学生校园消费行为分析项目分析报告

目录

[一，项目背景 3](#_Toc20020)

[二，项目目的 3](#_Toc1697)

[三，项目目标 3](#_Toc3327)

[四，项目分析过程 3](#_Toc16104)

[任务1：数据探索与预处理 3](#_Toc7762)

[任务2：食堂就餐行为分析 5](#_Toc5139)

[任务3：学生消费行为分析 9](#_Toc23688)

一，项目背景

南京理工大学教育基金会启动的名叫“暖心饭卡”的项目，目的是针对特困生的温饱问题进行“精准援助”。项目专门针对贫困本科生的“温饱问题”进行援助。在学校一卡通中心，教育基金会的工作人员找来了全校一万六千余名在校本科生9月中旬到11月中旬的刷卡记录，对所有记录进行了大数据分析。最终圈定了500余名“准援助对象”。

二，项目目的

某学校想要通过分析学生的校园消费行为，改进学校的服务，并为其相关部门的决策提供信息支持

三，项目目标

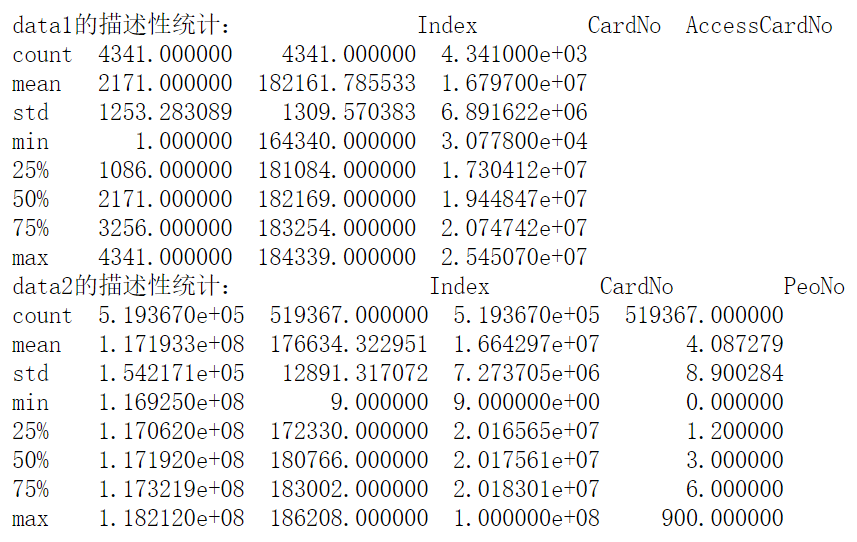
1，分析学生的消费行为和食堂的运营状况，为食堂运营提供建议。

2，构建学生消费的细分模型，为学生判定学生的经济状况提供参考意见。

四，项目分析过程

**任务1：数据探索与预处理**

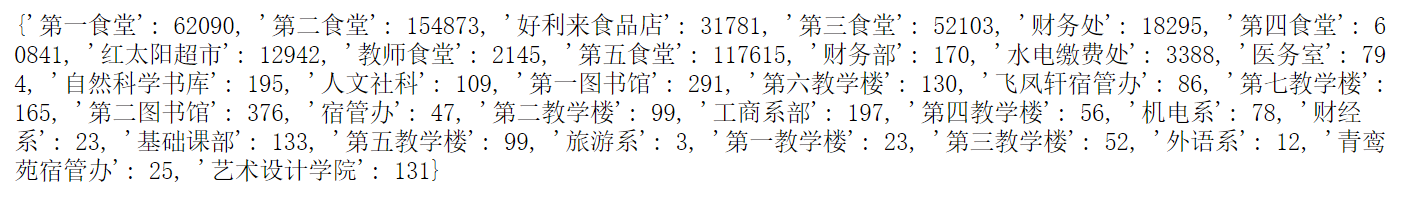
1.1，根据字段描述进行分析，对提供的数据data1和data2进行描述性统计、观察数据的维度和大小，分析数据的集中趋势、数据离散程度、数据的频数分布等

data2的大小：(519367,12)

1.2，对data2的数据查看缺失值，如若发现缺失值则判断是否删除有缺失值的数据，或者对数据进行填充。分析发现“消费项目的序列号”和“消费操作的编码”存在大量缺失值且为无关列，删除对应两列，其他列未发现缺失值。

|  |  |
| --- | --- |
| index | 0 |
| CardNo | 0 |
| PeoNo | 0 |
| Date | 0 |
| Money | 0 |
| FundMoney | 0 |
| Surplus | 0 |
| CardCount | 0 |
| Type | 0 |
| TermSerNo | 512106 |
| con0perNo | 519116 |
| Dept | 0 |

