=False)
```

③#分组求平均

```
df1=df.groupby("班级", as_index=False).mean()
```

④df2=df1.head(3) #获取前3条数据

(3) 观察图a可知, 每个班级都配有多个寝室。现要筛选出“高一(2)班”各个寝室的扣分情况, 以便进一步分析, 可在第(1)小题处理结果的基础上, 再运行如下语句, 请在划线处填入合适的代码。

```
df3=df[df.班级=="高一(2)班" 或 df["班级"]=="高一(2)班"]
```

(1) ①for循环处理合并的列, 然后根据列名将“寝室美化”列的数据拆分开来, 根据下面代码可知, 该处将“寝室美化”列的数据赋值给变量k, 故填入代码为k=df.at[i, "寝室美化"]。②在计算总分的地方需要从满分中减去扣分项, 故填入代码为s。 (2) 方框中需要对每个班级的总分进行平均计算, 选择前3名班级, 并进行相应排序。需要先对班级进行平均分的计算, 即df1=df.groupby("班级", as_index=False).mean(), 再降序排序, 即df1=df1.sort_values("总分", ascending=False), 最后取前3名, 即df2=df1.head(3)。 (3) 筛选出特定班级“高一(2)班”的数据, 筛选条件可以是df.班级=="高一(2)班"或df["班级"]=="高一(2)班"。

10. [2024.1 浙江选考] 某学院举行运动会，比赛设跳高、100 米等项目，每个项目分男子组和女子组。现要进行报名数据处理和比赛成绩分析。请回答下列问题：

	专业	学号	姓名	性别	项目
0	软件工程	S10111	钱*然	男	跳高
1	软件工程	S20212	石*如	女	100 米
2	软件工程	S30212	宋*尘	男	100 米
...
26	软件工程	S10622	王*娟	女	400 米
27	软件工程	S10919	王*栩	女	200 米
28	软件工程	S30110	叶*涛	男	100 米

图 a

专业	学号	姓名	性别	项目	名次	得分
软件工程	S30110	叶*涛	男	跳远	5	
计算机	D30101	朱*奕	女	100 米	2	
电子信息	C10522	赵*宇	男	铁饼	4	
...	
电子信息	C30211	郑*珥	女	跳远	15	
人工智能	A20109	裘*晨	女	跳高	16	
人工智能	A20109	裘*晨	女	跳远	7	

图 b

(1) 运动会报名规则为：对于每个项目的男子组和女子组，每个专业最多各报 5 人（如“软件工程”专业在男子跳高项目中最多报 5 人）。软件工程专业的报名数据保存在 DataFrame 对象 df 中，如图 a 所示。若要编写 Python 程序检查该专业男子跳高项目报名是否符合规则，下列方法中，正确的是 **A** （单选，填字母）。

A. 从 df 中筛选出性别为“男”的数据 dfs，再从 dfs 中筛选出项目为“跳高”的数据，判断筛选出的数据行是否超过 5 行

B. 对 df 中数据按性别排序并保存到 dfs 中，再从 dfs 中筛选出项目为“跳高”的数据，判断筛选出的数据行是否超过 5 行

C. 从 df 中筛选出项目为“跳高”的数据 dfs，判断 dfs 中是否有连续 5 行以上的男生数据

(2) 运动员比赛成绩的部分数据如图 b 所示。根据已有名次计算得分，第 1 名至第 8 名分别计 9, 7, 6, 5, 4, 3, 2, 1 分，第 8 名之后计 0 分。实现上述功能的部分 Python 程序如下，请在程序中划线处填入合适的代码。

```
import pandas as pd
import matplotlib.pyplot as plt
'''读取如图 b 所示数据，保存到 DataFrame 对象 df1 中，代码略'''
f=[9, 7, 6, 5, 4, 3, 2, 1]
for i in range(0, len(df1)):
    #通过行、列标签取单个值
    rank=df1.at[i, "名次"]
    score=0
    if rank<=8:
        score=f[rank-1] 或 score=f[int(rank)-1]
    df1.at[i, "得分"]=score
```

(3) 根据上述 df1 中的得分数据, 统计各专业总分, 绘制如图 c 所示的柱形图, 实现该功能的部分 Python 程序如下:

