

Excel表格数据处理

excel表格数据处理

- ◆ 加载表格
- ◆ 基础数据分析
- ◆ 表格数据保存
- ◆ 综合案例

加载表格

◆ 安装pandas

```
pip install pandas
```

加载表格

◆ 读取csv表格

```
1 import pandas as pd
2
3 data_loc = r'./resources/yolov5s.csv'
4 data = pd.read_csv(data_loc)
5 print(data)
```

	epoch	train/box_loss	train/obj_loss	train/cls_loss	metrics/precision	metrics/recall	metrics/mAP_0.5	metrics/mAP_0.5:0.95	val/box_loss	val/obj_loss	v
0	0	0.103060	0.037026	0.010070	0.14502	0.18253	0.065063	0.012609	0.084471	0.024975	
1	1	0.076159	0.039491	0.005065	0.42583	0.34571	0.302180	0.097310	0.065515	0.022369	
2	2	0.066722	0.033346	0.001560	0.30246	0.35623	0.258250	0.070605	0.060241	0.021680	
3	3	0.002203	0.033572	0.001129	0.43289	0.36005	0.381110	0.140980	0.050548	0.019933	
4	4	0.057136	0.030049	0.000997	0.54404	0.50147	0.503630	0.204920	0.053203	0.018728	
...
95	95	0.026012	0.018415	0.000055	0.71293	0.59470	0.617860	0.307080	0.046327	0.020765	
96	96	0.026545	0.018615	0.000042	0.71060	0.59372	0.611530	0.302590	0.046212	0.020430	
97	97	0.026161	0.020037	0.000041	0.69721	0.60255	0.628320	0.312580	0.046403	0.020721	
98	98	0.026270	0.023377	0.000041	0.68128	0.61335	0.615640	0.310900	0.045835	0.020823	
99	99	0.025782	0.018342	0.000040	0.71131	0.60255	0.619860	0.303100	0.046320	0.020630	

100 rows × 14 columns

加载表格

◆ 读取excel表格

```
1 import pandas as pd
2
3 data_loc = r'./resources/销售数据.xls'
4 data = pd.read_excel(data_loc)
5 print(data)
```

顾客编号	大类编码	大类名称	中类编码	中类名称	小类编码	小类名称	销售日期	销售月份	商品编码	规格型号	商品类型	单位	销售数量	销售金额	商品单价	是否促销
0	0.0	12.0	蔬菜	1201.0	蔬菜	120100.0	20150101.0	201501.0	DW-1201000311		生鲜	个	8.000	4.00	2.00	否
1	1.0	20.0	粮油	2014.0	蔬菜类	201401.0	20150101.0	201501.0	DW-2014010019	90g	一般商品	袋	6.000	3.00	0.50	否
2	2.0	15.0	日配	1505.0	冷藏食品	150502.0	20150101.0	201501.0	DW-1505020011	150g	一般商品	袋	1.000	2.40	2.40	否
3	3.0	15.0	日配	1503.0	冷藏食品	150305.0	20150101.0	201501.0	DW-1503050035	500g	一般商品	袋	1.000	6.50	6.50	否
4	4.0	15.0	日配	1505.0	冷藏食品	150502.0	20150101.0	201501.0	DW-1505020020	100g*8	一般商品	袋	1.000	11.90	11.90	否
...
42811	1605.0	12.0	蔬菜	1201.0	蔬菜	120106.0	20150430.0	201504.0	DW-1201060002	散称	生鲜	千克	0.217	3.91	18.00	否
42812	1572.0	12.0	蔬菜	1201.0	蔬菜	120102.0	20150430.0	201504.0	DW-1201020040	散称	生鲜	千克	0.440	0.86	1.96	否
42813	1170.0	30.0	冻化	3016.0	纸制品	301603.0	20150430.0	201504.0	DW-3016030007	10卷	一般商品	提	1.000	12.90	14.50	是
42814	2605.0	12.0	蔬菜	1201.0	蔬菜	120101.0	20150430.0	201504.0	DW-1201010023	散称	生鲜	千克	0.718	1.84	2.56	否
42815	2610.0	23.0	冻饮	2317.0	进口饮料	231702.0	20150430.0	201504.0	DW-2317020131	490ml	联营商品	罐	1.000	8.00	8.00	否

42816 rows x 17 columns

基础数据分析

◆ 数据分组与排序

```
1 import pandas as pd
2
3 data_loc = r'./resources/销售数据.xlsx'
4 data = pd.read_excel(data_loc)
5
6 print(data.describe())
```

