

1、创建1000条语、数、外、Python的考试成绩DataFrame，范围是0~150包含150，分别将数据保存到csv文件以及Excel文件，保存时不保存行索引。

提示：（说明没有这个库，安装一下，如果不出错，说明你的电脑上有这个库，直接过滤就行）

```
-> 2189     formatter.write(
2190         excel_writer,
2191         sheet_name=sheet_name,
2192
2193         c:\python38\lib\site-packages\pandas\io\formats\excel.py in write(self, writer, sheet_name, startrow, startcol, freeze_panes, engine, storage_options)
2194             # abstract class 'ExcelWriter' with abstract attributes 'engine',
2195             # 'save', 'supported_extensions' and 'write_cells' [abstract]
-> 2196             writer = ExcelWriter( # type: ignore[abstract]
2197                 writer, engine=engine, storage_options=storage_options
2198             )
2199
2200         c:\python38\lib\site-packages\pandas\io\excel\_openpyxl.py in __init__(self, path, engine, mode, storage_options, **engine_kwargs)
2201             28     ):
2202             29         # Use the openpyxl module as the Excel writer.
-> 2203             30         from openpyxl.workbook import Workbook
2204             31
2205             32         super().__init__(
2206
2207             ModuleNotFoundError: No module named 'openpyxl'
```

→ pip install openpyxl

```
import numpy as np
import pandas as pd

df = pd.DataFrame(np.random.randint(0,151,size = (1000,4)),columns=['语文','数学','英语','Python'])

# 查看前5个
display(df.head())

# 保存到csv文件中
df.to_csv('./score.csv',index=False)

# 保存到excel文件中
df.to_excel('./score.xlsx',index=False)
```

2、创建使用字典创建DataFrame，行索引是a~z，列索引是：身高（160-185）、体重（50-90）、学历（无、本科、硕士、博士）。身高、体重数据使用NumPy随机数生成，学历数据先创建数组edu = np.array(['无','本科','硕士','博士'])，然后使用花式索引从四个数据中选择26个数据。

```
import numpy as np
import pandas as pd

# 根据花式索引创建学位
edu = np.array(['无','本科','硕士','博士'])
index = np.random.randint(0,4,size = 26)
edu = edu[index]

df = pd.DataFrame({'身高':np.random.randint(160,186,size = 26),
                   '体重':np.random.randint(50,90,size = 26),
                   '学位':edu},index = list('qwertyuiopasdfghjklzxcvbnm'))

# 对行索引排序
df.sort_index()
```

3、使用题目二中的数据，进行数据筛选。

- 1、筛选索引大于 't' 的所有数据
- 2、筛选学历是博士，身高大于170或者体重小于80的学生

```

import numpy as np
import pandas as pd

# 根据花式索引创建学位
edu = np.array(['无', '本科', '硕士', '博士'])
index = np.random.randint(0, 4, size = 26)
edu = edu[index]

df = pd.DataFrame({'身高':np.random.randint(160, 186, size = 26),
                   '体重':np.random.randint(50, 90, size = 26),
                   '学位':edu}, index = list('qwertyuiopasdfghjk1zxcvbnm'))

# 问题一
cond = df.index > 't'
display(df[cond])

# 问题二
cond1 = df['学位'] == '博士'
cond2 = df['身高'] > 170
cond3 = df['体重'] < 80

cond = cond1 & (cond2 | cond3)
df[cond]

```

4、使用题目二中数据，开始学生们开始减肥

- 本科生减肥，减掉的体重统一是10
- 博士生减肥，减掉体重范围是5~10

```

import numpy as np
import pandas as pd

# 根据花式索引创建学位
edu = np.array(['无', '本科', '硕士', '博士'])
index = np.random.randint(0, 4, size = 26)
edu = edu[index]

df = pd.DataFrame({'身高':np.random.randint(160, 186, size = 26),
                   '体重':np.random.randint(50, 90, size = 26),
                   '学位':edu}, index = list('qwertyuiopasdfghjk1zxcvbnm'))

# 问题一
# 先找到本科生
cond = df['学位'] == '本科'
print('本科生减肥前: ')
display(df[cond])
print('本科生减肥后: ')
df.loc[cond, '体重'] -= 10
display(df[cond])

# 问题二
# 先找到博士生
cond = df['学位'] == '博士'
print('博士生减肥前: ')

```

```
display(df[cond])
print('博士生減肥后: ')
df.loc[cond, '体重'] -= np.random.randint(5,11, size = cond.sum())
display(df[cond])
```