目录

**[一、背景介绍 2](#_Toc3564)**

[1、 项目背景 2](#_Toc25520)

[2、 分析目标 2](#_Toc31063)

**[二、 项目任务 3](#_Toc9914)**

[1、任务一 数据预处理与表关联 3](#_Toc31228)

[2、任务二 对食堂就餐行为进行分析 4](#_Toc15012)

[3、任务三 对学生消费行为进行分析 11](#_Toc9043)

# 一、背景介绍

## 1、 项目背景

随着中国国内物质文化水平的提升，中国大学生的生活质量不断提高，越来越多人关注大学生的校园生活，把目光投向了校园环境。良好的校园环境可以给学生提供众多的便利条件，让学生更愉悦、更努力地学习。因此，学校也非常重视校园环境的质量，其中就包括了校园内的学生消费场所，例如饭堂、商店等。本项目应运而生，是某学校想要通过分析学生的校园消费行为，进而改进学校的服务，并为其相关部门的决策提供信息支持。

本项目通过分析该校学生消费的具体项目和地点，了解并掌握学生的消费水平、消费地点和消费时间，定位不同阶层的学生及其消费特征，分析其消费的未来走向，挖掘学生的深层次需求，给校方提供清晰准确的数据和有针对性的建议，提高学校的服务。

## 2、 分析目标

1.学生早中晚餐的就餐地点存在的显著区别。

2.食堂早中晚餐的就餐峰值。

3.不同专业间不同性别学生群体的消费特点。

4.每一类学生的消费特点，为学校判定学生的经济情况提供参考意见。

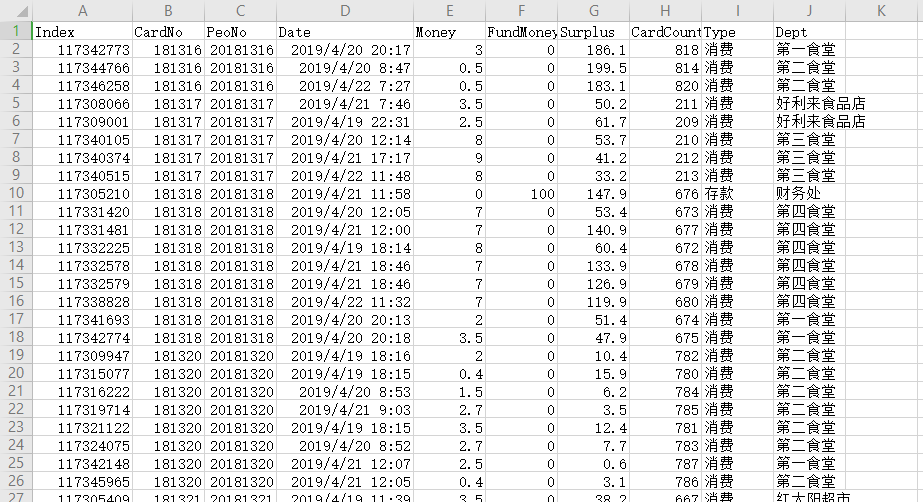
# 项目任务

### 1、任务一 数据预处理与表关联

#### 任务1.1 根据实际项目需求对数据进行必要的数据探索和数据预处理。

对data1.csv文件中数据的异常值进行处理，修改为正常值；删除data2.csv文件中数据的TermSerNo和conOperNo两列无意义数据；对data1.csv和data2.csv的数据进行描述性统计和缺失值统计；删除data2.csv中凌晨0点到早上6点的无效消费数据。