1.3，对data2的消费地点进行了词频统计，观察学生消费地点的分布情况。



1.4，对数据进行异常值的判断。观察消费金额、存储金额、消费次数有无出现负数的情况，从中取出数据的最小值，发现最小值非负。再对数据进行月份的筛选，观察数据是否存在超出4月的范围的情况，发现数据在时间上无异常。而由于可以对同一件物品做同样的消费，所以没有进行去重处理。

1.5，结合实际情况，半夜食堂不营业的客观因素，假定21点-06点(深夜/午夜-黎明)内，饭堂不营业。分别提取21点-06点的数据、1点-5点的数据22点-23点的数据。发现在1点-5点内存在饭堂营业的情况，剔除这些数据，由于项目是对食堂的数据进行研究，其他数据保留与否并无太大关系，此处是做了保留。发现在22点-23点存在较多数据，则不定义成异常值处理，重新定义提取的时间为6点-23点为该校食堂正常营业的时间。

1.6，最后将我们预处理之后的数据与data1数据进行关联，使用merge函数得到task\_1\_x.

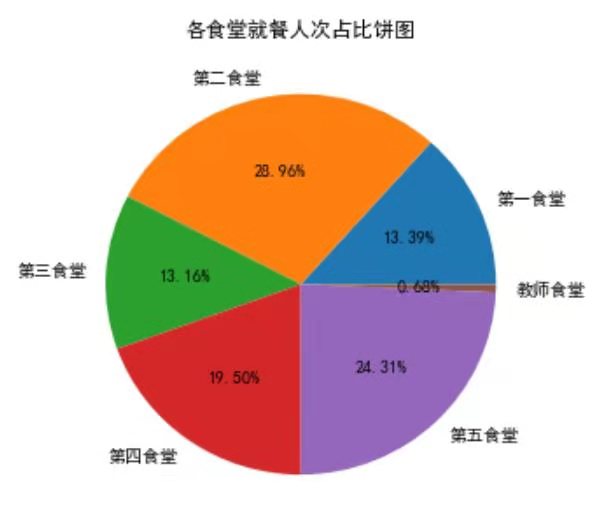
**任务2：食堂就餐行为分析**

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。

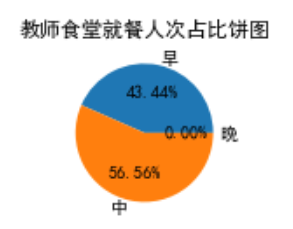
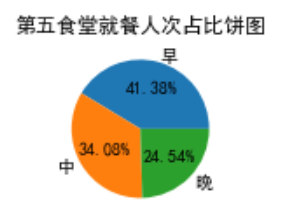
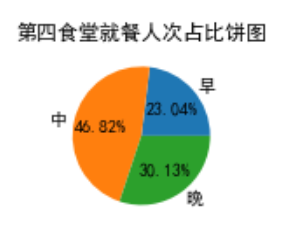
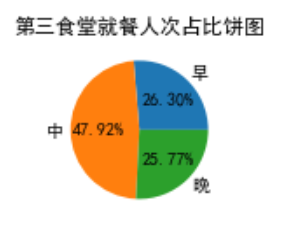
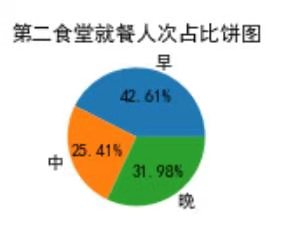
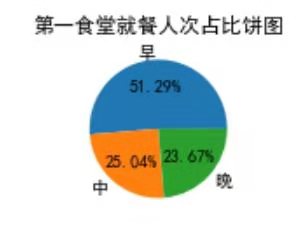
首先筛选出六个食堂的记录，然后根据不同食堂，按照当天一个小时以内的消费算作一次就餐行为来计算就餐人次，即根据校园卡号、天和时间进行一次特征重复去重，最后再统计去重之后的频次。

|  |  |
| --- | --- |
| 第一食堂 | 36142 |
| 第二食堂 | 78170 |
| 第三食堂 | 35529 |
| 第四食堂 | 52621 |
| 第五食堂 | 65615 |
| 教师食堂 | 1840 |

