目录

[1. 背景介绍 2](#_Toc4023)

[2. 项目目标 2](#_Toc31456)

[3. 数据说明 2](#_Toc17052)

[4. 数据预处理 2](#_Toc28916)

[4.1 处理data1 2](#_Toc12725)

[4.2 处理data2 3](#_Toc14018)

[5. 食堂就餐行为分析 4](#_Toc4509)

[5.1 数据处理 4](#_Toc17206)

[5.2 就餐行为分析 5](#_Toc10081)

[5.3 食堂运营建议 6](#_Toc4529)

[6. 学生消费行为分析 7](#_Toc18746)

[6.1 数据处理 7](#_Toc20065)

[6.2 学生消费行为分析 7](#_Toc19990)

[6.3 构建聚类模型与评价 9](#_Toc10791)

# 

# 背景介绍

校园一卡通是集身份认证、金融消费、数据共享等多项功能于一体的信息集成系统。在为师生提供优质、高效信息化服务的同时，系统自身也积累了大量的历史记录，其中蕴含着学生的消费行为以及学校食堂等各部门的运行状况等信息。因此，对校园一卡通数据进行分析，有利于改进学校服务并为相关部门的决策提供信息支持。某校通过分析学生的校园消费行为，然后改进学校的服务，并为其部门的决策提供信息支持。

# 项目目标

根据国内某高校校园一卡通系统一个月的运行数据，使用数据分析和建模的方法，挖掘数据中所蕴含的信息，分析学生在校园内的学习生活行为，为改进学校服务并为相关部门的决策提供信息支持。

# 数据说明

某学校2019年4月1日至4月30日的校园一卡通数据和学生个人信息数据。data1是学生个人信息数据，共4341条数据；data2是校园一卡通数据，共519367条数据。

