**目录**

1. **背景介绍**
2. 项目背景
3. 分析目标
4. **数据说明**
5. 数据来源
6. 数据量
7. 数据预处理
8. **不同时间、不同地点的消费数据分析**
9. 各食堂就餐人次所占比例
10. 分析早中晚餐就餐地点是否有显著差异
11. 工作日与非工作日学生就餐时间分布
12. **聚类分析**
13. 分析不同专业、不同性别群体的消费特点
14. 构建聚类模型分析不同群体消费特点
15. **小结**
16. **背景介绍**

**1.项目背景**

随着现代科学技术不断发展，学校管理现代化方向意识也不断增强。基于卡片应用的计算机管理系统已日益普及，信息化管理能够高速地传递信息，极大地推动了学生管理工作的高效运行，使学校管理更精细、周到。

校园一卡通系统作为高校数字化校园建设的重要组成部分，对学校管理和决策发挥着重要作用。校园一卡通动态掌握每一持卡人情况，极大提高学校的管理水平，系统每天产生大量消费流水数据，充分利用这些数据进行分析，能够很好地为学校管理与决策提供有效的建议。

**2.分析目标**

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

1. **数据说明**

**1.数据来源**

数据来源：某高校

数据期限：2019年4月1号——2019年4月30号

**2.数据量**

表data1： 18级学生个人信息数据，4341条

表data2： 学生校园一卡通数据，519367条

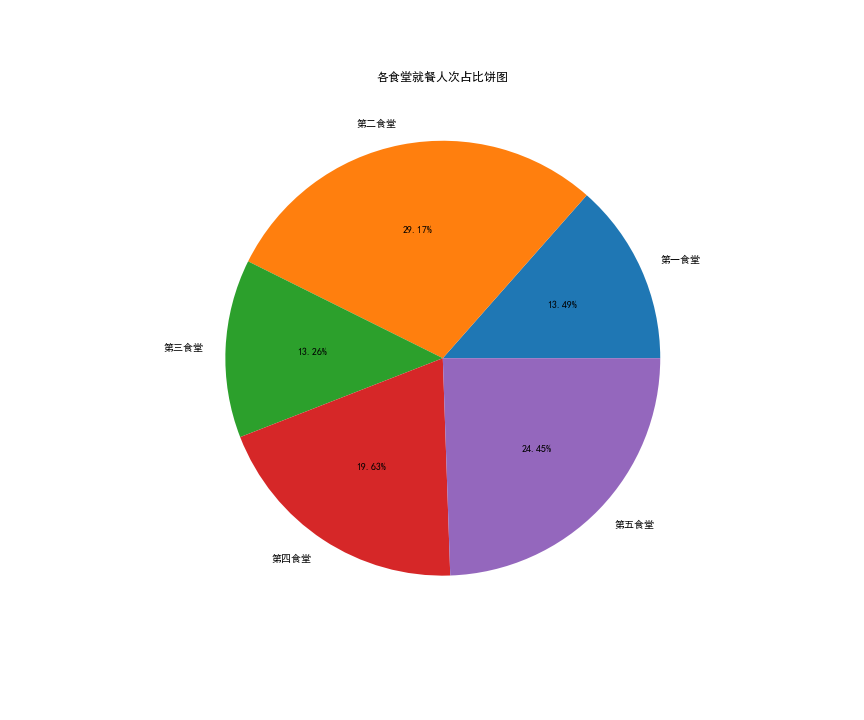
**3.数据预处理**

表data1：对数据进行去重（没有重复值）、删除缺失值（没有缺失值）

表data2：对数据进行去重（没有重复值）、删除缺失值（没有缺失值）、去掉异常值（去掉零点到六点的刷卡记录）

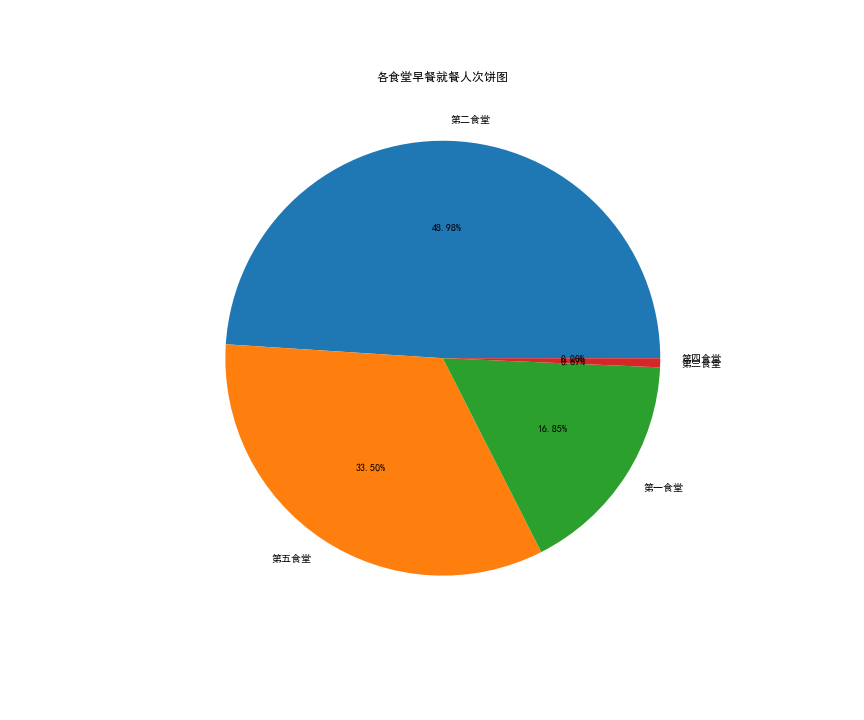
**三．不同时间、不同地点的消费数据分析**

**1.各食堂就餐人次所占比例**

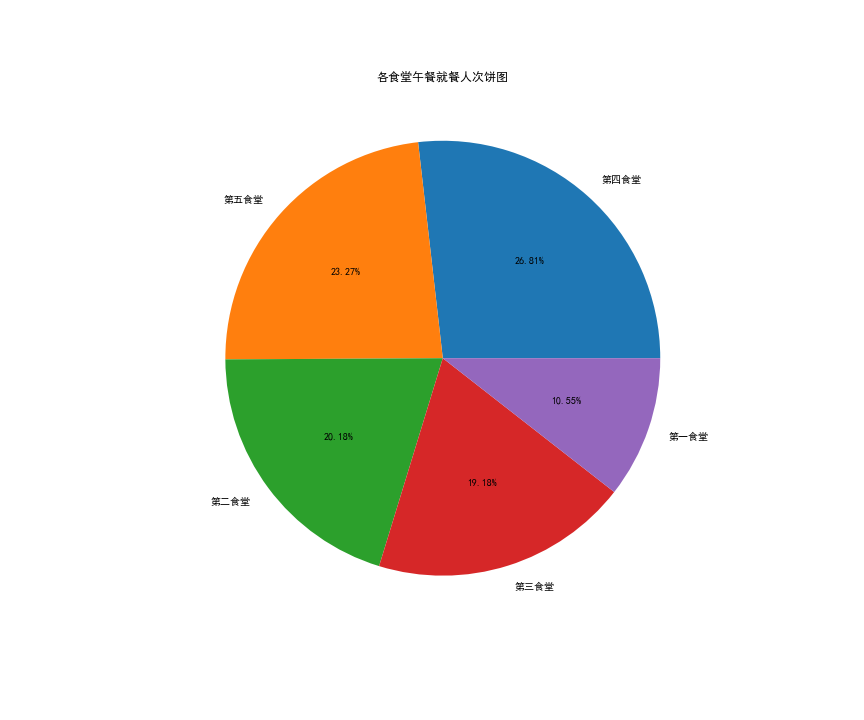


针对学生就餐地点进行分析，第二食堂就餐人次所占比例最高，说明整体上第二食堂比较受学生青睐。而第一和第三饭堂就餐人次相对较少，食堂可以适当增加菜色，饭菜分量充足，价格适中，迎合大多数人的口味，吸引更多的学生前来就餐。

