

```

1 import pandas as pd
2 from matplotlib import pyplot as plt

```

(1) 计算评论者中男性和女性的人数并绘制如下所示饼图

```

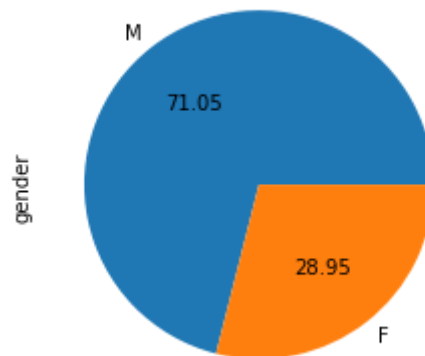
1 columns = ['user id', 'age', 'gender', 'occupation', 'zip code']
2 user = pd.read_csv('./ml-100k/u.user', delimiter='|', header=None,
3 names=columns)
4 user.gender.value_counts().plot(kind='pie', autopct='%.2f')

```

```

1 <matplotlib.axes._subplots.AxesSubplot at 0x189a83d32e8>

```



(2) 计算数据中不同种类电影的评论数并绘制柱状图

```

1 columns = ['unknown', 'Action', 'Adventure', 'Animation', "Children's",
2 'Comedy',
3 'Crime', 'Documentary', 'Drama', 'Fantasy', 'Film-Noir', 'Horror',
4 'Musical', 'Mystery', 'Romance', 'Sci-Fi', 'Thriller', 'War',
5 'Western']
6 item = pd.read_csv('./ml-100k/u.item', delimiter='|', header=None,
7 names=columns,
8 usecols=range(5, 24))

```

```

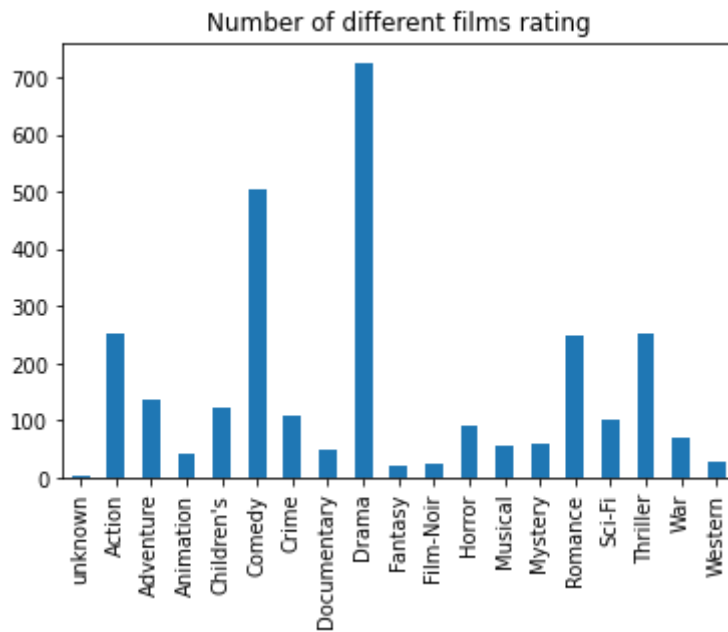
1 item.sum().plot(kind='bar')
2 plt.title('Number of different films rating')

```

```

1 Text(0.5, 1.0, 'Number of different films rating')

```



(3) 计算数据中评论最热门（评价数最多）的前10部电影的均分，绘制如图所示柱状图

```

1 item_names = ['item_id', 'title']
2 data_names = ['item_id', 'rating']
3 item = pd.read_csv('./ml-100k/u.item', delimiter='|', header=None,
4 names=item_names, usecols=(0, 1))
5 data = pd.read_csv('./ml-100k/u.data', delimiter='\t', header=None,
6 names=data_names, usecols=(1, 2))

```

```

1 table = pd.merge(item, data)
2 table = table.groupby('title').rating.agg(['count', 'mean'])
3 table.sort_values('count', ascending=False).iloc[:10, -1].plot(kind='bar')

```

```

1 <matplotlib.axes._subplots.AxesSubplot at 0x189abb3ada0>

```

