

※

[python之pandas](https://www.pypandas.cn/) Pandas统计学习积累 <https://www.pypandas.cn/>

pandas 在数据分析上很强大，主要数据结构是 Series（一维数据）与 DataFrame（二维数据），他是基于numpy开发的。

pandas 读写csv

1. 安装pandas: `conda install pandas`
2. 引入: `import pandas as pd`
3. 读取csv: `df = pd.read_csv('./gaode2lvtu.csv')`
4. 写入csv: `df.to_csv('out.csv', encoding = "utf-8-sig", index=False)`

pandas 统计

1. DataFrame: `index(行)`, `columns(列)`
2. Serries和DataFrame创建:
`s = pd.Series([1, 3, 5, np.nan, 6, 8])`
`dates = pd.date_range('20130101', periods=6)`
`df = pd.DataFrame(np.random.randn(6, 4), index=dates, columns=list('ABCD'))`
3. 查询所有列的内容: `df.dtypes`
4. 查看数据统计摘要: `df.describe()`
5. 排序
按轴排序: `df.sort_index(axis=1, ascending=False)`
按值排序: `df.sort_values(by='B')`
6. 合并:
`df = pd.concat([df, df_file])`
7. 一列切分几列:
`dfselect['typecode'] = dfselect['typecode'].astype('str')`
`data_find =`
`dfselect.loc[dfselect['typecode'].str.contains('|')]`
`if len(data_find)>0:`

```

gjj_pro = dfselect['typecode'].str.split('|', expand=True)
# 将数据按'|'拆分
typecount = gjj_pro.shape[1]
new_names = ['typecode' + str(x + 1) for x in
range(typecount)] # 为新生成的列取名
gjj_pro.columns = new_names # 重命名新生成的列名
dfselect['typecode'] = gjj_pro['typecode1']
dfselect['typecode'] = dfselect['typecode'].astype('int')
8. 两表合并:
dfselect = pd.merge(dfselect,gaode2CTA,how='left', on=
['typecode'])
9. 删除指定列:
dfselect = dfselect.drop('typecode', axis=1)
10. 重命名列名:
dfselect.rename(columns={'name':'名称','lng':'经度','lat':'纬
度','lvtu-name':'类型','tel':'电话','address':'地址','adname':'区
县','cityname':'城市'},inplace=True)
11. 分组统计:
f_typecode_file =
df_file.groupby(['cityname','adname','typecode'])['name'].count()
df_typecode_file = pd.DataFrame(df_typecode_file)
df_typecode_file.reset_index(inplace=True)
12. 列类型转换, 列计算:
df_class1['typecode'] = df_class1['typecode'].astype('int')/100
df_class1['typecode'] = df_class1['typecode'].astype('int')
13. 扩展一行数据:
df_ad_Area = df_ad_Area.append(a,ignore_index=True)
14. 空白填充0:
df_pois = df_pois.fillna(0)

```

pandas

- 1.
- 2.

3.