

# 学生消费行为分析

学校：惠州学院

---

班级：16 应用统计学一班

---

名字：李敏辉

---

背景与目标.....	1
背景.....	1
目标.....	1
学生消费行为分析.....	1
数据预处理.....	1
任务 1：数据探索和预处理.....	1
任务 1.1.....	1
任务 1.2.....	1
学生消费行为分析.....	2
任务 2：食堂就餐行为分析.....	2
任务 2.1.....	2
任务 2.2.....	3
任务 3：学生消费行为分析.....	4
任务 3.1.....	4
任务 3.2.....	5

## 背景与目标

### 背景:

随着中国高等教育的快速发展,高校在校学生人数逐年增加,导致管理压力剧增,很难及时发现学生经济存在的问题,影响了学生的健康发展。而校园一卡通作为在校师生的必备物品,记录了学生就餐、超市消费、其他缴费等数据,这些数据背后隐藏着大部分学生日常行为活动信息,而通过挖掘和分析学生日常行为活动可以提前了解学生的消费状况。本文利用数据挖掘技术对学生校园一卡通数据进行研究,分析了学生日常消费行为,分析和挖掘得到的结果可以帮助学校管理者及时为学业出现问题的学生提供引导和帮助以及为合理配置资源、制定科学决策提供支持和参考,促进学校的智能化管理。

目标:

目标一：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

目标二：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。

## 学生消费行为分析

## 数据预处理

## 任务 1: 数据探索和预处理

### 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

### 任务 1.2: 对学生个人信息表和消费记录表进行关联

结果:

本次研究对 data1,data2 数据进行空值处理，去除重复值，剔除异常值，将日期修改为时间序列等。

空值处理：根据索引 ‘AccessCardNo’ 对 data1 数据进行去重，发现处理前原有 4341 条数据，处理后变成 4336 条数据。

统计缺失值：经过统计 data1 和 data2 的数据，发现不存在缺失值。

修改成时间序列并剔除从凌晨 0 点的数据到早上 6 点前的数据。部分结果如下：

240803	2019-04-10	16:29:00	0.4	0.0	61.8	519	游標
240804	2019-04-12	17:43:00	0.6	0.0	56.5	523	游標
240805	2019-04-17	07:35:00	1.0	0.0	50.8	508	游標
240806	2019-04-20	17:44:00	0.4	0.0	64.2	517	游標
240807	2019-04-08	18:08:00	2.5	0.0	76.2	513	游標
240808	2019-04-09	07:34:00	1.2	0.0	86.9	508	游標
240809	2019-04-09	18:08:00	3.0	0.0	73.2	514	游標
240810	2019-04-09	18:07:00	0.6	0.0	78.7	512	游標
240811	2019-04-10	11:58:00	0.4	0.0	62.2	518	游標
240812	2019-04-10	11:53:00	0.6	0.0	51.3	513	游標
240813	2019-04-10	17:44:00	1.0	0.0	69.2	514	游標
240814	2019-04-09	07:36:00	0.6	0.0	85.3	510	游標
240815	2019-04-10	17:44:00	3.0	0.0	70.2	515	游標
240816	2019-04-17	16:32:00	0.0	80.0	81.7	549	浮球
240817	2019-04-19	18:01:00	4.5	0.0	71.8	553	游標