```
df2=df1.groupby(" 专业 ",as_index=False).sum() #分组求和
```

#设置绘图参数, 代码略

plt.bar(x,y) #绘制柱形图

①请在程序中划线处填入合适的代码。

②程序的方框中应填入的正确代码为

B (单选, 填字母)。

A. x=df1["专业"]

y=df1["总分"]

B. x=df2["专业"]

y=df2["得分"]

C. df1["专业"]="专业"

df1["总分"]="总分"

D. df2["专业"]="专业"

df2["得分"]="得分"

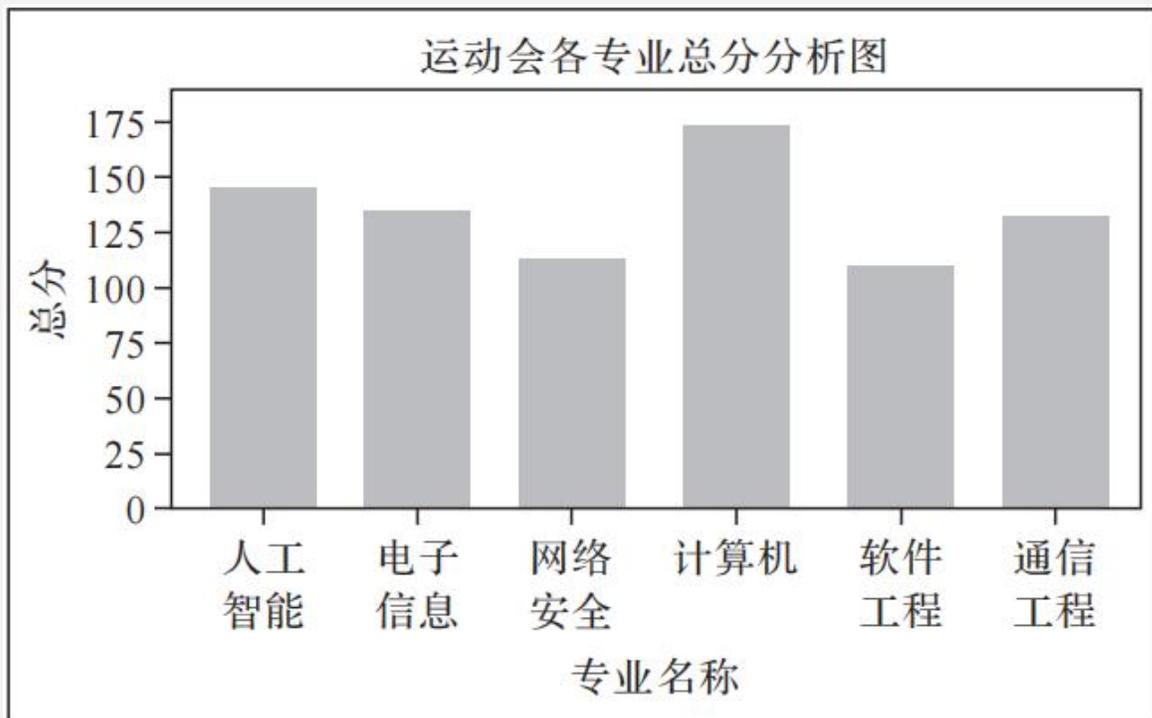


图 c

(1) 选项A首先从df中筛选出性别为“男”的数据，然后再次筛选出项目为“跳高”的数据，最后判断筛选出的数据行是否超过5行。选项B不需要对数据进行排序，因为关注点是性别为“男”且项目为“跳高”的数据行数，排序并不会改变数据行的数量。选项C只筛选出了项目为“跳高”的数据，没有进一步筛选出性别为“男”的数据。

(2) if语句判断当前名次是否小于或等于8（即是否在前8名之内），若是，则根据名次计算得分，使用数组索引，因为数组的索引是从0开始的，而名次是从1开始的，所以需要减去1，然后将对应的得分赋值给变量score，故填入代码为`score=f[rank-1]`或`score=f[int(rank)-1]`。

