# 食堂聚类分析

聚类分析可提供数量化的衡量学生消费情况相似程度的指标，利用这些指标可以将学生划分为不同的类别，从而揭示学生之间本质上的联系，分析学生的消费特征，挖掘贫困生等。

**一、数据处理**

为了进行有效的分析，我们利用SQL Server 2016 对数据进行处理，在Canteen\_data数据库下，创建CanteenDCluster视图，其中包含：学生学号（ID）、年级（Grade）、学院（School）、专业（Major）、各餐交易次数及各餐交易额，共12个变量。

**二、聚类分析**

为了分析我校学生食堂消费情况，我们选取各餐交易次数及各餐交易额作为聚类分析变量，聚为5类。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **表\_\_1020120100\_Canteen\_data\_CanteenDCluster – 聚类分析的挖掘模型文档** | | | | | |
| **模型信息** | | | | | |
| 模型名称 | 表\_\_1020120100\_Canteen\_data\_CanteenDCluster - 聚类分析 | | | | |
| 模型说明 | 根据 表\_\_1020120100\_Canteen\_data\_CanteenDCluster 表建立的 Microsoft\_Clustering 模型 | | | | |
| 算法 | [Microsoft\_Clustering](http://go.microsoft.com/fwlink/?LinkId=103230) | | | | |
| 上次处理时间 | 2016/8/18 13:24 | | | | |
| 分类数 | 5 | | | | |
| 支持分类“分类 1” | 2045 | | | | |
| 支持分类“分类 2” | 1946 | | | | |
| 支持分类“分类 3” | 1486 | | | | |
| 支持分类“分类 4” | 1274 | | | | |
| 支持分类“分类 5” | 1289 | | | | |
| 平均值(行的概率) | 0.74136883 | | | | |
| **挖掘模型列** | | | | | |
| **列名** | **用法** | | **数据类型** | **内容类型** | **值** |
| \_\_RowIndex | 输入 | | Long | 键 |  |
| 晚餐交易次数 | 输入 | | Long | 连续 | 1 - 60 |
| 晚餐交易额 | 输入 | | Double | 连续 | 1.1 - 20 |
| 夜宵交易次数 | 输入 | | Long | 连续 | 1 - 30 |
| 夜宵交易额 | 输入 | | Double | 连续 | 0.5 - 20 |
| 早餐交易次数 | 输入 | | Long | 连续 | 1 - 59 |
| 早餐交易额 | 输入 | | Double | 连续 | 0.6 - 18.5 |
| 中餐交易次数 | 输入 | | Long | 连续 | 1 - 61 |
| 中餐交易额 | 输入 | | Double | 连续 | 0.65 - 16.5 |
| **算法参数** | | | | | |
| **名称** | | **值** | | | |
| [CLUSTER\_COUNT](http://go.microsoft.com/fwlink/?LinkId=103230) | | 5 | | | |
| [CLUSTER\_SEED](http://go.microsoft.com/fwlink/?LinkId=103230) | | 0 | | | |
| [CLUSTERING\_METHOD](http://go.microsoft.com/fwlink/?LinkId=103230) | | 1 | | | |
| [MAXIMUM\_INPUT\_ATTRIBUTES](http://go.microsoft.com/fwlink/?LinkId=103230) | | 255 | | | |
| [MAXIMUM\_STATES](http://go.microsoft.com/fwlink/?LinkId=103230) | | 100 | | | |
| [MINIMUM\_SUPPORT](http://go.microsoft.com/fwlink/?LinkId=103230) | | 1 | | | |
| [MODELLING\_CARDINALITY](http://go.microsoft.com/fwlink/?LinkId=103230) | | 10 | | | |
| [SAMPLE\_SIZE](http://go.microsoft.com/fwlink/?LinkId=103230) | | 50000 | | | |
| [STOPPING\_TOLERANCE](http://go.microsoft.com/fwlink/?LinkId=103230) | | 10 | | | |

**1.分类关系图**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  | | --- | | **分类关系由强到弱：** | | 分类1-分类5 | | 分类1-分类3 | | 分类3-分类5 | | 分类2-分类3 | | 分类1-分类2 | | 分类4-分类5 | | 分类1-分类4 | |
| 以最强关系（分类1-分类5）对比，为例： |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **变量** | **值** | **倾向于 分类 1** | **倾向于 分类 5** |
| 晚餐交易额 | 1.0 - 8.5 | 100 |  |
| 晚餐交易额 | 8.5 - 20.0 |  | 100 |
| 中餐交易额 | 1.0 - 8.3 | 96.384 |  |
| 中餐交易额 | 8.3 - 20.0 |  | 96.384 |
| 早餐交易额 | 0.9 - 5.7 | 66.121 |  |
| 早餐交易额 | 5.7 - 18.5 |  | 65.826 |
| 夜宵交易额 | 0.5 - 8.6 | 35.85 |  |
| 夜宵交易额 | 8.6 - 20.0 |  | 35.85 |
| 夜宵交易次数 | 1 - 6 | 31.082 |  |
| 夜宵交易次数 | 7 - 30 |  | 29.274 |
| 早餐交易次数 | 1 - 9 |  | 10.432 |
| 早餐交易次数 | 10 - 59 | 10.432 |  |

