

```
1 import pandas as pd
2 from openpyxl import load_workbook
3
4 # 处理日期
5 # 打开文件
6 workbook = load_workbook(filename='日化.xlsx', data_only=True)
7 # 选择工作表
8 ws = workbook['销售订单表']
9 nrows = ws.max_row
10 # 遍历第二列单元格
11 for i in range(2, nrows + 1):
12     # 处理单元格数据
13     if isinstance(ws.cell(i, 2).value, str):
14         print(ws.cell(i, 2).value)
15         ws.cell(i, 2).value = ws.cell(i, 2).value.replace("#", "/")
16
17 # 保存文件
18 workbook.save(filename='日化_1.xlsx')
19 # 关闭文件
20 workbook.close()
21
22 # 读取"销售订单表"到DataFrame
23 df_dd = pd.read_excel('日化_1.xlsx', sheet_name='销售订单表')
24 print(df_dd.head(20))
25
26 # 读取"商品信息表"到DataFrame
27 df_xx = pd.read_excel('日化_1.xlsx', sheet_name='商品信息表')
28 print(df_xx.head(20))
29
30 # 处理缺失值
31 print(df_dd)
32 df_dd.dropna(inplace=True)
33 print(df_dd)
34
35 # 根据商品编号, 增加商品小类列
36 dict_xx = dict()
37 for index, row in df_xx.iterrows():
38     print(row['商品编号'], row['商品小类'])
39     dict_xx[row['商品编号']] = row['商品小类']
40 print(dict_xx)
41 df_dd['商品小类'] = df_dd['商品编号'].replace(dict_xx)
42 print(df_dd.head(20))
43
44 # 排序, 按商品小类排序也可以
```

```
45 df_dd.sort_values(by=['所在地市', '商品编号'], inplace=True)
46 pd.set_option('display.max_columns', None)
47 pd.set_option('display.max_rows', None)
48 print(df_dd.head(200))
49
50 # 查看分组后的信息，此操作是为了便于同学们理解，
   可以直接遍历df_dd输出到word
51 print(df_dd.groupby(['所在地市', '商品编号']).sum())
52 print((df_dd.groupby(['所在地市', '商品编号']).sum()).describe())
53 # 保存到文件，此操作是为了便于同学们理解，
   可以直接遍历df_dd输出到word
54 df_dd.to_excel('日化_2.xlsx', index=False)
55
```