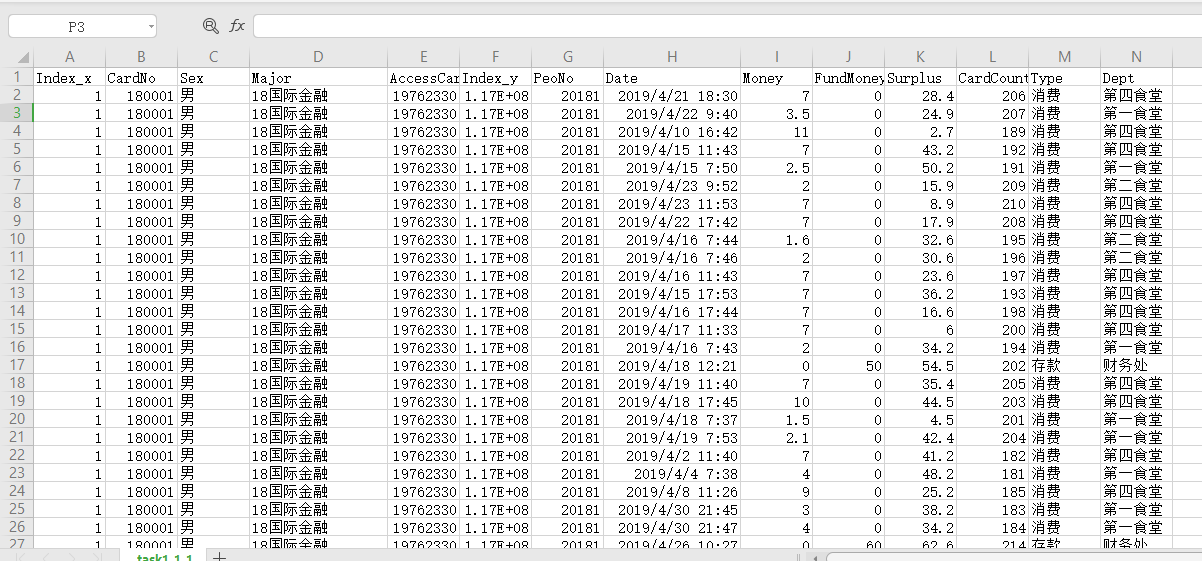
下图为处理后的部分数据的截图：



#### 任务1.2 对学生个人信息表和消费记录表进行关联

使用CardNo字段作为表关联的依据。

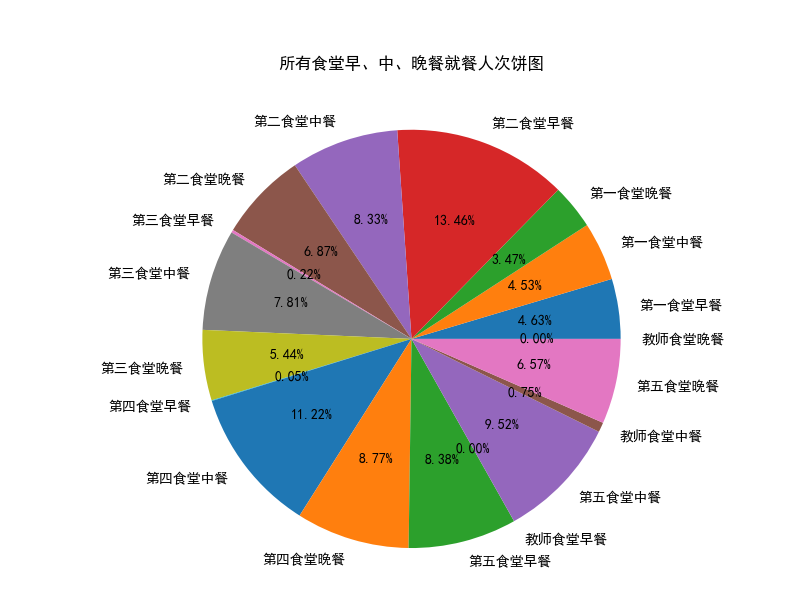
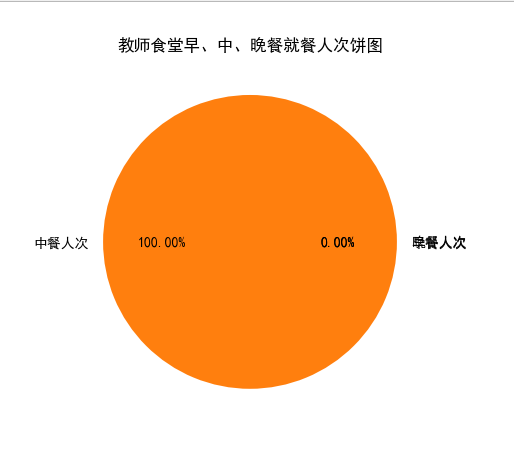