# 数据预处理

## 处理data1

1. 检查异常值

观察data1数据可以看到索引值为4340和4341的两名学生的校园卡号存在错误，将其分别改为184340，184341。

1. 删除字段

删除data1的AccessCardNo无关字段，简化数据结构。

1. 数据去重

对data1数据按校园卡号CardNo进行去重，可以得到4341条数据。说明data1共有4341名18级学生的信息。

## 处理data2

1. 检查缺失值

观察data2数据，发现TermSerNo和conOperNo字段存在大部分缺失值，进行删除处理。

1. 数据转换

将data2的Index,CardNo,PeoNo字段转换为字符串类型。

1. 数据去重

对data2数据按校园卡号CardNo进行去重，可以得到8636条数据。说明data2共有8636名学生的信息。

1. 修改字段

修改data2的Date字段为日期型数据，以便后续对时间列进行提取时分秒操作

1. 筛选数据

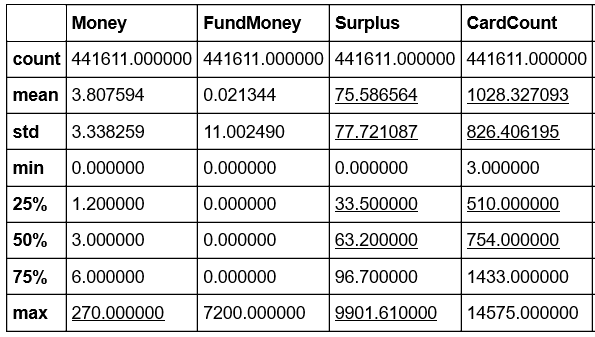
筛选消费地点为食堂的消费记录，共有449667条数据。

1. 检查异常值

筛选data2数据显示，将食堂的运营时间设定为6:00~23:00，删除其他异常值，得到441611条数据。

1. 数据描述性统计

可以看到，学生的消费金额，存储金额，余额和消费次数的极差相差很大，说明学生间的消费存在着很大的不同，在后续处理时需要缩小数据间的差距。



**表1 data2数据描述性统计**

1. 数据关联

对data1和data2按照CardNo字段进行主键合并，得到新的数据表，共207139条数据。

1. 数据统计

1）统计学生消费次数最多的消费地点，可以得到，学生消费次数最多的食堂分别为第二食堂和第五食堂，其次是第一食堂，第四食堂和第三食堂。

2）学生消费次数最多的时间为2019/04/12的15：05：00。

3）统计学生工作日和非工作日消费情况，可以看到，学生工作日消费次数远远大于非工作日消费次数。

4） 统计学生消费类型可知，学生主要消费类型为消费，存款次之。

# **食堂就餐行为分析**

## 数据处理

1.  筛选地点是食堂的记录，按小时划分中午晚餐的范围，规定早餐时间为6~11点，午餐时间为11~16点，晚餐时间为16~23点，并生成新的一列"meal"。

2. 按校卡、餐别和日期进行去重，得到235541条数据，并统计各食堂的就餐人次。

3. 划分工作日和非工作日，规定工作日对应1，非工作日对应0，并统计各自的数量；

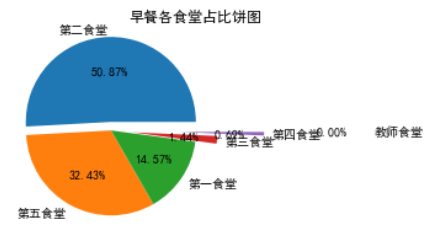
4. 因4月5号为清明假期，故筛选日期为5号的记录，将其对应的工作日改为非工作日；

5. 分别提取早午晚餐的记录，统计各食堂三餐的就餐人次，并绘制饼图；

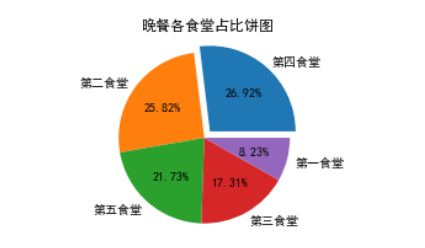
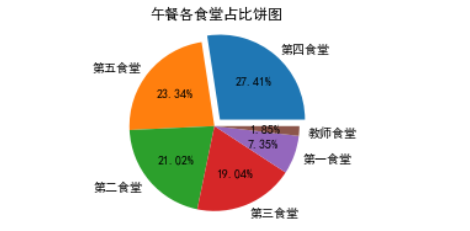
6. 分别提取各食堂的消费记录，在此基础上分别筛选工作日和非工作日的消费记录，再分别计算工作日与非工作日各食堂的平均就餐人次，绘制曲线图。

## 就餐行为分析

从饼图可以看出，三餐中学生消费次数前3的食堂分别为第二食堂，第五食堂和第四食堂。早午晚餐中占比最高的分别为第二食堂，第四食堂和第四食堂。值得注意的是，教师食堂只有早午餐的消费记录。