关联 data1 和 data2 数据表，部分结果如下：

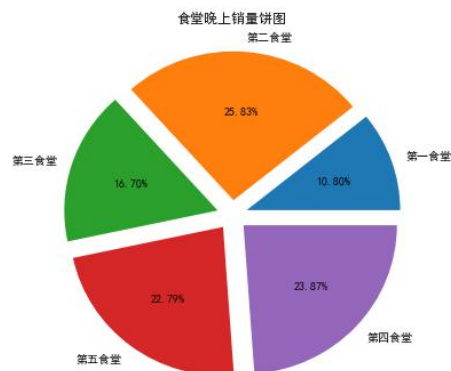
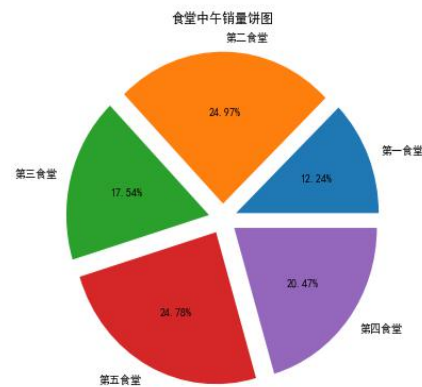
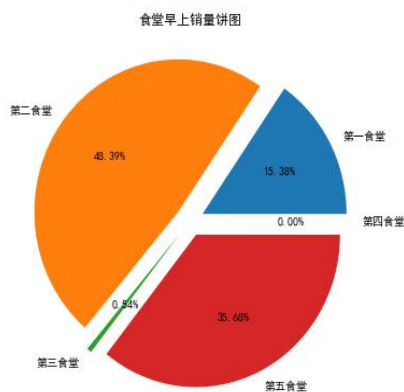
	Index_x	CardNo	Sex	Major	AccessCardNo	Index_y	PeoNo	\
0	1	180001	男	18国际金融	19762330	117331517	20181	
1	1	180001	男	18国际金融	19762330	117341866	20181	
2	1	180001	男	18国际金融	19762330	117154618	20181	
3	1	180001	男	18国际金融	19762330	117200032	20181	
4	1	180001	男	18国际金融	19762330	117206146	20181	
5	1	180001	男	18国际金融	19762330	117367323	20181	
6	1	180001	男	18国际金融	19762330	117378955	20181	
7	1	180001	男	18国际金融	19762330	117380758	20181	
8	1	180001	男	18国际金融	19762330	117222956	20181	
9	1	180001	男	18国际金融	19762330	117225959	20181	
10	1	180001	男	18国际金融	19762330	117229404	20181	
11	1	180001	男	18国际金融	19762330	117247387	20181	
12	1	180001	男	18国际金融	19762330	117248328	20181	

## 学生消费行为分析

### 任务 2：食堂就餐行为分析

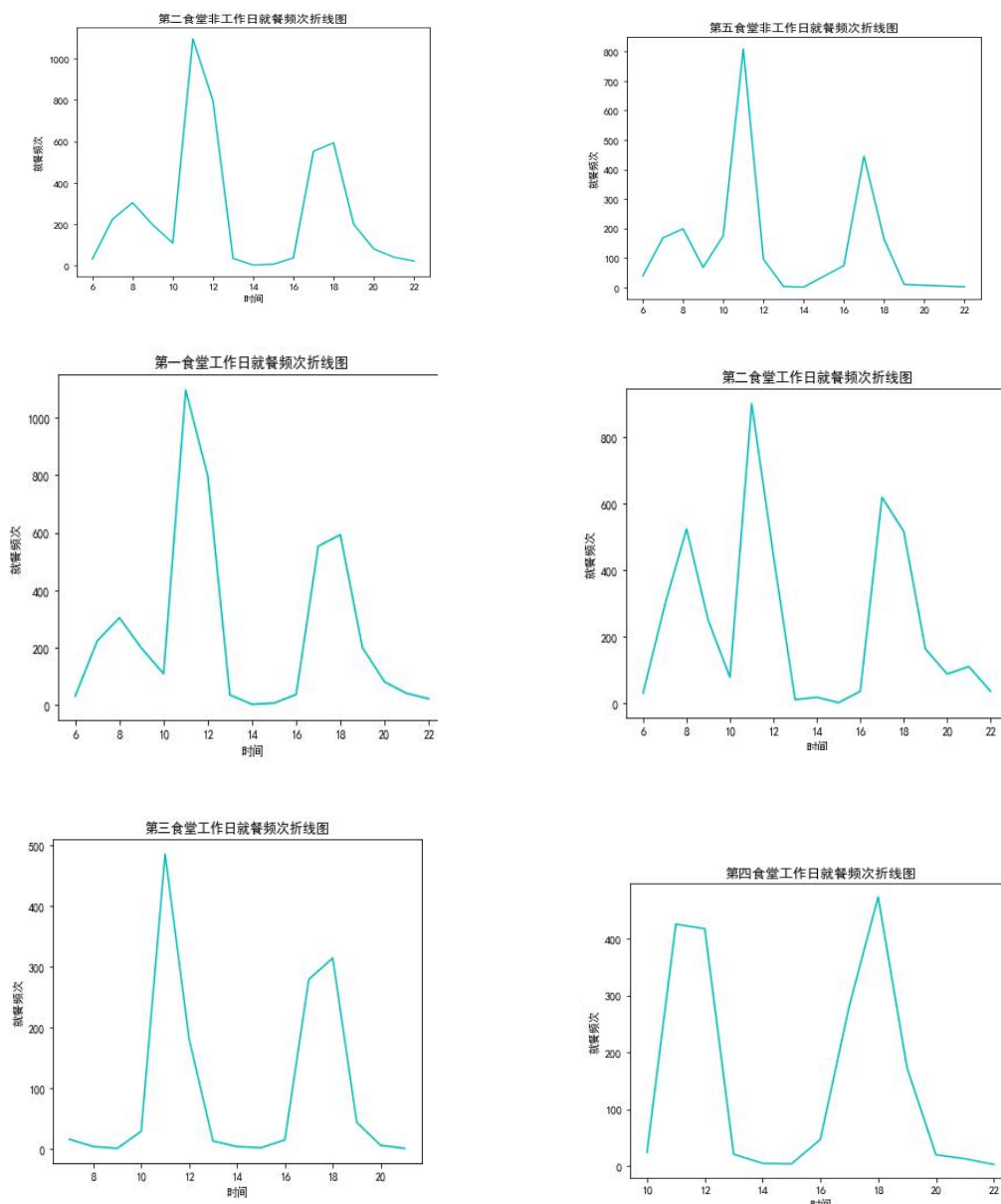
任务 2.1：绘制各食堂就餐人次的占比饼图，分析学生早中晚餐的就餐地点是否有显著差别