(3) ①根据图c可知，X轴数据根据专业进行分组求和，故填入代码为专业。②根据代码“`plt.bar(x,y)`”可知，需要对x和y进行赋值，而值为进行分组求和后的数据。

11. [2025宁波联考] 某教师收集全校学生某科目前后两次测试的成绩数据, 以了解学生的学习情况。成绩数据存储在“score.xlsx”文件中, 部分数据如图a所示。“学生编号”列中的第3~4位是该学生的班级号, 第5~6位是其学号。

	A	B	C	D
1	学生编号	测试一分数	测试二分数	分数增量
2	cx0101	96	96	0
3	cx0102	88	96	8
4	cx0103	98	98	0
5	cx0104	97	97	0
6	cx0105	96	99	3
7	cx0106	90	93	3
8	cx0107	89	92	3

图 a

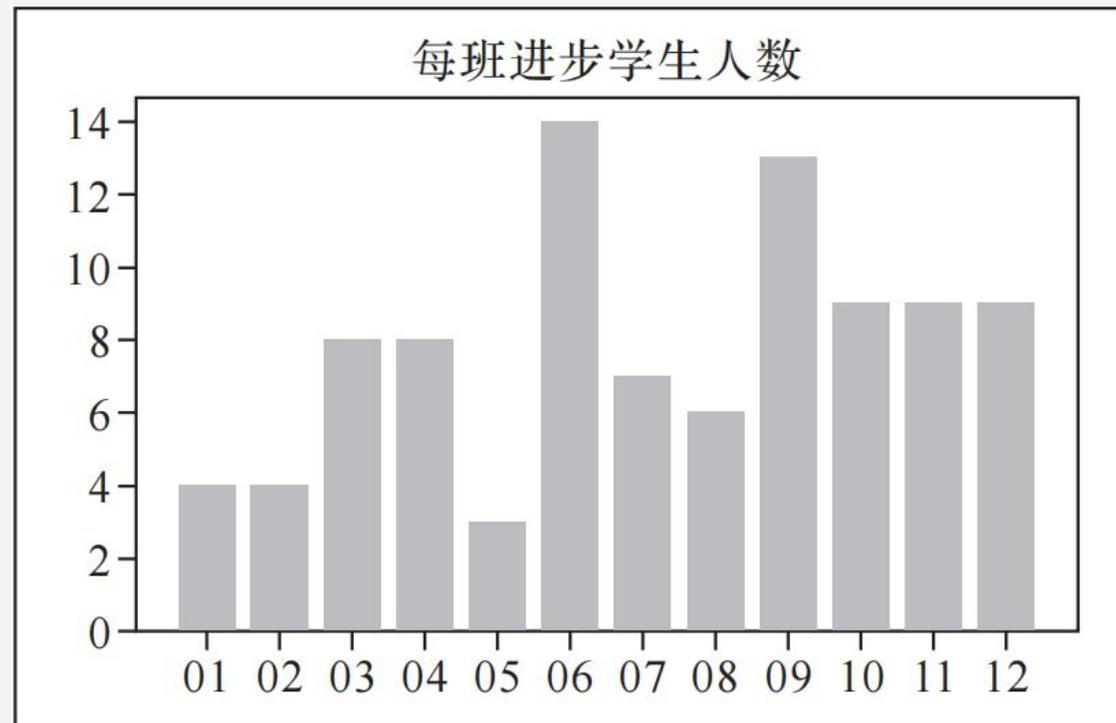


图 b

请回答下列问题：

(1) 为获取测试一到测试二分数有进步的学生数据（分数增量=测试二分数-测试一分数），划线处应填入的正确代码为_____ **A** _____（单选，填字母）。