**图1 早餐各食堂占比饼图**



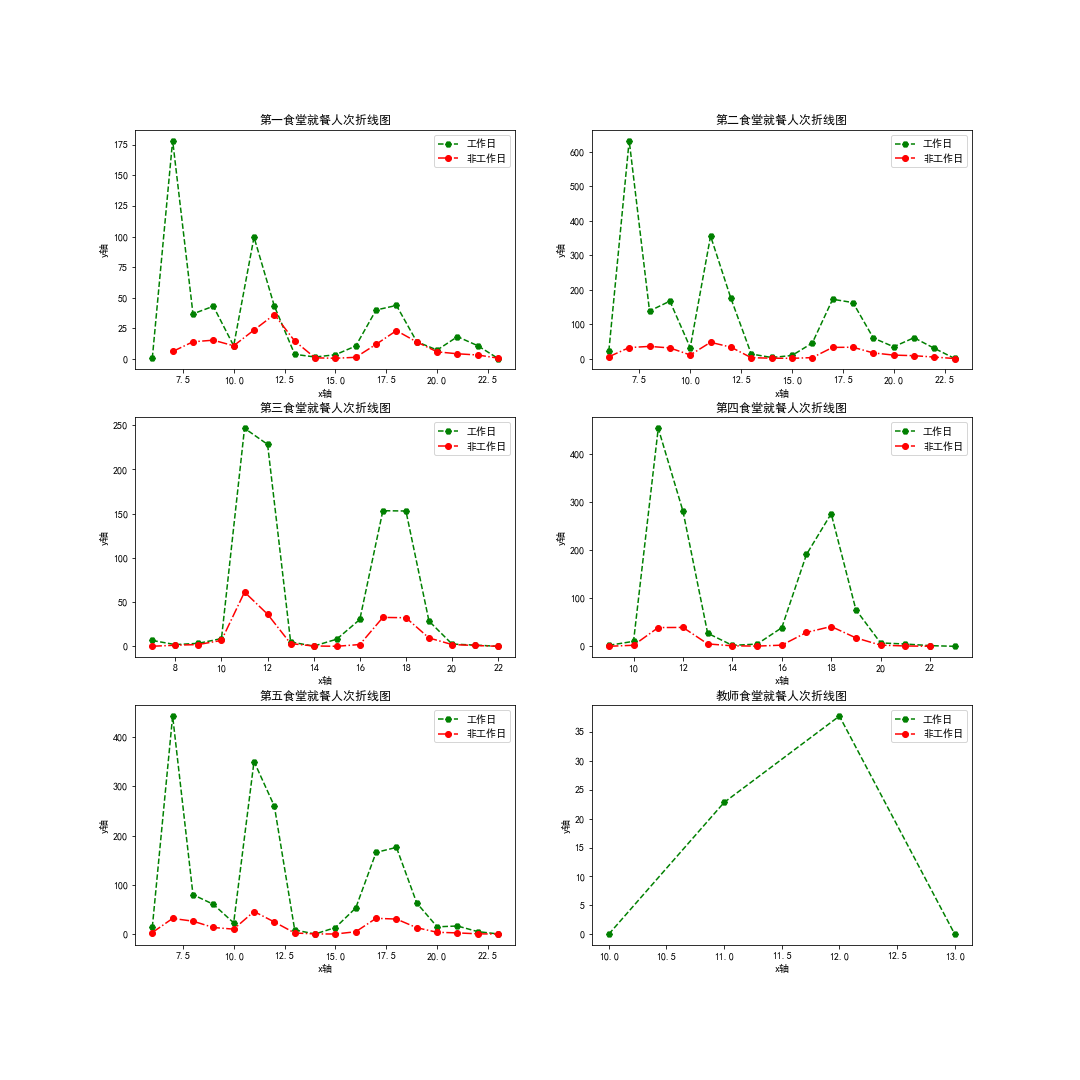
**图2 午餐各食堂占比饼图 图3 晚餐各食堂占比饼图**

从折线图可以看出，各食堂工作日的就餐人次均远高于非工作日的就餐人次。第一食堂，第二食堂和第五食堂有早午晚就餐高峰值，而第三食堂和第四食堂有午晚就餐高峰值，教师食堂只有午就餐高峰值。

第一食堂，第二食堂和第五食堂工作日的就餐早高峰值均出现在7点半，平均就餐人次第二食堂最多，第一食堂最少，午就餐高峰值均在11点左右，平均就餐人次与早高峰相同；非工作日的峰值，第一食堂出现在12点，而第二食堂和第五食堂出现在11点，平均就餐人次均不足100。

第三食堂和第四食堂工作日的午高峰值均在10点半，平均就餐人次第四食堂高于第三食堂，第三食堂晚高峰值在16点半到18点半，而第四食堂晚高峰值在18点，平均就餐人次高于第三食堂；第三食堂非工作日午高峰值在11点，而第四食堂午高峰值在11点半到12点，平均就餐人次在50左右；

教师食堂工作日的午高峰值在12点，平均就餐人次超过35，没有非工作日的消费记录。



**图4 各食堂就餐人次折线图**

## 食堂运营建议

1. 从工作日的就餐峰值来看，第一食堂，第二食堂和第五食堂的消费人次曲线图趋势一致，都有三餐的就餐峰值点。这三个食堂可以采用相同的运营模式，但在饭菜量的准备上，第二食堂需要准备的量多于第五食堂，第五食堂需要准备的量多于第一食堂。

2. 第三食堂和第四食堂只有午晚餐就餐峰值，两个食堂在运营时需要注意就餐时间的细微差别，并且第四食堂饭菜准备量大于第三食堂。

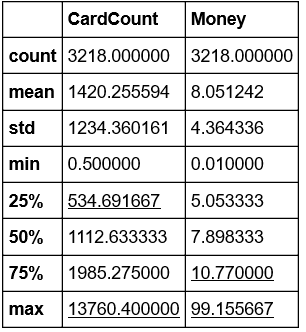
3. 非工作日的就餐，除了第一食堂和第三食堂的峰值波动明显，需要在午晚餐时段增加饭菜量，其他食堂波动不大。