**2.分析早中晚餐就餐地点是否有显著差异**

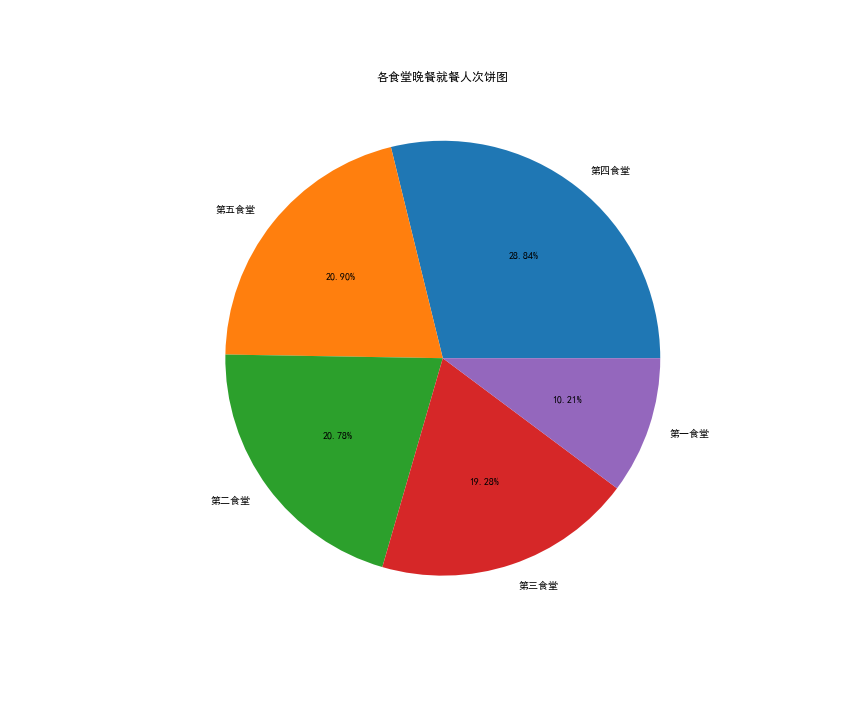
****早餐就餐地点各食堂人次

|  |  |  |
| --- | --- | --- |
| 排名 | 食堂 | 人次 |
| 1 | 第二食堂 | 25826 |
| 2 | 第五食堂 | 17664 |
| 3 | 第一食堂 | 8883 |
| 4 | 第三食堂 | 352 |
| 5 | 第四食堂 | 2 |
| 合计 |  | 52727 |



午餐就餐地点各食堂人次

|  |  |  |
| --- | --- | --- |
| 排名 | 食堂 | 人次 |
| 1 | 第四食堂 | 26336 |
| 2 | 第五食堂 | 22860 |
| 3 | 第二食堂 | 19822 |
| 4 | 第三食堂 | 18841 |
| 5 | 第一食堂 | 10363 |
| 合计 |  | 98222 |

晚餐就餐地点各食堂人次

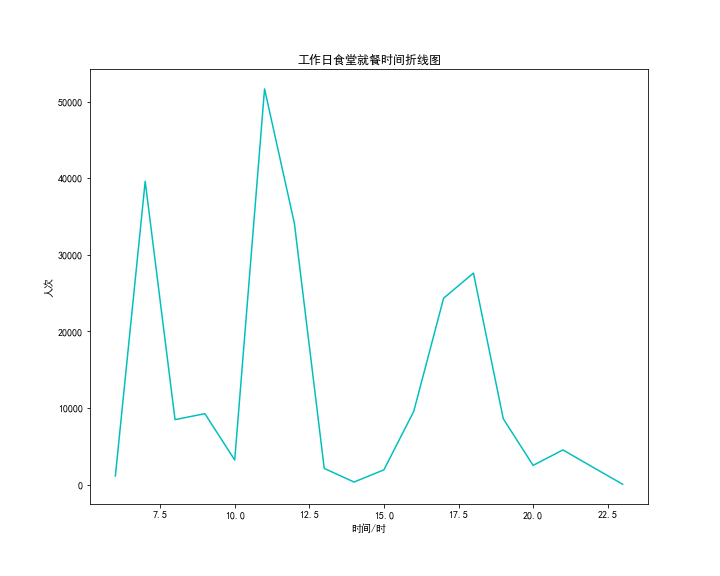
|  |  |  |
| --- | --- | --- |
| 排名 | 食堂 | 人次 |
| 1 | 第四食堂 | 17804 |
| 2 | 第五食堂 | 12901 |
| 3 | 第二食堂 | 12831 |
| 4 | 第三食堂 | 11902 |
| 5 | 第一食堂 | 6303 |
| 合计 |  | 61741 |

