

## 数据分析项目

依据提供的 65536 条记录的超市销售数据，根据自己的需求进行分析，最终出具报告，报告包括数据的分析，并将数据可视化，要求最终提交学号姓名.ipynb 和学号姓名.pdf。

过程要求：

- 1、新建 ipynb 文档，使用 `df=pd.read_csv(open('超市销售数据.csv'))` 读入数据
- 2、用 markdown 语句在 ipynb 中分块制定分析计划。

分析目标：

大框架：

- 数据读取 参考 `pd.read_csv(open('超市销售数据.csv'))`
- 数据整理（如果觉得数据量太大了，可以提取局部数据）参考：`df.dropna()`
- 数据分析 包括（分组统计、常用函数应用、排序）
- 数据可视化 用 `matplotlib` 绘图，有标题、坐标轴、坐标轴刻度和图例等设置。

- 3、在讲义中补充学习经验。

时间:下周一下课时交。

提示：时间有限，目标不要太远大。

补充知识：

`df.to_excel('文件名.xlsx',sheet_name='sheet1',index=False)`

`del df['售价']` 可以永久删除 `df` 对象中的“售价”列数据。

`df.shape[0]` 可以读取多少行 `df.shape[1]` 可以读取多少列

`plot()` 线形图 `scatter()` 散点图

`Series` 是一种一维的数据结构，包含一个数组的数据和一个与数据关联的索引（`index`）。

列表和字典可以用来创建，索引可以指定，可以是字符串类型。有 `index` 、`values` 两个属性。

`s2=pd.Series([166,178,180],index=["s01","s02","s03"])`

补充说明：

1、可以缩小数据规模。

2、周一课还有做的时间，如果来不及可以下周第二节课占用半节。

3、如果完全不知道怎么做，也可以已经完成的第一个图书销售管理任务再独立做一次。

4、如果出现 UTF-8 的编码问题，可以用记事本打开.csv 文件，选择另存为，在窗体的编码项中，由 ANSI 改为 UTF-8，如果直接可以打开就不必做此处理。不想通过文件编码修改的也可以使用 `encoding` 的参数设置，可以自行百度。