4. 教师食堂重点做好工作日午餐时段的供餐需求。

# **学生消费行为分析**

## 数据处理

导入关联后的数据表，按照CardNo字段计算本月人均刷卡频次和人均消费，并对数据进行描述性统计，得到下表，可以看到，消费金额与刷卡频次的波动范围很大。

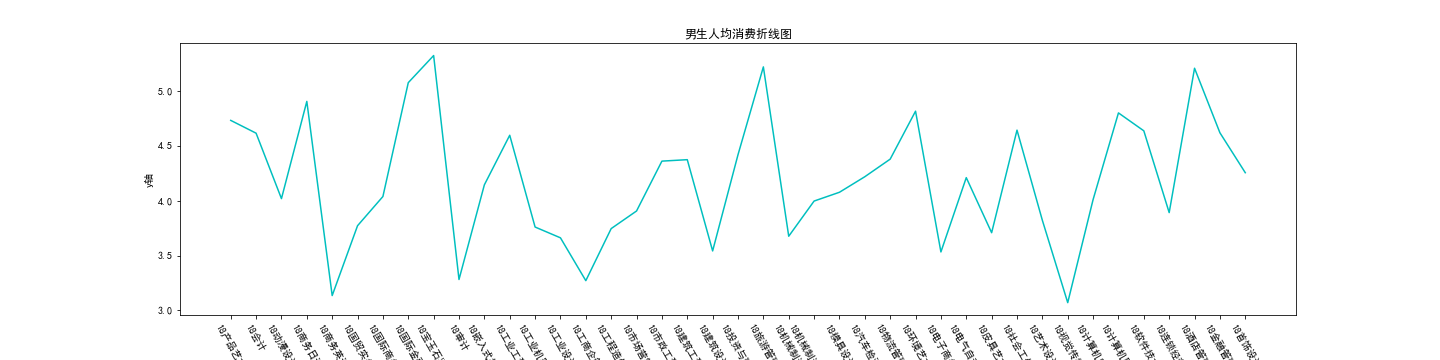


**表2 关联后数据描述性统计**

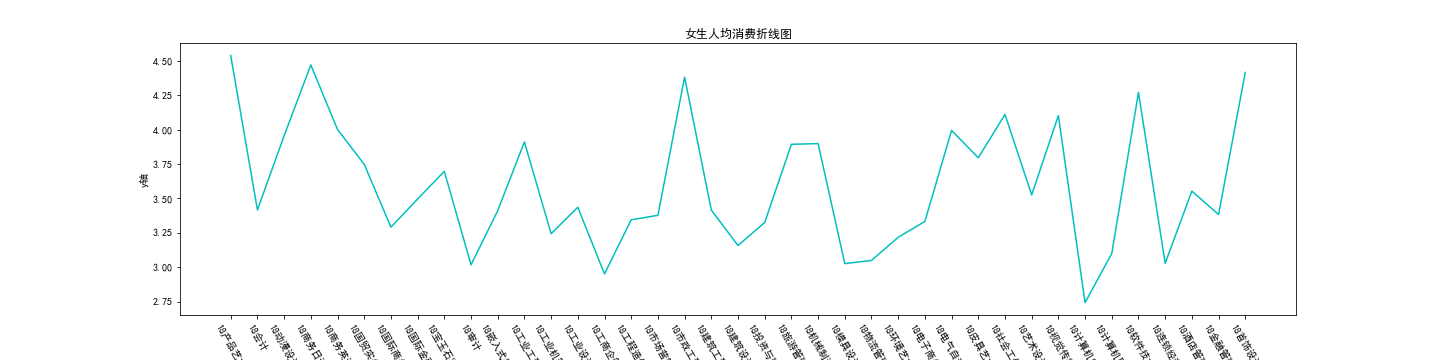
## 学生消费行为分析

按专业统计男女学生的人均消费，分别得到以下的折线图，可以看到，男生的人均消费区间在（3，5.5）之间，女生的人均消费区间在（2.75，4）之间。总体上，男生的人均消费高于女生。男生中，专业为宝石鉴定，投资与管理和金融管理的学生的人均消费在5元以上，商务英语，视觉传媒的学生人均消费在3元左右。