对学生的早中晚餐就餐地点进行分析，由以上三个饼图可知，午餐和晚餐的就餐地点没有显著差异，说明学生对午餐和晚餐就餐地点的选择整体上变化不大。

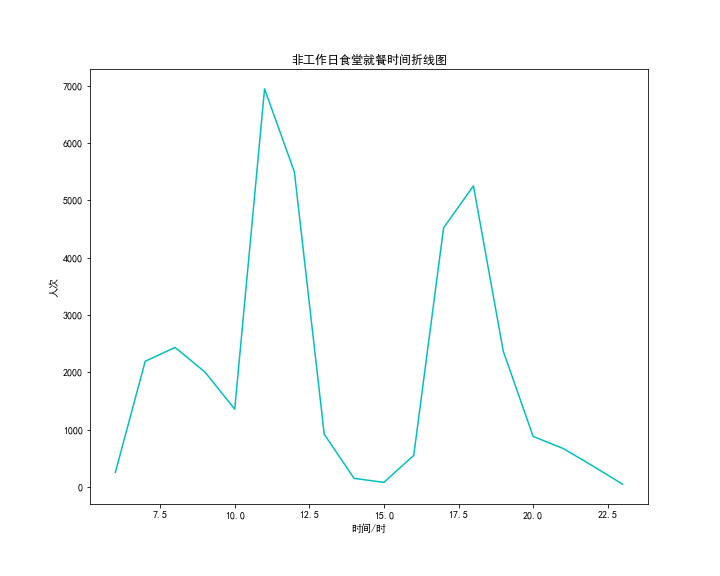
而早餐的就餐地点分布有显著差异，而且早餐食堂就餐总人数少于午餐食堂就餐总人数的1/2，说明很多人都没有在食堂吃早餐。特别是第四食堂，一个月只有两条早餐就餐记录，与午餐、晚餐记录形成鲜明对比（两个极端），早餐人数过少可能与学生个人生活习惯有关，许多学生没有吃早餐的习惯。第四食堂可以不提供早餐，专注于午餐和晚餐，提高食堂资源利用率和竞争力。

**3.工作日与非工作日学生就餐时间分布**

工作日时间折线图：



非工作日时间折线图:

由以上两个折线图可知，工作日和非工作日时间折线图趋势大致相同，工作日就餐人次明显多于非工作日。工作日早餐就餐峰值在7点半左右，午餐就餐峰值在11点到12点之间，晚餐就餐峰值大约在18点。非工作日早餐就餐峰值在8点左右，午餐就餐峰值在11点到12点之间，晚餐就餐峰值也大约在18点。非工作日早餐峰值比工作日晚了一点，其他大致相同。

