# 一、背景介绍

某学校存在需求，希望通过分析学生的校园消费，改进学校的服务，并为其相关部门的决策提供信息支持。通过分析了解学生的消费情况，判断学生的经济状况，精准援助，减少各种审核流程。

# 二、项目目标

1、分析学生的消费行为和食堂的运营状况，为食堂运营提供建议；

2、构建学生消费细分模型，为学校判定学生的经济状况提供参考意见。

# 三、数据说明

## 3.1、数据来源

数据来源：泰迪云课堂

数据起始至截止时间：2019-4-1至2019-4-30

## 3.2、数据量

学生消费信息：近52万条

学生基础信息：4342条

# 四、任务一：数据预处理过程

## 1.判断是否存在缺失值、异常值、无效值

1. 删除字段内容全为空的数据，这些字段无分析的价值，只会影响运行的速度；
2. 判断消费金额字段是否为负数，按常识来说消费金额为非负数，则为异常数据，若出现负数应当删除；
3. 判断时间字段是否出现异常时间，如：深夜凌晨时间，学校食堂的营业时间是固定的，通常凌晨和深夜是不对学生进行开放的。

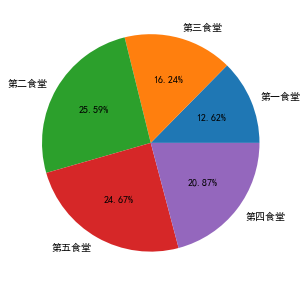
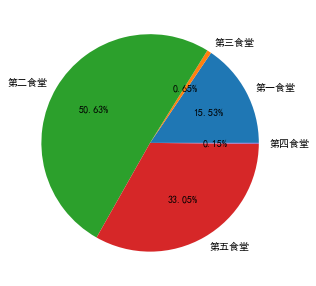
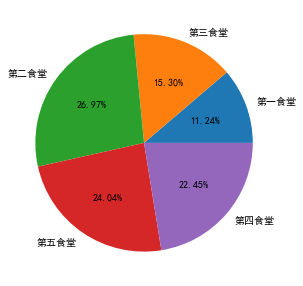
## 进行学生个人信息和消费记录关联

由于两个表共同拥有CardNo字段，且是数组中的唯一值，故将其作为关联表的连接字段，将两个表进行关联。

# 五、任务二：就餐行为分析

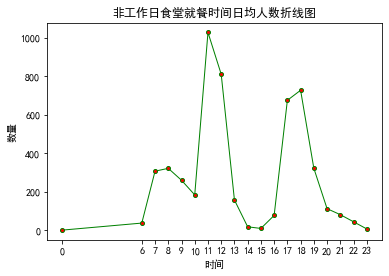
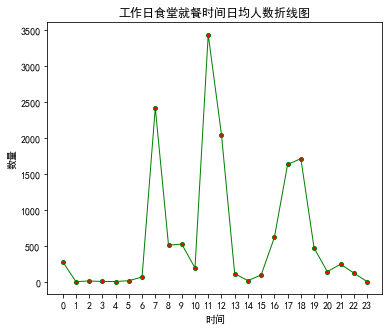
## 分析学生早中晚餐的就餐地点是否有显著区别

将数据中Dept字段数据为：食堂一、食堂二、食堂三、食堂四、食堂五的进行分组，统计这五个食堂四月份的累计消费刷卡次数，按早中晚时间段进行统计，绘制出三个时间段的各个食堂学生就餐人次饼图，结果如下：



## 分析食堂早中晚餐的就餐峰值

将数据中Dept字段数据为食堂的数据，统计这五个食堂四月份的每日消费刷卡次数，按工作日和非工作日时间段进行统计，绘制出两个时间段中不同时间的日均就餐人数折线图，结果如下：



## 提供食堂运营建议

根据早中晚餐的各食堂学生就餐人次饼图，可以分析出早餐时间段学生更倾向于在第二食堂进行就餐，其次是第五食堂，最少人的是第四食堂，人次占比低至0.15%；午餐时间段各食堂的就餐人次相对平均，其中第二食堂的就餐人次最高；晚餐时间段主要集中在第五食堂和第二食堂，占比分别为24.07%和26.87%。建议学校可以调整食堂早中晚餐的开放地点，早上只开放第一食堂、第二食堂、第五食堂，可以减少学校食堂的运营成本，中餐和晚餐时间段由于就餐人次较多，全部食堂都开放。

根据工作日和非工作日的不同时间日均就餐人数折线图，可以看出在工作日时，食堂的就餐峰值集中在6点-8点，10点-13点，16点-19点，这三个峰值符合早中晚餐时间段；非工作日食堂就餐峰值出现在10点-13点和16点-20点，可能是非工作日学生起床时间延后和时间比较自由，就餐峰值持续时间比工作日更长。建议学校食堂在工作日的6点-8点，10点-13点，16点-19点时间段增加食堂工作人员，为学生提供良好的就餐服务；在非工作日时，学校食堂的工作时间可以延迟开放，不仅经营成本上可以减少，而且食堂工作人员可以有比较充足的休息，后面的工作效率也会提高。

# 六、任务三：消费行为分析

## 分析不同专业间不同性别学生群体的消费特点

根据18级学生的整体校园消费数据，计算出4月份人均刷卡频次和人均消费金额，按不同专业和不同性别学生进行分组，统计各专业的男女生的人均刷卡频次和人均消费金额。

## 分析不同类的学生群体的消费特点

根据学生的整体校园消费行为，选择学生的日平均刷卡频次、日平均消费金额，以及月平均刷卡频次、月平均消费金额为聚类特征，构建聚类模型。

## 提供学校判定学生经济状况建议

由于各专业学生群体的消费特点不同，学校在精准扶贫时，对于不同专业可以进行不同扶贫名额分配，规定在某一消费范围内，判定该学生为需要扶贫对象。