	顾客编号	大类编码	中类编码	小类编码	销售日期	销售月份	销售数量	销售金额	商品单价
count	42814.000000	42814.000000	42814.000000	42814.000000	4.281400e+04	42814.000000	42814.000000	42814.000000	42814.000000
mean	876.917667	17.934741	1799.498178	179958.383613	2.015026e+07	201502.443383	1.199202	10.608968	12.541444
std	895.070907	6.061268	608.945570	60897.176238	1.151576e+02	1.149798	2.519062	41.827726	16.658760
min	0.000000	10.000000	1001.000000	100101.000000	2.015010e+07	201501.000000	-16.000000	-145.000000	0.000000
25%	279.000000	12.000000	1201.000000	120106.000000	2.015013e+07	201501.000000	0.530000	2.900000	3.960000
50%	729.000000	20.000000	2001.000000	200103.000000	2.015022e+07	201502.000000	1.000000	5.700000	6.900000
75%	1350.000000	22.000000	2203.000000	220306.000000	2.015040e+07	201504.000000	1.000000	10.960000	14.997500
max	2611.000000	34.000000	3436.000000	343699.000000	2.015043e+07	201504.000000	216.000000	5340.000000	880.000000

基础数据分析

◆ 输出头几行

```
1 import pandas as pd
2
3 data_loc = r'./resources/销售数据.xlsx'
4 data = pd.read_excel(data_loc)
5
6 data.head()
```

	顾客编号	大类编码	大类名称	中类编码	中类名称	小类编码	小类名称	销售日期	销售月份	商品编码	规格型号	商品类型	单位	销售数量	销售金额	商品单价	是否促销
0	0.0	12.0	蔬果	1201.0	蔬菜	120109.0	其它蔬菜	20150101.0	201501.0	DW-1201090311		生鲜	个	8.0	4.0	2.0	否
1	1.0	20.0	粮油	2014.0	酱菜类	201401.0	榨菜	20150101.0	201501.0	DW-2014010019	60g	一般商品	袋	6.0	3.0	0.5	否
2	2.0	15.0	日配	1505.0	冷藏乳品	150502.0	冷藏加味酸乳	20150101.0	201501.0	DW-1505020011	150g	一般商品	袋	1.0	2.4	2.4	否
3	3.0	15.0	日配	1503.0	冷藏料理	150305.0	冷藏面食类	20150101.0	201501.0	DW-1503050035	500g	一般商品	袋	1.0	6.5	8.3	否
4	4.0	15.0	日配	1505.0	冷藏乳品	150502.0	冷藏加味酸乳	20150101.0	201501.0	DW-1505020020	100g*8	一般商品	袋	1.0	11.9	11.9	否

基础数据分析

◆ 输出尾几行

```
1 import pandas as pd
2
3 data_loc = r'./resources/销售数据.xlsx'
4 data = pd.read_excel(data_loc)
5
6 data.tail() # 默认输出尾5行
```

	顾客编号	大类编码	大类名称	中类编码	中类名称	小类编码	小类名称	销售日期	销售月份	商品编码	规格型号	商品类型	单位	销售数量	销售金额	商品单价	是否促销
42811	1605.0	12.0	蔬果	1201.0	蔬菜	120106.0	菌菇类	20150430.0	201504.0	DW-1201060002	散称	生鲜	千克	0.217	3.91	18.00	否
42812	1572.0	12.0	蔬果	1201.0	蔬菜	120102.0	根茎	20150430.0	201504.0	DW-1201020040	散称	生鲜	千克	0.440	0.86	1.96	否
42813	1170.0	30.0	洗化	3016.0	纸制品	301603.0	无芯纸	20150430.0	201504.0	DW-3016030007	10卷	一般商品	提	1.000	12.90	14.50	是
42814	2605.0	12.0	蔬果	1201.0	蔬菜	120101.0	叶菜	20150430.0	201504.0	DW-1201010023	散称	生鲜	千克	0.718	1.84	2.56	否
42815	2610.0	23.0	酒饮	2317.0	进口饮料	231702.0	进口果汁	20150430.0	201504.0	DW-2317020131	490ml	联营商品	罐	1.000	8.00	8.00	否

基础数据分析

◆ 定位第一行

```
1 import pandas as pd
2
3 data_loc = r'./resources/销售数据.xlsx'
4 data = pd.read_excel(data_loc)
5
6 # 读取第一行的值
7 # 索引第一行的值，行标签是“0”
8 data_0 = data.loc[0]
9 data_0
```

```
顾客编号      0.0
大类编码      12.0
大类名称      蔬果
中类编码      1201.0
中类名称      蔬菜
小类编码      120109.0
小类名称      其它蔬菜
销售日期      20150101.0
销售月份      201501.0
商品编码      DW-1201090311
规格型号
商品类型      生鲜
单位          个
销售数量      8.0
销售金额      4.0
商品单价      2.0
是否促销      否
Name: 0, dtype: object
```

基础数据分析

◆ 定位指定列

```
1 # 读取“大类编码”列全部值
2 data_1 = data.loc[:, "大类编码"]
3 data_1
```

```
0      12.0
1      20.0
2      15.0
3      15.0
4      15.0
...
42811   12.0
42812   12.0
42813   30.0
42814   12.0
42815   23.0
Name: 大类编码, Length: 42816, dtype: float64
```

