

In [1]:

```
import csv
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
```

In [2]:

```
# csv_file = csv.reader(open('a.txt', 'r', encoding='utf-8'))
# a=[]
```

In [3]:

```
# for i in csv_file:
#     a.append(i)
# b = pd.DataFrame(a)
```

In [4]:

```
# b.to_csv('target.csv')
```

In [5]:

```
c = pd.read_csv('target.csv', header=1, keep_default_na=0, index_col=0)
c
```

```
C:\ProgramData\Anaconda3\lib\site-packages\IPython\core\interactiveshell.py:3058:  
DtypeWarning: Columns (6) have mixed types. Specify dtype option on import or set  
low_memory=False.  
    interactivity=interactivity, compiler=compiler, result=result)
```

Out[5]:

	读者条码	姓名	读者级别	单位代码	单位	条形码	读者条码.1	处理时间	操作类型	主键码
0										
1	20131201021010	* 思其	学生	15	保险系	200541770	20131201021010	2016/12/27 17:25:48	J	9360
2	20131201021010	* 思其	学生	15	保险系	200541770	20131201021010	2017/1/3 14:26:29	H	9360
3	20121205021068	* 欢欢	学生	16	会计系	200592885	20121205021068	2013/3/5 14:14:35	H	8143
4	20121205021068	* 欢欢	学生	16	会计系	200592885	20121205021068	2013/1/17 10:53:04	J	8143
5	20110912043	* 欣瑜	学生	38	国际教育学院	200346287	20110912043	2013/12/12 11:51:29	J	1344
...

	读者条码	姓名	读者级别	单位代码	单位	条形码	读者条码.1	处理时间	操作类型	主键码
0										
374818	20120203022053	* 凯敏	学生	20	管理系	207848105	20120203022053	2014/9/22 11:26:05	H	162985
374819	20131203021037	* 蕾	学生	20	管理系	207848105	20131203021037	2014/6/23 11:02:29	J	162985
374820	20131204011057	* 悦	学生	38	国际教育学院	207848105	20131204011057	2014/4/4 18:06:20	J	162985
374821	20131205011011	* 丽婷	学生	16	会计系	207848105	20131205011011	2014/4/3 10:11:31	H	162985
374822	20131205011011	* 丽婷	学生	16	会计系	207848105	20131205011011	2014/2/24 11:22:43	J	162985

374822 rows × 17 columns



In [6]:

```
c.head()
```

Out[6]:

	读者条码	姓名	读者级别	单位代码	单位	条形码	读者条码.1	处理时间	操作类型	主键码	题名	出版者
0												
1	20131201021010	* 思其	学生	15	保险系	200541770	20131201021010	2016/12/27 17:25:48	J	9360	国企改革中的难点与对策	新华出版社
2	20131201021010	* 思其	学生	15	保险系	200541770	20131201021010	2017/1/3 14:26:29	H	9360	国企改革中的难点与对策	新华出版社
3	20121205021068	* 欢欢	学生	16	会计系	200592885	20121205021068	2013/3/5 14:14:35	H	8143	企业会计核算精要	中国计量出版社
4	20121205021068	* 欢欢	学生	16	会计系	200592885	20121205021068	2013/1/17 10:53:04	J	8143	企业会计核算精要	中国计量出版社
5	20110912043	* 欣瑜	学生	38	国际教育学院	200346287	20110912043	2013/12/12 11:51:29	J	1344	世界经济政治与国际关系	北京广播学院出版社

In [7]:

```
c.describe(include='all')
```

Out[7]:

	读者条码	姓名	读者级别	单位代码	单位	条形码	读者条码.1	！
count	3.748220e+05	374822	374822	374822.000000	374822	374822.0	3.748220e+05	
unique	NaN	7929	6	NaN	97	76420.0	NaN	
top	NaN	*静	学生	NaN	金融系	200826335.0	NaN	21
freq	NaN	2850	335560	NaN	56487	244.0	NaN	
mean	1.508459e+13	NaN	NaN	26.460773	NaN	NaN	1.508459e+13	
std	8.691563e+12	NaN	NaN	42.015644	NaN	NaN	8.691563e+12	
min	1.000000e+00	NaN	NaN	1.000000	NaN	NaN	1.000000e+00	
25%	2.016126e+12	NaN	NaN	15.000000	NaN	NaN	2.016126e+12	
50%	2.013121e+13	NaN	NaN	17.000000	NaN	NaN	2.013121e+13	
75%	2.014121e+13	NaN	NaN	20.000000	NaN	NaN	2.014121e+13	
max	2.018221e+13	NaN	NaN	543.000000	NaN	NaN	2.018221e+13	