**四．聚类分析**

**1.分析不同专业、不同性别群体的消费特点**

根据18级校园整体消费数据进行计算：

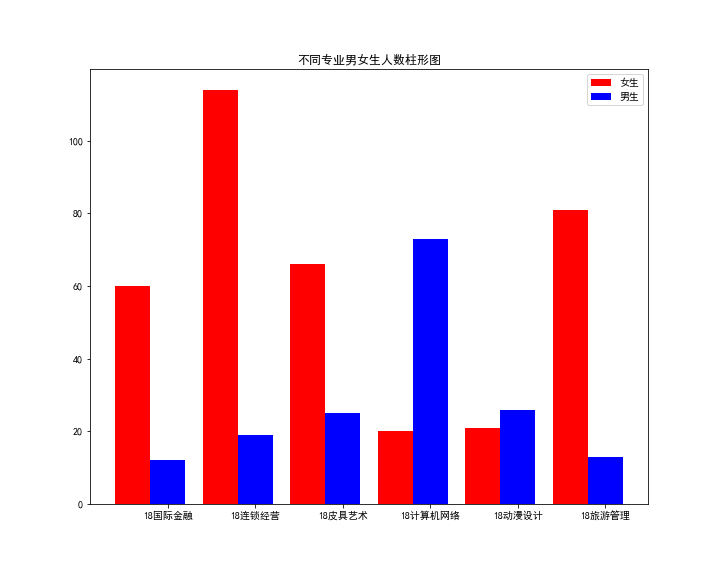
本月人均刷卡频次为：71.62160474555104

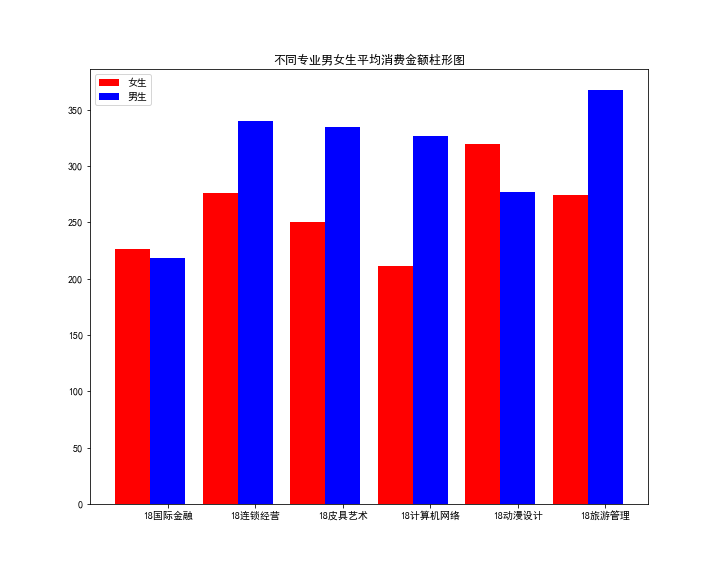
人均消费金额为：285.2007867624103

选取六个专业不同性别群体的消费数据进行分析：

专业：18国际金融、18连锁经营、18皮具艺术

18计算机网络、18动漫设计、18旅游管理





选取的六个专业中，四个专业女多男少，两个专业男多女少。

从整体来看，男生的校园平均消费要高于女生。同一性别不同专业之间的平均消费有一定差异。女生平均消费多数在200--300之间，其中18动漫设计的平均消费金额最高。男生平均消费多数在250--350之间，其中18旅游管理专业人数很少，但平均消费金额最高。

**2.构建聚类模型分析不同群体消费特点**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 群体 | dept\_mean均值 | 消费金额均值 | 刷卡次数均值 | 样本个数 | 样本数占比 |
| 群体一 | 42.86 | 214.6946 | 55.5488 | 1208 | 38％ |
| 群体二 | 25.1366 | 361.711 | 101.419 | 1033 | 32.5％ |
| 群体三 | 25.0497 | 150.0768 | 37.4786 | 936 | 29.5％ |

根据18级学生食堂消费数据进行聚类分析：

分别给第一食堂、第二食堂、第三食堂、第四食堂、第五食堂赋值为：

10、20、30、40、50

群体一：该群体学生人数最多，1208人，占总体的38％，就餐食堂均值为42.86，消费金额均值为214.7元，刷卡频次均值为55.5次。该类人群常去的就餐地点偏向第四食堂，消费金额处于中等水平。

群体二：该群体学生学生人数为1033，占总体32.5％，就餐食堂均值为25.1，消费金额均值为361.7元，刷卡频次均值为101.4次。该类人群常去的就餐地点偏向第二食堂和第三食堂，消费金额均值最高，刷卡次数最多，属于高消费水平。

群体三：该群体学生学生人数为936，人数最少，占总体29.5％，就餐食堂均值为25.05，消费金额均值为150.08元，刷卡频次均值为37.5次。该类人群常去的就餐地点偏向第二食堂和第三食堂，消费金额均值最低，刷卡次数最少，处于低消费水平。

**五．小结**

为顺应高校信息化管理的发展，将学生校园一卡通数据运用python数据分析软件进行整合与分析，主要针对某校学生校园一卡通的消费数据进行分析。首先， 通过数据预处理手段将数据进行优化，提高数据的准确性。然后，通过饼图对各个食堂的运营状况进行比较，分析早中晚餐的就餐情况。再从工作日和非工作日方面对消费行为进行对比。最后，根据学生整体消费行为，选用合适的特征，采用一种优化的K-means算法进行聚类分析，所得结果将学生分为几个群体，并分析不同群体的行为特征。此分析可为学校食堂制定经营策略提供依据，为学校判定学生经济情况提供参考意见。