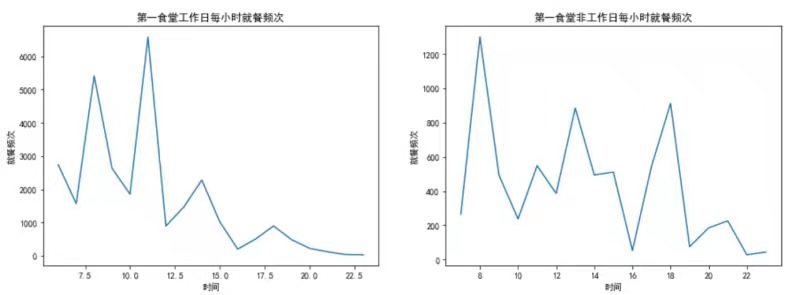
最后画出六个食堂的就餐人次占比饼图：

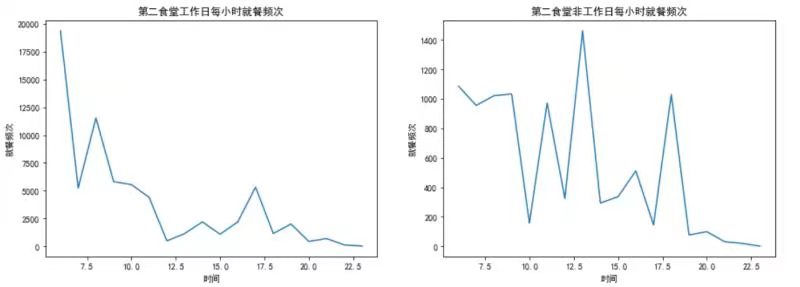


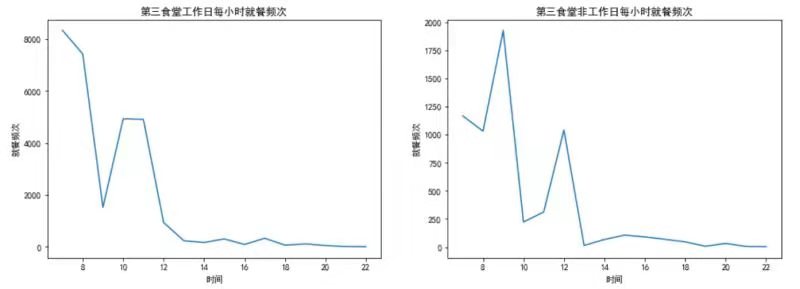
2.2，根据提取数据时的时间，划分早中晚时间区域，这里将6点-12点表示'早',将12点-18点表示'中',将18点-24点表示'晚'，得出各食堂早中晚的数据大小，画出饼图：

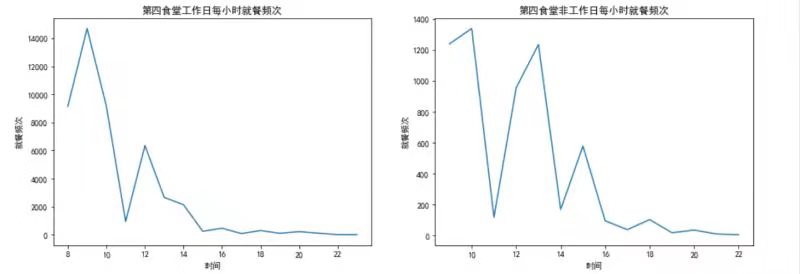


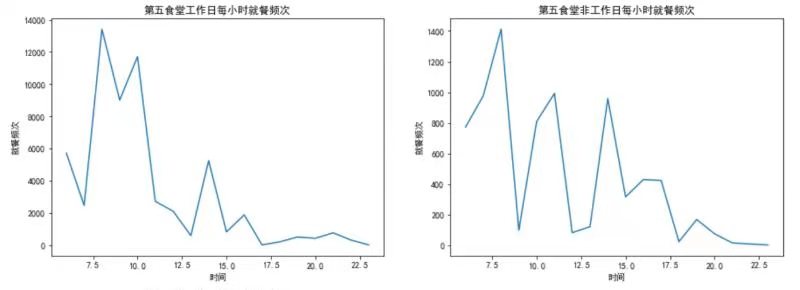
2.3，使用上面提取到的小时后的数据，划分工作日(周一至周五，周六日，清明假期)，并计算每个食堂每小时的就餐频次，绘制工作日和非工作日食堂就餐时间曲线图。

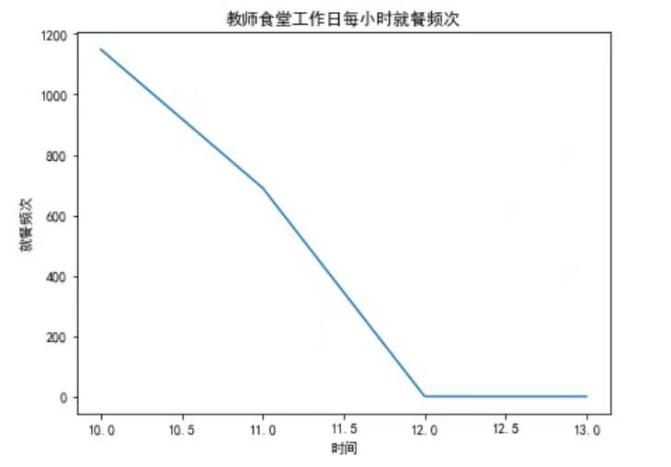












2.4，根据上述的分析结果，为食堂的运营提供建议。

根据观察，除了教师食堂在非工作日时无人就餐，其余食堂无论是在工作日和非工作日，就餐峰值几乎都在‘早’，然后随着时间逐渐减少。由此可以建议该校食堂，可以增加‘早’的食品供给量，其次是在别的就餐高峰期之后，适当减少食品供给量。