In [8]:

```
c.shape
```

Out[8]:

(374822, 17)

In [9]:

```
c.to_csv('complet.csv')
```

In [10]:

```
c.columns
```

Out[10]:

Index(['读者条码', '姓名', '读者级别', '单位代码', '单位', '条形码', '读者条码.1', '处理时间', '操作类型', '主键码', '题名', '出版者', '责任者', 'Unnamed: 14', 'Unnamed: 15', 'Unnamed: 16', 'Unnamed: 17'], dtype='object')

In [11]:

```
c.duplicated().sum()
```

Out[11]:

32

In [12]:

```
c.iloc[141]
```

Out[12]:

```
读者条码      2009002
姓名          *卫民
读者级别      教师
单位代码      1
单位          河北金融学院
条形码        200420231
读者条码. 1    2009002
处理时间      2016/12/19 08:00:48
操作类型      J
主键码        4470
题名          从计划经济到市场经济
出版者        中国财政经济出版社
责任者        盛洪、张宇燕主编
Unnamed: 14
Unnamed: 15
Unnamed: 16
Unnamed: 17
Name: 142, dtype: object
```

In [13]:

```
del c['Unnamed: 14']
del c['Unnamed: 15']
del c['Unnamed: 16']
del c['Unnamed: 17']
```

In [14]:

```
c
```

Out[14]:

	读者条码	姓名	读者级别	单位代码	单位	条形码	读者条码.1	处理时间	操作类型	主键码
0										
1	20131201021010	* 思其	学生	15	保险系	200541770	20131201021010	2016/12/27 17:25:48	J	9360
2	20131201021010	* 思其	学生	15	保险系	200541770	20131201021010	2017/1/3 14:26:29	H	9360
3	20121205021068	* 欢欢	学生	16	会计系	200592885	20121205021068	2013/3/5 14:14:35	H	8143
4	20121205021068	* 欢欢	学生	16	会计系	200592885	20121205021068	2013/1/17 10:53:04	J	8143
5	20110912043	* 欣瑜	学生	38	国际教育学院	200346287	20110912043	2013/12/12 11:51:29	J	1344
...

	读者条码	姓名	读者级别	单位代码	单位	条形码	读者条码.1	处理时间	操作类型	主键码
0										
374818	20120203022053	* 凯敏	学生	20	管理系	207848105	20120203022053	2014/9/22 11:26:05	H	162985
374819	20131203021037	* 蕾	学生	20	管理系	207848105	20131203021037	2014/6/23 11:02:29	J	162985
374820	20131204011057	* 悦	学生	38	国际教育学院	207848105	20131204011057	2014/4/4 18:06:20	J	162985
374821	20131205011011	* 丽婷	学生	16	会计系	207848105	20131205011011	2014/4/3 10:11:31	H	162985
374822	20131205011011	* 丽婷	学生	16	会计系	207848105	20131205011011	2014/2/24 11:22:43	J	162985

374822 rows × 13 columns

In [15]:

```
c.iloc[141]
```

Out[15]:

读者条码	2009002
姓名	*卫民
读者级别	教师
单位代码	1
单位	河北金融学院
条形码	200420231
读者条码.1	2009002
处理时间	2016/12/19 08:00:48
操作类型	J
主键码	4470
题名	从计划经济到市场经济
出版者	中国财政经济出版社
责任者	盛洪、张宇燕主编

Name: 142, dtype: object

In [16]:

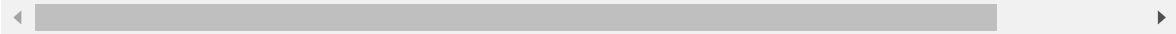
```
c.rename({'读者条码.1': '读者条码'})
```

Out[16]:

	读者条码	姓名	读者级别	单位代码	单位	条形码	读者条码.1	处理时间	操作类型	主键码
0										
1	20131201021010	* 思其	学生	15	保险系	200541770	20131201021010	2016/12/27 17:25:48	J	9360
2	20131201021010	* 思其	学生	15	保险系	200541770	20131201021010	2017/1/3 14:26:29	H	9360
3	20121205021068	* 欢欢	学生	16	会计系	200592885	20121205021068	2013/3/5 14:14:35	H	8143
4	20121205021068	* 欢欢	学生	16	会计系	200592885	20121205021068	2013/1/17 10:53:04	J	8143
5	20110912043	* 欣瑜	学生	38	国际教育学院	200346287	20110912043	2013/12/12 11:51:29	J	1344
...

	读者条码	姓名	读者级别	单位代码	单位	条形码	读者条码.1	处理时间	操作类型	主键码
0										
374818	20120203022053	* 凯敏	学生	20	管理系	207848105	20120203022053	2014/9/22 11:26:05	H	162985
374819	20131203021037	* 蕾	学生	20	管理系	207848105	20131203021037	2014/6/23 11:02:29	J	162985
374820	20131204011057	* 悦	学生	38	国际教育学院	207848105	20131204011057	2014/4/4 18:06:20	J	162985
374821	20131205011011	* 丽婷	学生	16	会计系	207848105	20131205011011	2014/4/3 10:11:31	H	162985
374822	20131205011011	* 丽婷	学生	16	会计系	207848105	20131205011011	2014/2/24 11:22:43	J	162985

374822 rows × 13 columns



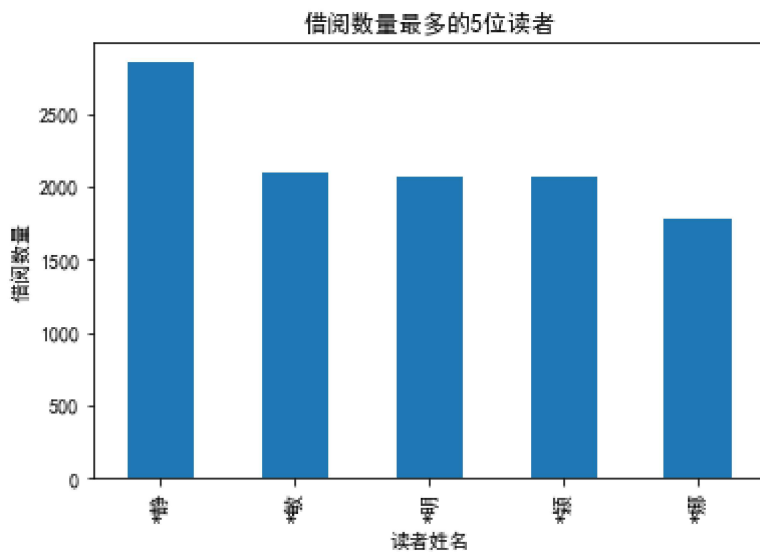
In [17]:

```
mpl.rcParams['font.sans-serif']=['SimHei']  
c['姓名'].value_counts().head(5).plot('bar')  
plt.title('借阅数量最多的5位读者')  
plt.xlabel('读者姓名')  
plt.ylabel('借阅数量')
```

C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:2: FutureWarning:
`Series.plot()` should not be called with positional arguments, only keyword arguments. The order of positional arguments will change in the future. Use `Series.plot(kind='bar')` instead of `Series.plot('bar',)`.