女生中，专业为动漫设计，市政工程和软件工程的学生人均消费为4.5左右，专业为计算机网络的学生人均消费仅为2.75左右。

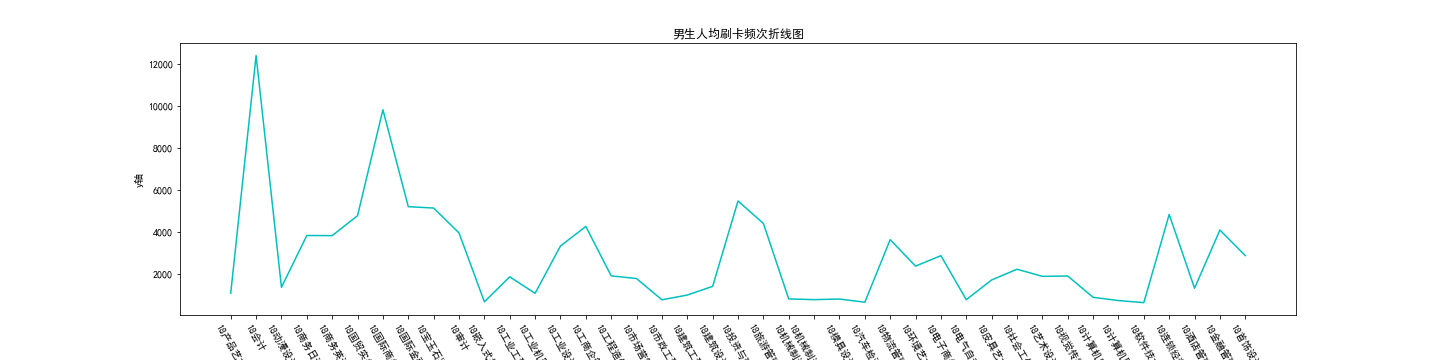


**图5 男生人均消费**

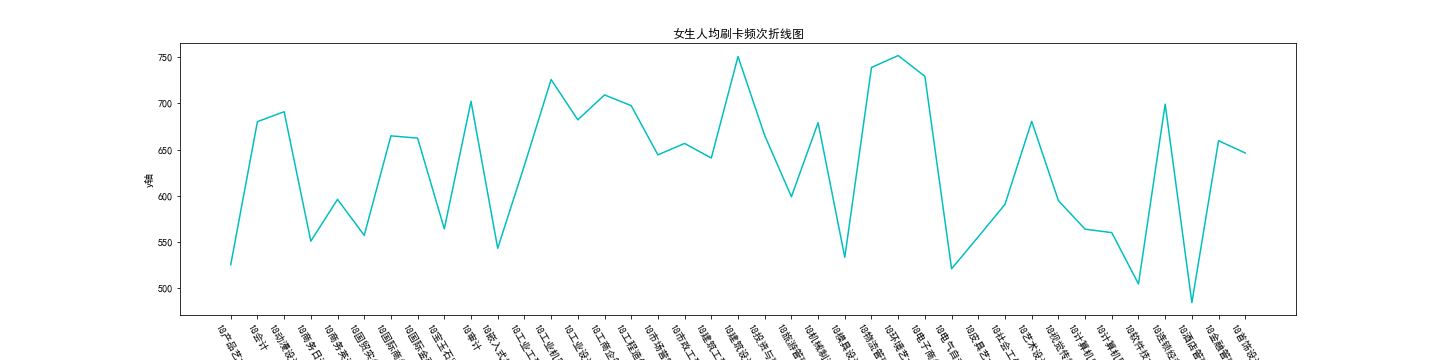


**图6 女生人均消费**

分别提取男女学生的消费记录，按专业统计男女学生的人均刷卡频次，分别得到以下的折线图，可以看到，男生人均刷卡频次的波动区间在（200，12000）之间，女生人均刷卡频次的波动区间相对于男生，范围更小，在（500，750）之间。由此可见，女生的人均刷卡频次更趋于稳定。会计，国际商务和投资与管理专业的男生人均刷卡频次在10000以上，投资理财，环境艺术和酒店管理专业的女生人均刷卡频次在650以上。