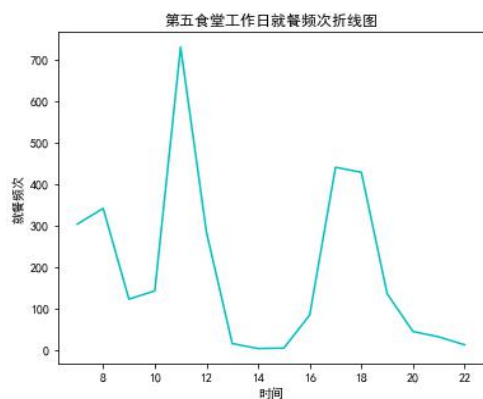
结果：



分析：可以看出，在早上最受学生欢迎的食堂是第二食堂，销量占比总体的 48.39%，其次是第五食堂，占 35.68%，最低的是第三食堂，销量惨淡；在中午，依旧是第二和第五食堂销量最高，从总体看，五个食堂销量差距不是很大；在晚上，第二食堂占 25%，第四食堂占 24%。由此可以分析出，面对高流量的第二食堂，可以提高第二食堂的菜品数量和丰富程度，扩大食堂的用餐空间，提供多个取餐窗口等，方便学生打饭和就餐。而面对早上销量极少的第三食堂，建议关闭此食堂的早上供应，以减少费用成本。

**任务 2.2：** 分别绘制工作日和非工作日食堂的就餐时间曲线图，分析食堂早中晚的就餐峰值





分析：根据以上就餐频次折线图，可得出以下表格：

	早上	中午	下午
第二食堂	360 (8: 00am)	1100 (11: 00am)	600 (18: 00pm)
第五食堂	200 (8: 00am)	800 (11: 00am)	450 (17: 00pm)

	早上	中午	下午
第一食堂	350 (8: 00am)	1100 (11: 00am)	400 (18: 00pm)
第二食堂	500 (8: 00am)	900 (11: 00am)	600 (17: 00pm)
第三食堂	40 (7: 00am)	480 (11: 00am)	300 (18: 00pm)
第四食堂	NAN	420 (11: 00am)	480 (18: 00pm)
第五食堂	350 (8: 00am)	750 (11: 00am)	450 (17: 00pm)

### 任务 3：学生消费行为分析

**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。

结果:

