Python数据分析实战

第十二课 pandas读取数据

本节课程目标

- pandas概述
- pandas读取数据

pandas概述

- 什么是pandas?
- 如何安装pandas环境?

```
# pip install pandas
```

pandas数据读取

- CSV
- TXT
- Excel

pandas读取纯文本文件

CSV

```
import pandas as pd
```

```
#读取数据
filePath = './datas/ratings.csv'
ratings = pd.read_csv(filePath)
```

```
#获取数据后可以进行操作了
```

```
#查看前几行数据
ratings.head()
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	userId	movieId	rating	timestamp
0	1	1	4.0	964982703
1	1	3	4.0	964981247
2	1	6	4.0	964982224
3	1	47	5.0	964983815
4	1	50	5.0	964982931

```
#查看数据的形状、返回（行数、列数）
ratings.shape
```

```
(100836, 4)
```

```
#查看列名列表
ratings.columns
```

```
Index(['userId', 'movieId', 'rating', 'timestamp'], dtype='object')
```

```
#查看索引列
ratings.index
```

```
RangeIndex(start=0, stop=100836, step=1)
```

```
#查看每列的数据类型
ratings.dtypes
```

```
userId          int64
movieId         int64
rating        float64
timestamp       int64
dtype: object
```

TXT

```
filePath = './datas/access_pvuv.txt'
pvuv = pd.read_csv(filePath,sep='\t',header=None,names=['pdata','pv','uv'])
```

```
pvuv
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	pdata	pv	uv
0	2019-09-10	139	92
1	2019-09-09	185	153
2	2019-09-08	123	59
3	2019-09-07	65	40
4	2019-09-06	157	98
5	2019-09-05	205	151
6	2019-09-04	196	167
7	2019-09-03	216	176
8	2019-09-02	227	148
9	2019-09-01	105	61

```
pvuv.head()
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	pdata	pv	uv
0	2019-09-10	139	92
1	2019-09-09	185	153
2	2019-09-08	123	59
3	2019-09-07	65	40
4	2019-09-06	157	98

```
pvuv.shape
```

```
(10, 3)
```

```
pvuv.columns
```

```
Index(['pdata', 'pv', 'uv'], dtype='object')
```

Excel

```
filePath = './datas/access_pvuv.xlsx'
pvuv = pd.read_excel(filePath)
```

```
pvuv
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	日期	PV	UV
0	2019-09-10	139	92
1	2019-09-09	185	153
2	2019-09-08	123	59
3	2019-09-07	65	40
4	2019-09-06	157	98
5	2019-09-05	205	151
6	2019-09-04	196	167
7	2019-09-03	216	176
8	2019-09-02	227	148
9	2019-09-01	105	61

```
#mysql   read_sql()
```

```
pd.read
```

```

```