**任务3：学生消费行为分析**

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。

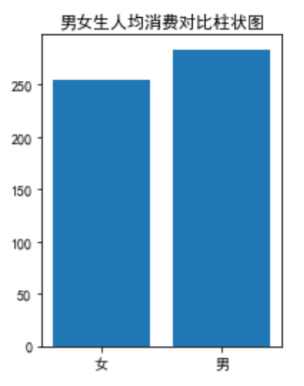
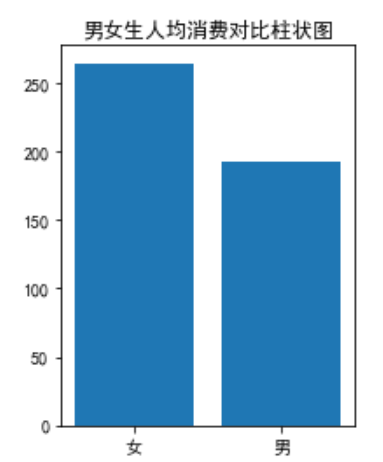
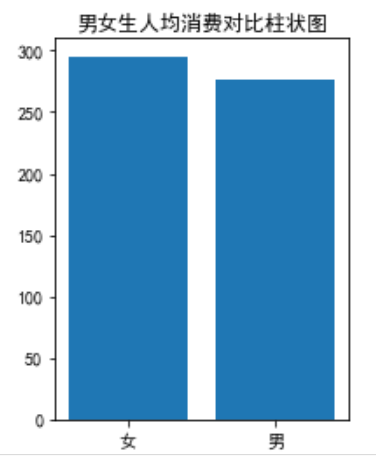
首先是计算本月人均刷卡频次和人均消费，使用最后合并的数据task1\_1\_x，算出总消费金额和该月消费的人数。最后得到人均消费为285.27，人均刷卡频次为74.16。

计算各专业不同性别学生的本月人均刷卡频次和人均消费，根据消费数据提取出各专业的数据，之后提取出各个专业男女生人均消费和人均刷卡频次做对比柱状图。

18会计： 18国际商务： 18市政工程：

女生9505人 女生7777人 女生351人

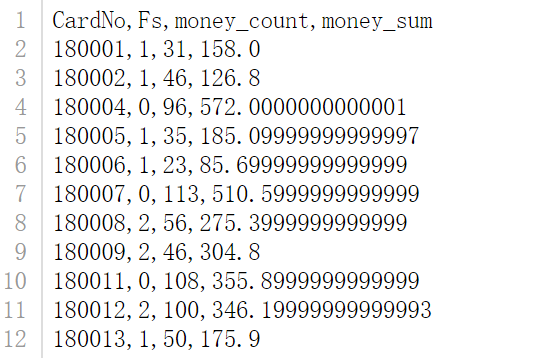
男生506人 男生519人 男生1442人

由于柱状图很多，这里展示比较特别的几种类型。18会计男生比女生少得多，但是男生人均消费却比女生的高；18国际商务的女生比男生多得多，人均消费也是；18市政工程的女生比男生少得多，但是女生人均消费却比男生的高。

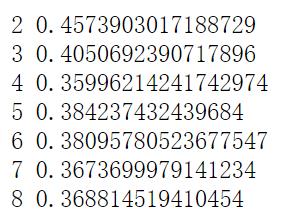
3.2根据学生的整体校园消费行为，选择合适的特征，构建聚类模型，分析每一类学生群体的消费特点，为学校判定学生的经济状况提供参考意见。

首先是选择我们的聚类指标，这里我选择用消费次数和消费总额作为我的聚类指标。先计算每个学生的消费次数和消费总额，将数据整合为task3\_X1.csv，然后对数据进行均值-方差标准化，虽然这样会降低最后聚类评价指数，但是能方便我们进行不同类别的分类。最后将聚类结果保存成task3\_X2.csv.



发现数据分成了三类，分别用0，1，2表示。分析得出： 1为校园卡消费次数少，且消费总额低的同学；2为校园卡消费次数较多，且消费金额较大的同学；0为校园卡消费次数多，且消费金额大的同学

再对我们的聚类进行轮廓系数评价法，使用for循环找出最优的分类k值。



由结果看出，分成2类时聚类效果最好，但是评价得分并不高，分成2类也只有0.457左右。说明聚类指标的选取可能不够贴切实际，所用的数据还需要更进一步的精准提取。