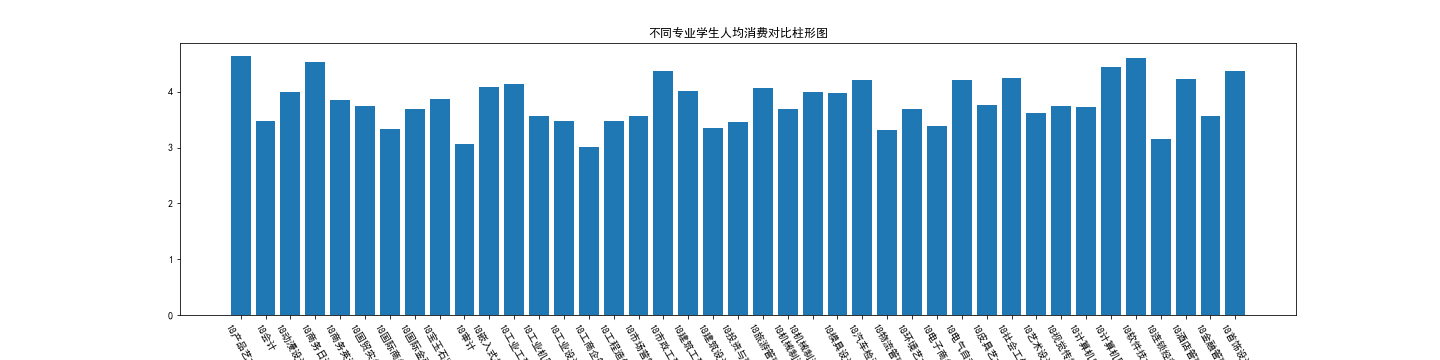


**图7 男生人均刷卡频次**

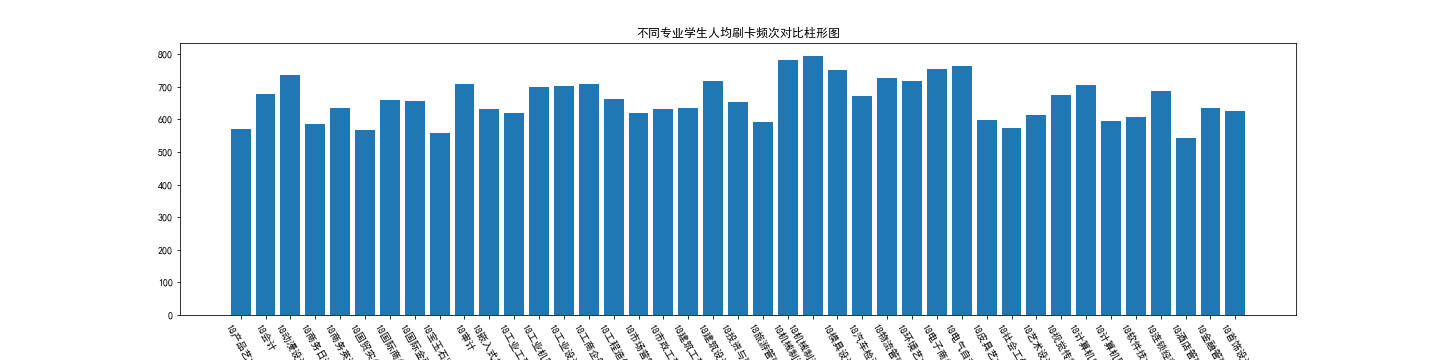


**图8 女生人均刷卡频次**

绘制不同专业学生人均消费对比柱形图，得到如下柱形图，各个专业的人均消费和人均刷卡频次分布不尽相同。专业为产品艺术，市政工程，软件工程和加算计网络专业的学生人均消费在所有专业中最高，专业为机械制造，机械制造（学徒）的学生刷卡频次最高。



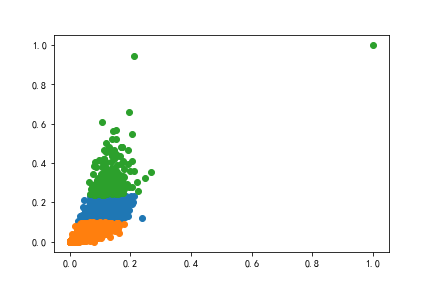
**图9 不同专业学生人均消费对比柱形图**



**图10 不同专业学生人均刷卡频次对比柱形图**

## 构建聚类模型与评价

经过上述的分析，可以知道，学生的消费金额和刷卡频次的方差波动极大，因此对关联后数据的Money和CardCount字段进行离差标准化，再构建聚类模型，并得到如下的聚类图形。



**图11 聚类图**

可以看到数据集中程度高，离散点很少。为进一步了解聚类模型效果，使用轮廓系数评价法查看维度评价得分，结果显示，2维的轮廓系数最大。