Major	
18 产品艺术	312.686190
18 会计	284.225000
18 动漫设计	278.196077
18 商务日语	322.000000
18 商务英语	401.063333
18 国贸实务	286.736364
18 国际商务	192.750000
18 国际金融	218.916667
18 宝石鉴定	251.327273
18 审计	282.115385
18 嵌入式技术	291.357895
18 工业工程	283.830435
18 工业机器人	350.466444
18 工业设计	243.508261
18 工商企管	208.905000
18 工程造价	248.100789
18 市场营销	224.272500
18 市政工程	278.214000
18 建筑工程	285.372833
18 建筑设计	313.868333
18 投资与理财	315.650000
18 旅游管理	350.085714
18 机械制造	327.356529
18 机械制造(学徒)	454.107143
18 模具设计	409.335065
18 汽车检测	343.312381
18 物流管理	300.516000
18 环境艺术	310.928571
18 电子商务	318.957895
18 电气自动化	368.067424
18 皮具艺术	340.577600
18 社会工作	309.425000
18 艺术设计	197.705000
18 视觉传播	379.260000
18 计算机应用	340.823721
18 计算机网络	321.149600
18 软件技术	268.394231
18 连锁经营	346.700000
18 酒店管理	406.224286
18 金融管理	231.987407
18 首饰设计	279.109091

Major	
18 产品艺术	289.985000
18 会计	255.403835
18 动漫设计	331.290476
18 商务日语	234.817407
18 商务英语	269.338421
18 国贸实务	250.519444
18 国际商务	266.128283
18 国际金融	223.305968
18 宝石鉴定	266.048505
18 审计	247.530149
18 嵌入式技术	205.114286
18 工业工程	298.254167
18 工业机器人	293.344483
18 工业设计	324.882759
18 工商企管	278.633585
18 工程造价	310.225273
18 市场营销	222.678182
18 市政工程	335.300000
18 建筑工程	290.905556
18 建筑设计	292.519412
18 投资与理财	288.689655
18 旅游管理	271.367590
18 机械制造	331.900000
18 模具设计	202.741667
18 物流管理	287.606765
18 环境艺术	285.858696
18 电子商务	271.760000
18 电气自动化	304.200000
18 皮具艺术	257.816667
18 社会工作	253.863030
18 艺术设计	217.012766
18 视觉传播	333.495417
18 计算机应用	217.364286
18 计算机网络	212.494500
18 软件技术	236.036364
18 连锁经营	276.154957
18 酒店管理	229.923421
18 金融管理	249.922200
18 首饰设计	313.158205

左图:男 右图:女

分析:由图知,男生最高消费的专业是 18 机械制造(学徒),人均消费金额约 454 元,最低的是 18 国际商务,人均约 193 元。女生消费最高的专业是 18 使市政工程,人均消费金额约 335 元,最低是 18 模具设计,约 202 元。

任务 3.2:根据学生的整体校园的消费行为,选择合适的特征,构架聚类模型,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。

CardNo	moneysum	moneycount	Fs
180001	161.6	35	0
180002	126.8	47	0
180004	572	100	1
180005	185.1	36	0
180006	85.7	23	0
180007	512.6	117	1
180008	275.4	59	2
180009	304.8	48	2
180011	385.9	112	1
180012	346.2	103	2
180013	175.9	51	0
180014	436.5	140	1
180015	345.9	86	2
180016	299.9	118	2
180017	289.7	87	2
180018	139.9	31	0
180021	172	61	2
180022	85.9	17	0
180023	80.5	1	0
180024	142.7	31	0
180025	202.2	76	2
180026	295.5	109	2
180027	268.5	73	2
180029	54.3	10	0
180030	99.1	27	0
180031	269.8	104	2
180032	186.2	30	0
180035	54.5	18	0
180036	502.7	119	1
180037	171.6	32	0
180038	217.1	87	2
180039	159.3	1	0
180040	229.6	64	2
180041	20.2	6	0
180042	326.7	93	2

分析:

如图,将校园学生分成三个大类,分别为 0,1,2。 0 为校园卡消费次数较少的同学; 1 为校园卡消费次数较多,且消费金额较大的同学,为高消费水平; 2 为 1 为校园卡消费次数较多,且消费金额较小的同学,经济状况较差。根据三个不同类别,针对类别 0 和 2,学校可以提供相对适应的助学贷款,解决学生的经济困难。