Out[17]:

Text(0, 0.5, '借阅数量')



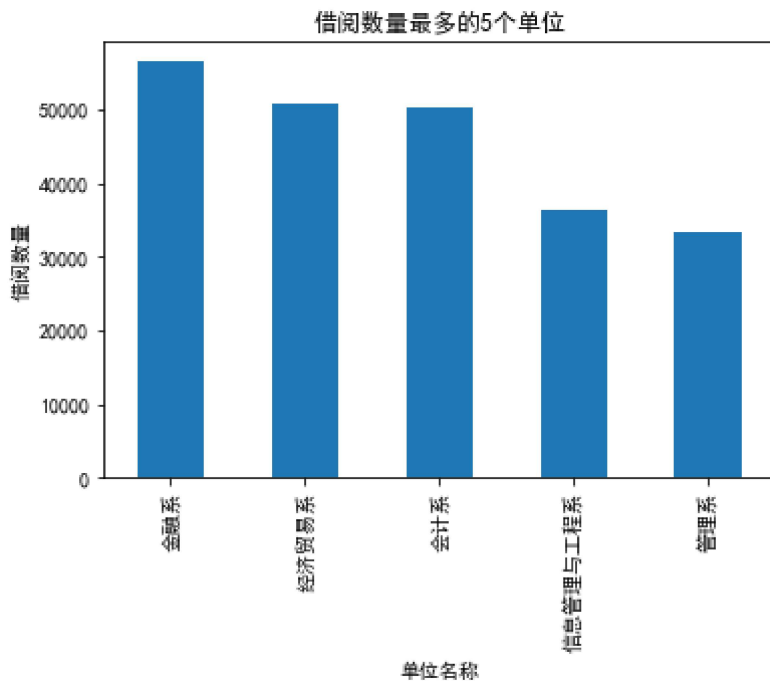
In [18]:

```
mpl.rcParams['font.sans-serif']=['SimHei']  
c['单位'].value_counts().head(5).plot('bar')  
plt.title('借阅数量最多的5个单位')  
plt.xlabel('单位名称')  
plt.ylabel('借阅数量')
```

C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:2: FutureWarning:
`Series.plot()` should not be called with positional arguments, only keyword arguments. The order of positional arguments will change in the future. Use `Series.plot(kind='bar')` instead of `Series.plot('bar',)`.

Out[18]:

Text(0, 0.5, '借阅数量')



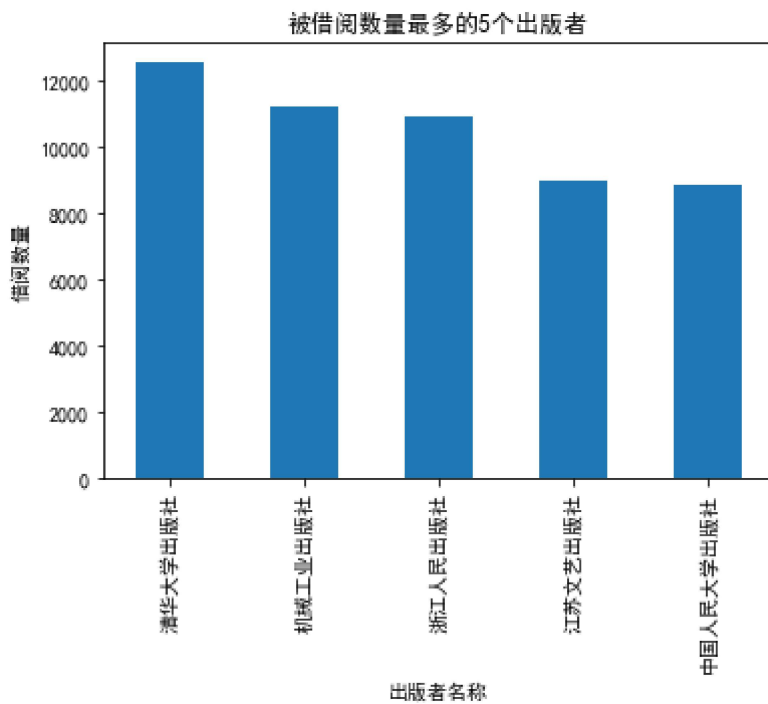
In [19]:

```
mpl.rcParams['font.sans-serif']=['SimHei']
c['出版者'].value_counts().head(5).plot('bar')
plt.title('被借阅数量最多的5个出版者')
plt.xlabel('出版者名称')
plt.ylabel('借阅数量')
```

C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:2: FutureWarning: `Series.plot()` should not be called with positional arguments, only keyword arguments. The order of positional arguments will change in the future. Use `Series.plot(kind='bar')` instead of `Series.plot('bar',)`.