**2.分类详情（剖面图）**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **变量** | **状态** | **总体(全部)** | **分类 1** | **分类 2** | **分类 3** | **分类 4** | **分类 5** |
| 大小 |  | 8040 | 2096 | 1644 | 1556 | 1367 | 1377 |
| 早餐交易次数 | 平均值 | 20 | 19.41 | 32.23 | 25.65 | 5.95 | 13.7 |
| 早餐交易次数 | 偏差 | 12.31 | 8.48 | 9.74 | 11.45 | 4.12 | 7.97 |
| 早餐交易额 | 平均值 | 4.53 | 4.11 | 3.31 | 4.48 | 5.54 | 5.83 |
| 早餐交易额 | 偏差 | 1.75 | 1.02 | 0.77 | 1.14 | 2.56 | 1.79 |
| 中餐交易次数 | 平均值 | 28 | 24.77 | 39.04 | 38.32 | 10.23 | 24.4 |
| 中餐交易次数 | 偏差 | 11.88 | 6.16 | 6.93 | 6.32 | 4.76 | 7.14 |
| 中餐交易额 | 平均值 | 7.23 | 6.69 | 5.42 | 7.46 | 8.33 | 9.04 |
| 中餐交易额 | 偏差 | 1.86 | 1.17 | 0.92 | 1.09 | 2.27 | 1.42 |
| 晚餐交易次数 | 平均值 | 26 | 22.6 | 37.56 | 36.94 | 9.2 | 22.95 |
| 晚餐交易次数 | 偏差 | 12.08 | 6.43 | 7.85 | 6.98 | 4.46 | 7.48 |
| 晚餐交易额 | 平均值 | 7.47 | 6.86 | 5.67 | 7.82 | 8.55 | 9.28 |
| 晚餐交易额 | 偏差 | 1.89 | 1.18 | 0.95 | 1.08 | 2.36 | 1.42 |
| 夜宵交易次数 | 平均值 | 4 | 3.35 | 4.26 | 6.8 | 1.89 | 5.14 |
| 夜宵交易次数 | 偏差 | 3.61 | 2.1 | 3.23 | 5.27 | 1.09 | 3.53 |
| 夜宵交易额 | 平均值 | 6.29 | 5.71 | 4.45 | 6.5 | 7.47 | 8.16 |
| 夜宵交易额 | 偏差 | 2.85 | 2.19 | 1.89 | 2.29 | 3.59 | 2.87 |

根据聚类分析，将学生分为5类，各类学生消费情况及特征见下表：

|  |  |  |
| --- | --- | --- |
| 分类 | 消费特点 | 分类名称 |
| 分类1 | 早餐交易次数少于总体平均值，早餐交易额较低；  中餐交易次数少于总体平均值，中餐交易额低于总体平均值；  晚餐交易次数少于总体平均值，晚餐交易额低于总体平均值；  夜宵交易次数少于总体平均值，夜宵交易额低于总体平均值。 | 较贫困 |
| 分类2 | 早餐交易次数最多，早餐交易额最低；  中餐交易次数最多，中餐交易额最低；  晚餐交易次数最多，晚餐交易额最低；  夜宵交易次数高于总体平均值，夜宵交易额最低。 | 贫困类 |
| 分类3 | 早餐交易次数多于总体平均值，早餐交易额低于总体平均值；  中餐交易次数多于总体平均值，中餐交易额高于总体平均值；  晚餐交易次数较多，晚餐交易额较高；  夜宵交易次数最多，夜宵交易额高于总体平均值。 | 普通类 |
| 分类4 | 早餐交易次数最少，早餐交易额高于总体平均值；  中餐交易次数最少，中餐交易额较高；  晚餐交易次数最少，晚餐交易额高于总体平均值，  夜宵交易次数最少，夜宵交易额较高。 | 宽裕类 |
| 分类5 | 早餐交易次数较少，早餐交易额最高；  中餐交易次数较少，中餐交易额最高；  晚餐交易次数低于总体平均值，晚餐交易额最高；  夜宵交易次数较高，夜宵交易额最高。 | 较宽裕 |