基础数据分析

◆ 通过指定行和列，定位指定元素

```
1 # 读取第1行， "大类编码"列对应的值
2 data_2 = data.loc[ 1, "大类编码"]
3 data_2
```

1 | 20

基础数据分析

◆ 对行和列使用切片

```
1 # 读取某个区域
2 # 读取第1行到第3行， "大类编码"列到"中类名称"列这个区域内的值
3 data_3 = data.loc[ 1:3, "大类编码":"中类名称"]
4 data_3
```

	大类编码	大类名称	中类编码	中类名称
1	20.0	粮油	2014.0	酱菜类
2	15.0	日配	1505.0	冷藏乳品
3	15.0	日配	1503.0	冷藏料理

基础数据分析

◆ 提取符合要求的元素

```
1 # 根据条件进行提取
2 # 读取"销售数量"列中大于10的值
3 data_4 = data.loc[ data["销售数量"] > 10]
4 data_4
```

销售编号	大类编码	大类名称	中类编码	中类名称	小类编码	小类名称	销售日期	销售月份	商品编码	规格型号	商品名称	单位	销售数量	销售金额	商品单价	是否促销
22	10.0	15.0	1518.0	常温乳品	151805.0	利乐枕纯奶	20150101.0	201501.0	DW-1518050007	240ml	一般商品	袋	10.0	33.9	2.7	是
51	33.0	15.0	1518.0	常温乳品	151805.0	利乐枕纯奶	20150101.0	201501.0	DW-1518050001	240ml	一般商品	袋	18.0	43.2	2.7	否
437	118.0	15.0	1518.0	常温乳品	151813.0	利乐砖乳酸菌	20150102.0	201501.0	DW-1518130016	205g	一般商品	盒	12.0	48.0	5.5	是
478	60.0	15.0	1518.0	常温乳品	151807.0	利乐枕营养奶	20150102.0	201501.0	DW-1518070001	240ml	一般商品	袋	10.0	44.8	2.8	否
488	80.0	20.0	2015.0	罐装	201501.0	八宝粥	20150102.0	201501.0	DW-2015010036	360g	一般商品	个	12.0	38.4	3.2	否
...
42444	2601.0	15.0	1518.0	常温乳品	151806.0	利乐枕加味奶	20150430.0	201504.0	DW-1518060018	230ml	一般商品	袋	10.0	29.9	2.8	是
42500	1073.0	15.0	1518.0	常温乳品	151813.0	利乐砖乳酸菌	20150430.0	201504.0	DW-1518130011	200ml	一般商品	盒	12.0	58.9	5.3	是
42509	1631.0	15.0	1521.0	蛋类	152101.0	新鲜蛋品	20150430.0	201504.0	DW-1521010022	散卵	一般商品	kg	10.1	141.4	14.0	否
42746	421.0	15.0	1505.0	冷藏乳品	150502.0	冷藏加味酸乳	20150430.0	201504.0	DW-1505020011	150g	一般商品	袋	15.0	26.9	2.4	是
42757	74.0	15.0	1518.0	常温乳品	151803.0	利乐砖营养奶	20150430.0	201504.0	DW-1518030008	250ml	一般商品	盒	24.0	79.2	3.3	否

682 rows × 17 columns

基础数据分析

◆ 提取符合要求的元素

```
1 # 切片操作
2 data_5 = data.loc[ data["销售数量"] > 10, ["小类编码","小类名称"]]
3 data_5
```

	小类编码	小类名称
22	151805.0	利乐枕纯奶
51	151805.0	利乐枕纯奶
437	151813.0	利乐砖乳酸菌
478	151807.0	利乐枕营养奶
488	201501.0	八宝粥
...
42444	151806.0	利乐枕加味奶
42500	151813.0	利乐砖乳酸菌
42509	152101.0	新鲜蛋品
42746	150502.0	冷藏加味酸乳
42757	151803.0	利乐砖营养奶

682 rows × 2 columns

基础数据分析

◆ 数据分组与排序

```
1 import pandas as pd
2
3 file_loc = 'resources/销售数据.xlsx'
4
5 data = pd.read_excel(file_loc)
6
7 data_extract = data.groupby('商品类型')['销售金额'].sum()
8 data_extract = data_extract.reset_index()
9
10 print(data_extract)
```

	商品类型	销售金额
0	一般商品	336529.74
1	生鲜	111808.70
2	联营商品	5873.81

表格数据保存

◆ 将处理好的数据保存成excel

```
1 import pandas as pd
2
3 file_loc = 'resources/销售数据.xlsx'
4
5 data = pd.read_excel(file_loc)
6
7 data_extract = data.groupby('商品类型')['销售金额'].sum()
8 data_extract = data_extract.reset_index()
9
10 data_extract.to_csv("处理好的表格.csv", encoding='gbk', index=False)
11 data_extract.to_excel("处理好的表格.xlsx", encoding='gbk', index=False)
```


综合案例

◆ 实操演示



慕课网
imooc.com



慕课网
imooc.com



慕课网
imooc.com



慕课网
imooc.com