```
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_excel("score.xlsx")
df=_____
print("进步学生数据：")
print(df)
```

- A. df[df["分数增量"]>0]
- B. df[df."分数增量">0]
- C. df[df["分数增量"]>=0]

(2) 统计每班测试一到测试二分数有进步的学生人数并绘制如图b所示的柱形图，输出进步人数最多的班级号（不存在并列），部分Python程序如下，请在划线处填入合适的代码。

```
df.insert(0, "班级", "") #在最前面插入"班级"列
for i in df.index:
    s=df.at[i, _____ ①"学生编号"
    df.at[i, "班级"]=s[2:4] #取2位班级号
df1=df.groupby("班级", as_index=False).分数增量.count()
x=df1.班级
y=_____ ②df1.分数增量
plt.bar(x, y) #绘制柱形图
#其他绘图参数，代码略

#取df1中第一行班级列的值
bj=df1.values[0][0]
print("进步人数最多的班级为", bj, "班")
```

(3) 加框处应填入的代码有 **AD**（多选，填字母）。

- A. df1=df1.sort_values("分数增量", ascending=False) #降序排序
- B. df1=df1.sort_values("分数增量", ascending=True)
- C. df1=df1.sort_values("分数增量", ascending=False).tail(1)
- D. df1=df1.sort_values("分数增量", ascending=True).tail(1)

(4) 由图b可知, 进步人数10人以上的班级数量为 2。

(1) 为获取测试一到测试二分数有进步的学生数据, 此处需要筛选出分数增量大于0的学生数据, 故填入代码为`df[df["分数增量"]>0]`。(2) ①处获取学生编号列的数据, 故填入代码为"学生编号"。②处使用“分数增量”列的计数结果作为Y轴数据, 故填入代码为`df1.分数增量`。(3) 根据输出语句可知, `bj`为进步人数最多的班级, 可以按分数增量降序排序, 此时, 进步人数最多的班级在第一行位置, 即`df1=df1.sort_values("分数增量", ascending=False)`; 也可以按分数增量升序排序并取最后一行, 得到进步人数最多的班级, 即`df1=df1.sort_values("分数增量", ascending=True).tail(1)`。(4) 由图b可知, 进步人数10人以上的班级为6班和9班。

12. [2023.1浙江选考] 小红收集了部分城市2021年全年每天PM2.5、PM10、CO浓度数据，每天的数据分别保存在以8位日期字符串命名的CSV文件中，部分文件如图a所示，每个文件记录了一天24小时的监测数据，示例如图b所示。

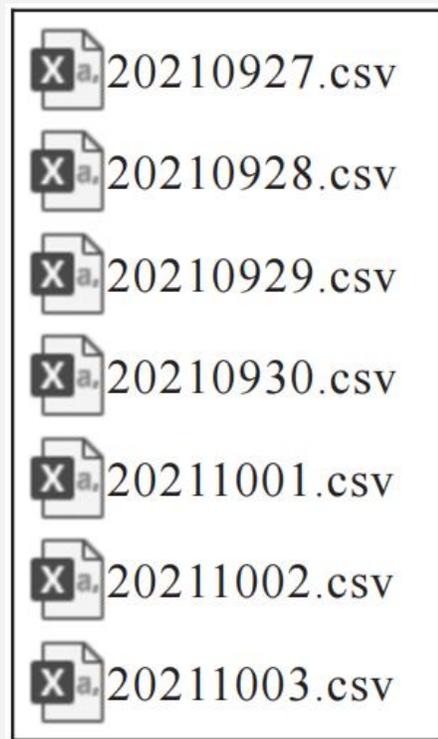


图 a

小时	类型	城市 A	城市 B	城市 C	城市 D
0	PM2.5	61	22	51	48
0	PM10	95	71	98	66
0	CO	1.8	0.3	1.7	1.5
1	PM2.5	57	26	45	57
1	PM10	101	70	188	35
...					
22	CO	2	0.3	1.4	1.9
23	PM2.5	46	25	48	46
23	PM10	89	81	107	80
23	CO	1.5	0.3	1.5	1.7

图 b

为统计分析城市A全年各月份PM2.5的月平均浓度（当月的日平均浓度的平均值），编写Python程序。请回答下列问题：

（1）定义pmday函数，功能为：读取某天的CSV文件，返回城市A当天 PM2.5的日平均浓度。函数代码如下，划线处应填入的代码为_____ **D** _____（单选，填字母）。

A. `df['类型'] == 'PM2.5'`

B. `df['类型' == 'PM2.5']`

C. `df[df['类型']] == 'PM2.5'`

D. `df[df['类型'] == 'PM2.5']`

```
import pandas as pd
```

```
def pmday(dayfile):
```

```
    #读取文件dayfile中的数据
```

```
    df=pd.read_csv(dayfile)
```

```
    df=_____
```

```
    #返回城市A当天PM2.5的日平均浓度
```

```
    return df['城市A'].mean()
```

(2) 统计城市A各月份PM2.5的月平均浓度并绘制线形图，部分Python程序如下，请在划线处填入合适的代码。

