

```
import pandas as pd

df = pd.read_excel("/Users/republic/考研信息/浙江工业大学/计算机学院各专业2022年硕士研究生复试考生名单.xlsx")

df[:5]
```

	序号	考生类型	学生代码	姓名	第一志愿	大类专业	录取专业	专业类型	外语	政治	业务课一	业务课二	总分	录取代码
0	1	第一批一志愿考生	103372210006660	陈天涯	浙江工业大学	081200计算机科学与技术	081200计算机科学与技术	全日制	80	73	110	120	383	ZG0812-001
1	2	第一批一志愿考生	103372210009385	伊秋华	浙江工业大学	081200计算机科学与技术	081200计算机科学与技术	全日制	74	66	111	130	381	ZG0812-002
2	3	第一批一志愿考生	103372210006633	吴康惠	浙江工业大学	081200计算机科学与技术	081200计算机科学与技术	全日制	78	66	117	120	381	ZG0812-003
3	4	第一批一志愿考生	103372210009382	严涵婷	浙江工业大学	081200计算机科学与技术	081200计算机科学与技术	全日制	77	77	99	121	374	ZG0812-004
4	5	第一批一志愿考生	103372210006653	沈哲辉	浙江工业大学	081200计算机科学与技术	081200计算机科学与技术	全日制	69	73	128	104	374	ZG0812-005

```
df.columns

Index(['序号', '考生类型', '学生代码', '姓名', '第一志愿', '大类专业', '录取专业', '专业类型', '外语', '政治',
      '业务课一', '业务课二', '总分', '录取代码'],
      dtype='object')

# pattern=r'^((?!士兵专项).)*$'
# df[df['录取代码'].str.extract(pattern)]
df = df[~df['录取代码'].str.contains('士兵')]

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 390 entries, 0 to 389
Data columns (total 14 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   序号         390 non-null    int64
 1   考生类型     390 non-null    object
 2   学生代码     390 non-null    int64
 3   姓名         390 non-null    object
 4   第一志愿     390 non-null    object
 5   大类专业     390 non-null    object
 6   录取专业     390 non-null    object
 7   专业类型     390 non-null    object
 8   外语         390 non-null    int64
 9   政治         390 non-null    int64
10   业务课一     390 non-null    int64
11   业务课二     390 non-null    int64
12   总分         390 non-null    int64
13   录取代码     390 non-null    object
dtypes: int64(7), object(7)
memory usage: 45.7+ KB
```

```
df.groupby(by=['录取专业'])['总分'].describe()
```

	count	mean	std	min	25%	50%	75%	max
录取专业								
081200计算机科学与技术	97.0	330.185567	21.762128	300.0	313.0	327.0	345.0	383.0
083500软件工程	41.0	317.365854	19.852904	284.0	302.0	320.0	329.0	360.0
085400电子信息	252.0	360.456349	18.752797	318.0	350.0	360.0	372.0	407.0

```
!jupyter nbconvert --to markdown process_data.ipynb
```

```
[NbConvertApp] Converting notebook process_data.ipynb to markdown
[NbConvertApp] Writing 5349 bytes to process_data.md
```