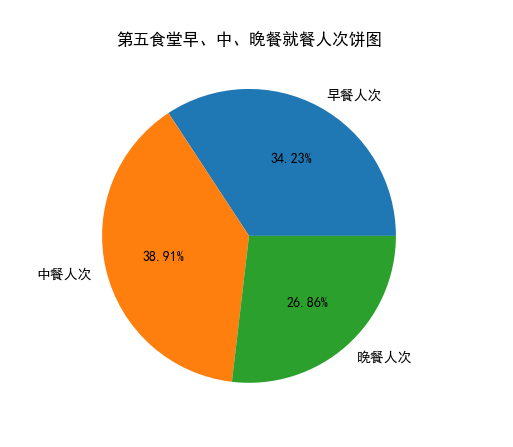
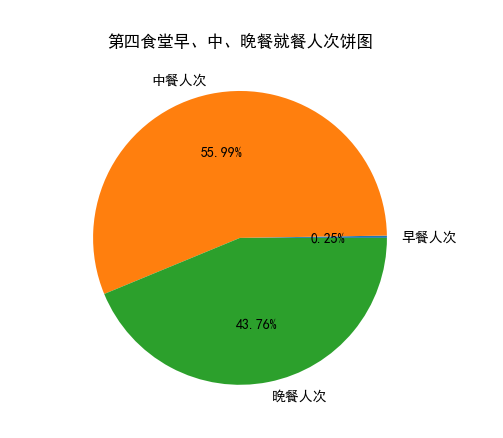
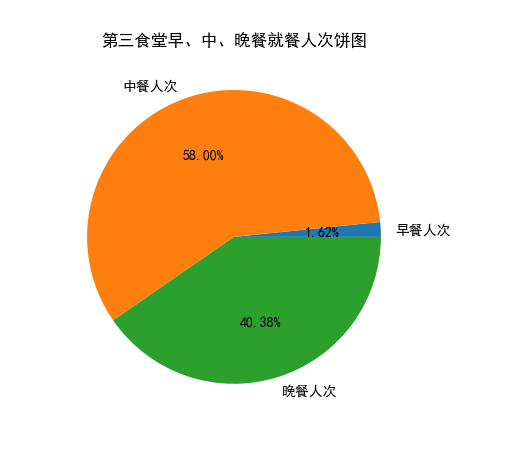
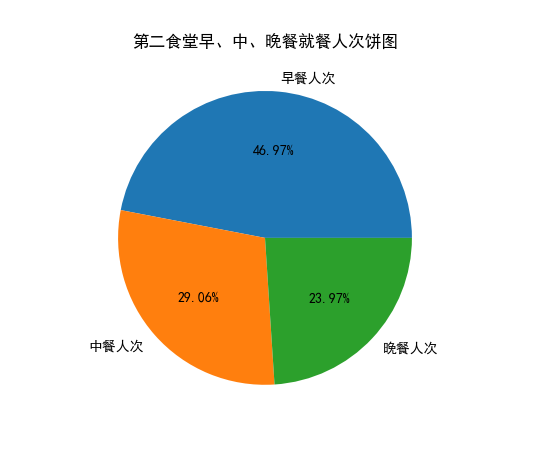
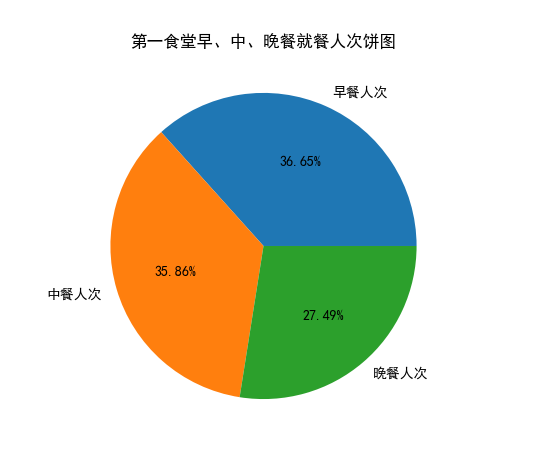
下图为处理后的部分数据的截图：