Out[19]:

Text(0, 0.5, '借阅数量')



In []:

In []:

In []:

In []:

In []:

In [20]:

```
c['读者级别'].value_counts()
```

Out[20]:

```
学生      335560
教师      33366
研究生    5851
特殊人员    28
校外人员    14
大额借书     3
Name: 读者级别, dtype: int64
```

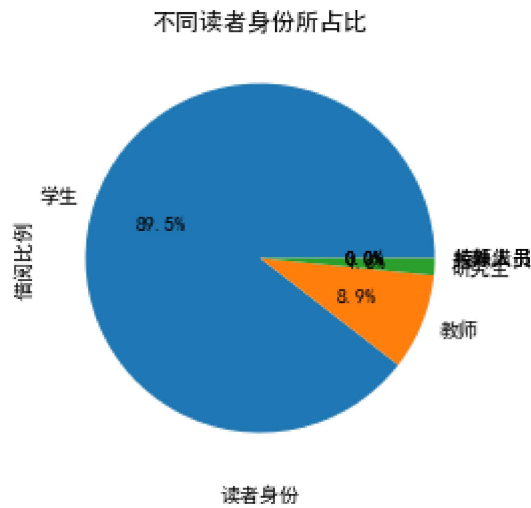
In [21]:

```
mpl.rcParams['font.sans-serif']=['SimHei']
c['读者级别'].value_counts().plot('pie', autopct='%1.1f%%')
plt.title('不同读者身份所占比')
plt.xlabel('读者身份')
plt.ylabel('借阅比例')
```

C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:2: FutureWarning: `Series.plot()` should not be called with positional arguments, only keyword arguments. The order of positional arguments will change in the future. Use `Series.plot(kind='pie')` instead of `Series.plot('pie',)`.

Out[21]:

Text(0, 0.5, '借阅比例')



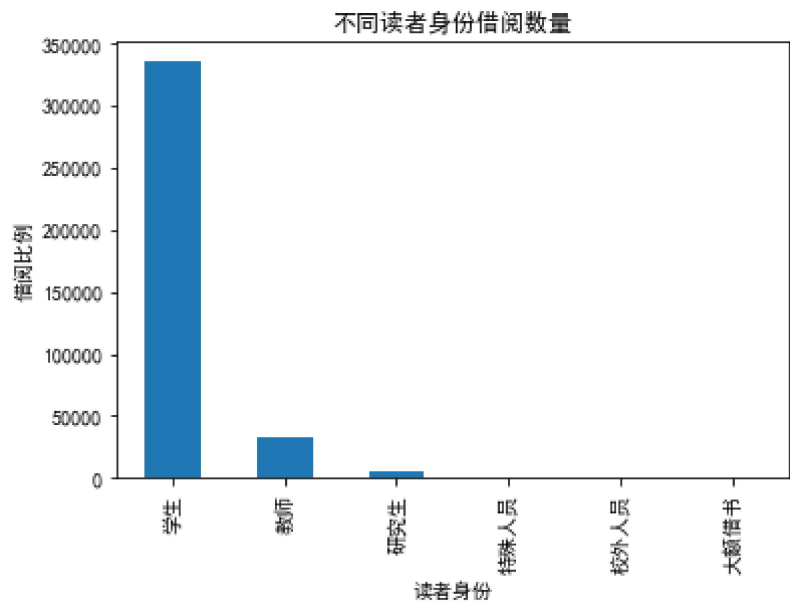
In [22]:

```
mpl.rcParams['font.sans-serif']=['SimHei']
c['读者级别'].value_counts().plot('bar')
plt.title('不同读者身份借阅数量')
plt.xlabel('读者身份')
plt.ylabel('借阅比例')
```

C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:2: FutureWarning: `Series.plot()` should not be called with positional arguments, only keyword arguments. The order of positional arguments will change in the future. Use `Series.plot(kind='bar')` instead of `Series.plot('bar',)`.

Out[22]:

Text(0, 0.5, '借阅比例')



In []:

In []: