

【课堂练习 2】

1.小明将一些 App 的活跃人数（单位：万人）数据存在文件“app.xlsx”中，部分界面如图所示。

```
import pandas as pd
#读取文件“app.xlsx”数据
book1= ① _____
#计算 10 月人数之和
book1_sum= ② _____
#计算 11 月人数平均值
book1_aver= ③ _____
#按应用领域分组统计，分别对各月人数求和
book1_g= ④ _____
#按 11 月人数降序排序
book1_sort= ⑤ _____
print("10 月人数之和：",book1_sum)
print("11 月人数平均值：",book1_aver)
print("各应用领域各月人数和：",book1_g)
print("11 月人数降序排序情况：",book1_sort)
```

	A	B	C	D	E
1	APP名称	应用领域	10月人数	11月人数	12月人数
2	360手机卫士	安全管理	13855.8	14422.3	16017.2
3	腾讯手机管家	安全管理	17992.5	18135.2	19688.5
4	微信	即时通讯	74257	75101.6	75616.5
5	QQ	即时通讯	53931.5	54913.7	55216.7
6	QQ浏览器	浏览器	20759.1	20540.3	22438.6
7	手机百度	搜索	18401.2	18571.3	19960.2
8	QQ音乐	移动音乐	13707.9	13709.1	13691.5
9	酷狗音乐	移动音乐	19407.5	19237.8	20847.05
10	酷我音乐	移动音乐	4935.37	4864.13	5480.75
11	支付宝	支付	29897	31408.9	33146.8
12	京东	综合电商	6441.72	7596.09	6854.08
13	淘宝	综合电商	27153.9	27737.6	29807.4
14	爱奇艺视频	综合视频	20319.4	20543.8	22047.9
15	芒果tv	综合视频	5283.03	5165.8	5108.08
16	腾讯视频	综合视频	19833.5	20412.6	21332.2
17	优酷视频	综合视频	12576.4	12264.5	13840.4

2.小明将某天 24 小时的楼层停靠数据导出，并保存在“data.csv”文件中，部分数据如图所示（时间格式为“时:分:秒”）。现分析各小时时段停靠次数最多的楼层（1 楼不参与统计）的部分 Python 代码如下，请在划线处填入合适的代码。

```
#导入模块，代码略
df=pd.read_csv("data.csv")
df.insert(0,"小时","")
for i in df.index:
    t=df.at[i,"时间"]
    df.at[i,"小时"]=int(t[0:2])
xs=[]
cnt=[]
for i in range(24):
    dft=df[df["小时"]==i]
    if len(dft)>0 :
        dfg=dfg.groupby(② "楼层" _____,as_index=False).count()
        dfg=③ dfg[dfg.楼层!=1]或dfg[dfg["楼层"]!=1] #筛出除 1 楼外楼层
        dfg=dfg.sort_values(④ _____)
        xs.append(i) "小时",ascending=False
        cnt.append(dfg["楼层"].values[0]) #获取第 i 小时停靠最多的楼层数
```

时间,楼层
01:17:00,6
01:33:00,6
01:45:00,5
02:11:00,5
02:23:00,6
02:25:00,5

【课后练习 2】pandas 综合练习

1. 小李收集了某奶茶门店 2024 年 1 月的销售数据，如图 a 所示。

	A	B	C	D	E	F	G	H
1	日期	订单号	点单状态	饮品名称	品类	单价	数量	金额
2	2024/1/1	1	成功	芋泥厚厚牛乳	芋泥	14		0
3	2024/1/1	1	成功	龙井香青团	轻牛乳	15	1	15
4	2024/1/1	1	成功	杨枝甘露	果茶	15	1	15
5	2024/1/1	2	退单	茉莉奶芙	轻牛乳	15	2	30
6	2024/1/1	2	退单	大叔奶茶	奶茶	11	1	11
7	2024/1/1	3	成功	生椰榴莲	果茶	20	1	20
8	2024/1/1	4	成功	芝士莓莓	果茶	22	1	22
312	2024/1/30	16	成功	酒酿小丸子	奶茶	12	1	12
313	2024/1/30	17	成功	云岭茉莉椰	轻牛乳	12	2	24
314	2024/1/30	18	成功	布蕾脆腕奶芙	奶茶	16	1	16

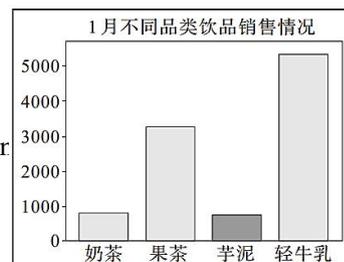
为统计分析该门店不同品类饮品的销售情况。编写 Python 程序，请回答下列问题：

(1) 读取文件，筛选出点单成功的数据，代码如下。

```
import pandas as pd
df=pd.read_excel("salelist.xlsx")
df1=  B df[df.点单状态=="成功"] 或 df[df["点单状态"]=="成功"] 
```

划线处应填入的代码为_____（单选，填字母）。

- A.df[df["金额"]>0]
- B.df[df.点单状态=="成功"]
- C.df1[df1["金额"]>0]
- D.df1[df1.点单状态=="成功"]



(2) 统计该月不同品类饮品销售数量，并绘制柱形图（图 b），部分 Python 入合适的代码。

```
import matplotlib.pyplot as plt
df2=df1.groupby("  品类①  ",as_index=False)
```

```
df2=df2.数量 .sum()
plt.figure()
plt.bar(df2.品类,  ② df2["金额"] )
```

index	品类	日期	订单号	点单状态	饮品名称	单价	数量	金额
0	奶茶	800
1	果茶	3200
2	芋泥	810
3	轻牛乳	5200

“设置绘图参数，显示结果如图 b 所示，代码略”
plt.show()

2. 小明收集了浙江省某些热门景点 3 月 15 日的温度数据，存储在“weather.xlsx”文件中，部分数据如图 a 所示。现利用 pandas 模块对该表进行数据分析，按“地级市”分类统计各市各时间点热门景点的平均温度，最后绘制线形图如图 b 所示。请在划线处填入合适的代码。

	A	B	C	D	E	F	G	H	I	J
1	地级市	景点	08时	11时	14时	17时	20时	23时	02时	05时
2	宁波	松兰山	12	15	16	14	13	12	12	12
3	杭州	宋城	13	18	22	21	17	17	16	15
4	绍兴	沈园	13	20	22	20	17	17	16	15
5	杭州	龙门古镇	12	20	25	24	18	17	15	15
6	绍兴	鲁迅故里	13	20	22	20	17	17	16	15
7	温州	楠溪江	14	21	24	20	15	14	14	14
8	杭州	西湖	14	22	24	24	18	16	16	14
9	宁波	五龙潭	11	20	22	20	13	12	12	12
10	绍兴	西施故里	13	20	24	24	20	18	16	16
11	温州	江心屿	18	21	23	21	17	16	17	16
12	宁波	天一阁	13	18	20	19	16	15	14	14
13	温州	雁荡山	12	18	21	17	13	12	12	12

图 a

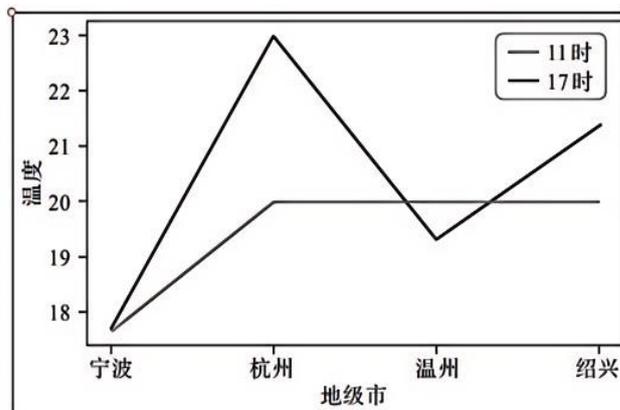
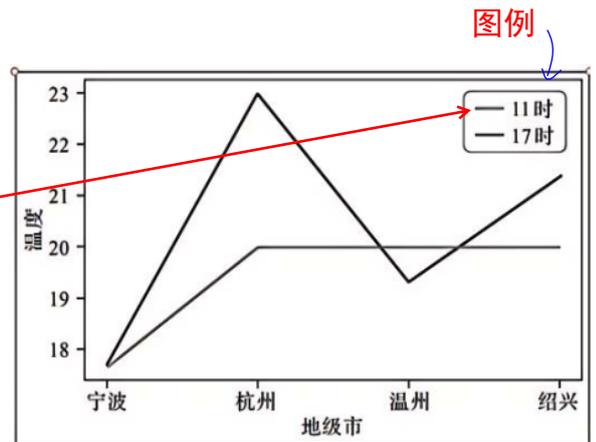


图 b

```
import pandas as pd
import matplotlib.pyplot as plt
plt.rc("font",family="SimSun")
df=pd.read_excel("weather.xlsx")
df_group=df.groupby("地级市",as_index=False)
df_ave=df_group.mean() #把groupby分两行代码操作
x=df_ave["地级市"]
y1=df_ave["11 时"]
y2=df_ave["17 时"]
plt.plot(x,y1,label="11 时")
plt.plot(x,y2,label="17 时")
plt.xlabel("地级市")
plt.ylabel("温度")
plt.legend() #显示图例 具体下节课讲解
plt.show()
```



3. 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*翔	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

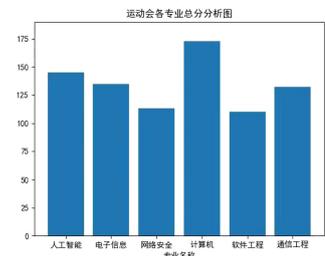


图 c

(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_____ (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,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)):
    rank=df1.at[i,"名次"] #通过行、列标签取单个值
    score=0
    if rank<=8:
        _____
        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