### 2、任务二 对食堂就餐行为进行分析

任务2.1 绘制各食堂就餐人次的占比饼图，分析学生早中晚餐的就餐地点是否有显著差别。其中，时间间隔非常接近的多次刷卡记录可能为一次就餐行为。

早餐时间定为6点-10点，中餐定为11点-14点，晚餐定为17点-21点。分别提取数据表中不同食堂的早中晚餐学生刷卡的次数，去除一次就餐行为中的多次刷卡记录，根据数据绘制饼图。

如下所示为各食堂就餐人次的占比饼图：

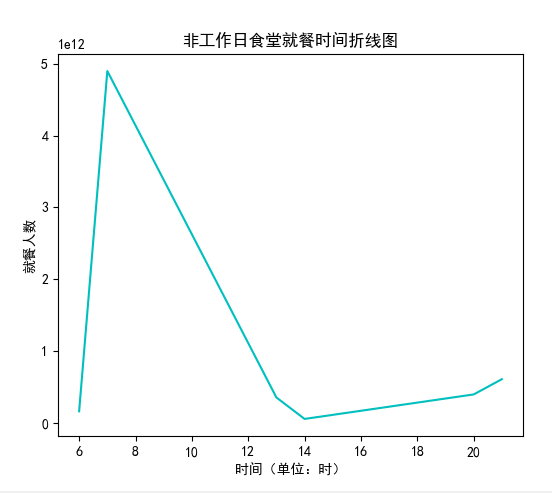
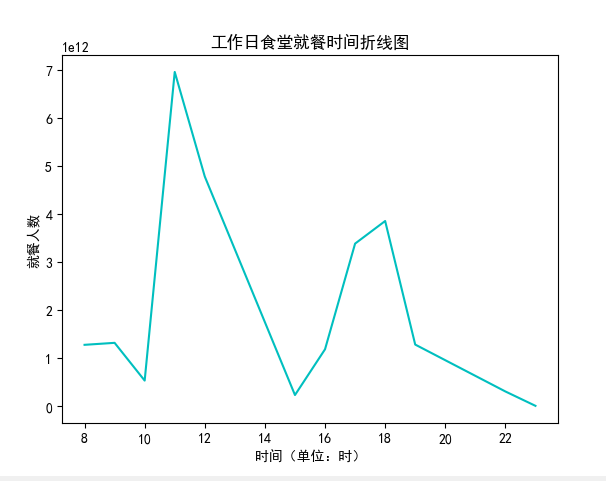
根据所有食堂早、中、晚餐就餐人次饼图可以得出结论：