```
import matplotlib.pyplot as plt
def tstr(t):
    if t<10:
        return '0'+str(t)
    else:
        return str(t)
pm=[0]*12
#2021年每月天数
mdays=[31,28,31,30,31,30,31,31,30,31,30,31]
for m in range(12):
    sm=0
    mstr=tstr(m+1)
    for d in range(____① mdays[m] 或 0,mdays[m]
        dstr=tstr(d+1)
        dayfile='2021'+mstr+dstr+'.csv'
        sd=pmday(dayfile)
        ____② sm+=sd 或 sm=sm+sd
    pm[m]=sm/mdays[m]
x=[1,2,3,4,5,6,7,8,9,10,11,12]
y=____③ pm
plt.plot(x,y) #绘制线形图
#设置绘图参数，显示如图c所示线形图，代码略
```

(3) 城市A 2021年PM2.5年平均浓度为34.6微克/立方米。由图c可知，城市A 2021年PM2.5月平均浓度超过年平均浓度的月份共 5 个。

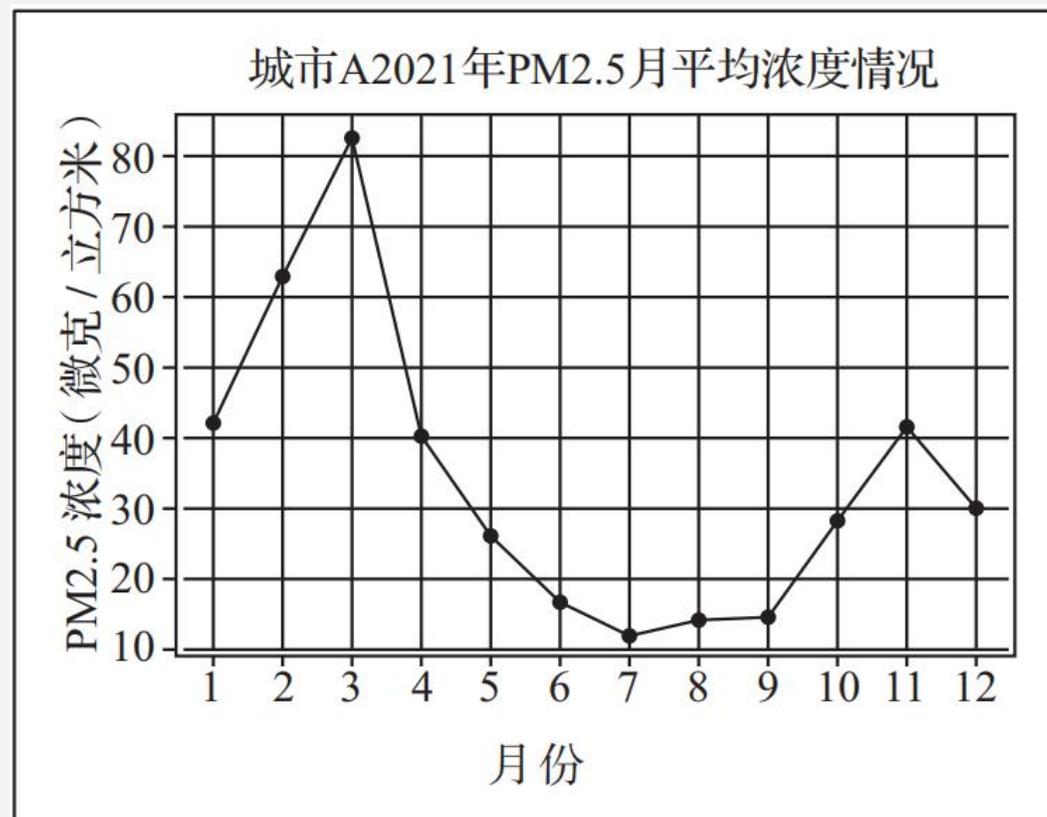


图 c

(1) 观察程序可知, `pmday`函数实现读取某天的CSV文件, 返回城市A当天PM2.5的日平均浓度, 筛选条件为`df['类型']=='PM2.5'`, 筛选对象为`df`, 故填入代码为`df[df['类型']=='PM2.5']`, 正确的是选项D。

(2) ①处变量`m`表示一年的12个月, `d`的数值表示12个月中的每个月的天数, 故填入代码为`mdays[m]`或`0,mdays[m]`。②处由代码“`sd=pmday(dayfile)`”和“`pm[m]=sm/mdays[m]`”可知, `sd`为PM2.5的日平均浓度, `sm`为每个月中每一天平均值的累加, 故填入代码为`sm+=sd`或`sm=sm+sd`。③处观察图像可知, Y轴的数据来自每个月的PM2.5的平均浓度, 故填入代码为`pm`。

(3) 城市A2021年PM2.5年平均浓度为34.6微克/立方米。观察图c可知, 城市A2021年PM2.5月平均浓度超过年平均浓度的月份共有5个, 分别是1、2、3、4、11月。