学生早餐就餐地点人数排名：第二食堂>第五食堂>第一食堂>第三食堂>第四食堂>教师食堂。

学生中餐就餐地点人数排名：第四食堂>第五食堂>第二食堂>第三食堂>第一食堂>教师食堂。

学生晚餐就餐地点人数排名：第四食堂>第二食堂>第五食堂>第三食堂>第一食堂>教师食堂。

任务2.2 通过食堂刷卡记录，分别绘制工作日和非工作日食堂就餐时间曲线图，分析食堂早中晚餐的就餐峰值。

工作日和非工作日食堂就餐时间曲线图如下所示：

由图可得出结论：工作日食堂早中晚餐的就餐峰值分别是9点、11点、18点，非工作日食堂早中晚餐的就餐峰值分别是7点、11点、21点。

任务2.3 根据上述分析的结果，为食堂的运营提供建议。

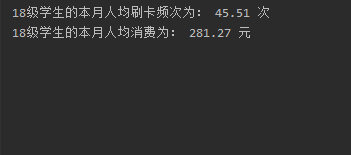
对食堂的营运提供的建议：

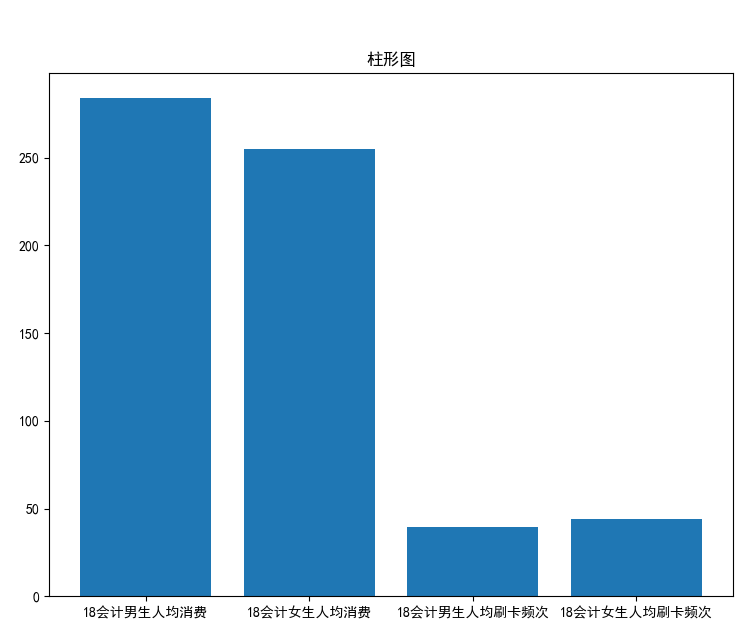
1. 第二食堂早餐人数占比为一天中的最大值，建议第二食堂增加早餐种类更好地满足学生需求和增加食堂营收；
2. 第三食堂、第四食堂的中餐和晚餐人数基本为当天营收主要收入，建议第三、第四食堂增加中餐和晚餐的菜品满足学生需求和促进食堂营收。
3. 建议参考工作日与非工作日的食堂早中晚餐就餐峰值调整食堂厨房的炒菜时间，让更多学生可以吃上热饭菜。

### 3、任务三 对学生消费行为进行分析

任务3.1 根据18级学生的整体校园消费数据，计算本月人均刷卡频次和人均消费，分析不同专业间不同性别学生群体的消费特点。

筛选数据表中属于“消费”类型的数据，去除一次就餐行为中多次刷卡记录，计算4月该校学生总刷卡频次，进一步计算出本月人均刷卡频次和人均消费数据。筛选数据表中属于“会计”专业的数据，并且筛选男生女生，分别计算出18级会计专业男生和女生的消费总金额和刷卡频次，最后计算出不同性别学生群体人均消费。

本月人均刷卡频次和人均消费如下图所示：

18级会计专业不同性别学生群体人均消费对比柱形图如下所示：

任务3.2 根据学生的整体校园消费行为，选择合适的特征，构建聚类模型，分析每一类学生群体的消费特点，为学校判定学生的经济情况提供参考意见。

选择消费金额和刷卡消费频次作为聚类特征，分析学生的整体校园消费行为。例如，学生A和学生B的消费金额相同，学生A的刷卡消费频次比学生B的多，则说明学生A的经济情况比较差；学生A和学生B的刷卡消费频次相同，学生A的消费金额比学生B的大，则说明学生A的经济情况比较好。

通过构建聚类模型后可以看出，学生群体分为两类：一类学生经常在食堂消费，消费金额较大；另一类学生较少在食堂消费，并且消费金额较小。