# Python实验环境介绍

## 实验目标

## 数据集介绍

## 实验步骤

## 实验小结

# 回归算法原理与实践

## 实验目标

## 数据集介绍

## 实验步骤

## 实验小结

# 决策树原理与实践

## 实验目标

## 数据集介绍

## 实验步骤

## 实验小结

# 支持向量积

## 实验目标

本实验对已经从单卡转合约的客户的数据分析，提取相关特性来进行数据挖掘，通过svm算法预测在网具有单卡转合约倾向的客户。

## 数据集介绍

本数据时通过模拟运营商的数据，共10000条,12个字段

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **序号** | **名称** | **说明** | **类型** | **备注** |
| 1 | user\_id | 用户标识 | int |  |
| 2 | service\_kind | 业务类型 | string | 2G\3G\4G |
| 3 | call\_duration | 主叫时长（分） | double |  |
| 4 | called\_duration | 被叫时长（分） | double |  |
| 5 | in\_package\_flux | 免费流量 | double |  |
| 6 | out\_package\_flux | 计费流量 | double |  |
| 7 | month\_duration | 月均上网时长（分） | double |  |
| 8 | net\_duration | 入网时长（天） | long |  |
| 9 | last\_recharge\_value | 最近一次缴费金额（元） | double |  |
| 10 | total\_recharge\_value | 总缴费金额(元) | double |  |
| 11 | total\_recharge\_count | 缴费次数 | int |  |
| 12 | contractuser\_flag | 是否潜在合约用户 | int | 1：是，0：不是 |

## 支持向量积介绍

## 实验步骤

* + 1. 导入相应模块

读取训练数据集

数据探索

分割数据集

训练简单模型

调试参数

训练最优参数的模型

模型评估

## 实验小结

# 聚类和降维算法原理和实践

## 实验目标

使用提供的小区画像识别数据和大妈

## 数据集介绍

## 实验步骤

## 实验小结

# 关联规则原理和实践

## 实验目标

## 数据集介绍

## 实验步骤

## 实验小结

# 推荐算法原理与实践

## 实验目标

采用原来电影推荐算法，套用运营商的应用领域（如移动的咪咕视频等）

## 数据集介绍

## 实验步骤

## 实验小结

# 集成算理论和实践

## 实验目标

## 数据集介绍

## 实验步骤

## 实验小结

# 大案例研讨

## 实验目标

本案例使用电信用户的通信行为数据集，对电信用户进行分群并进行行为分析。由于是没有标注的训练样本，我们可以使用聚类等无监督方法将用户进行分群，又由于特征数较多，我们可以进行降维的方法，去除可能存在的噪声数据，以提升分群的效果。

最后通过行业专家对不同群体数据进行分析，可以确定群体的行为特点，进行客户群体画像。

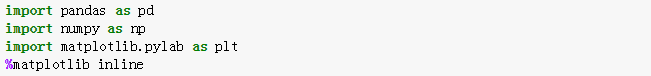
## 数据集介绍

本数据集来自公开数据集，共30000数据, 7个维度的特征，分别是：

入网时间，套餐价格，每月流量，每月话费，每月通话时长，欠费金额，欠费月份数

## 实验步骤

#### 读取数据

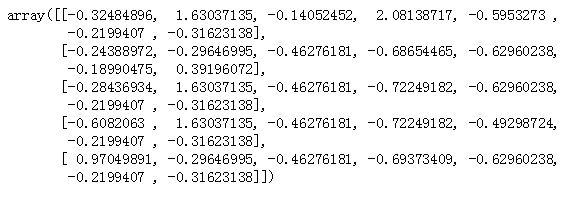




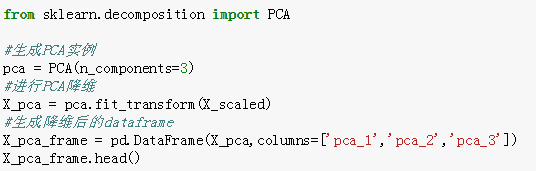


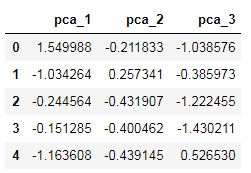
#### 数据标准化





#### 进行PCA数据降维





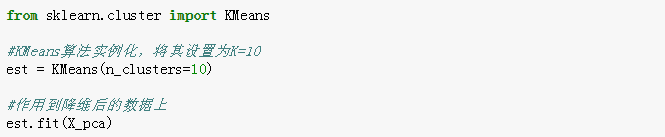
查看这三个维度坐标轴（主成分）与原始的7个维度坐标轴之间的关系：

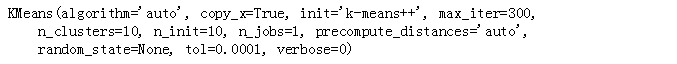


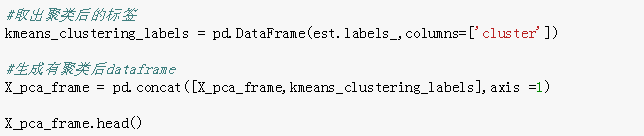


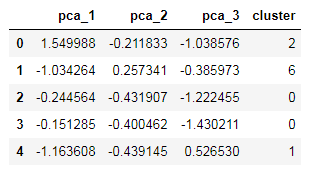
#### K-means 聚类

对降维后的三维数据进行聚类, 将其设置为K=10,使用Kmeans算法将其聚类为10类



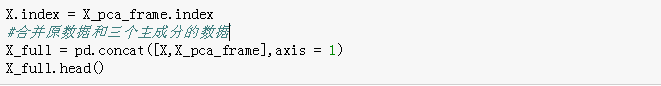






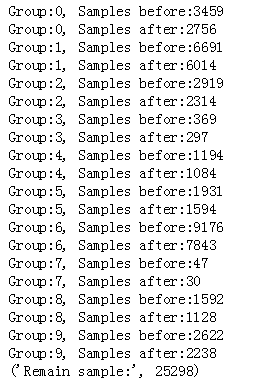
#### 样本筛选

通过对聚类数据进行筛选，使用统计中的1.5\*IQR法则，删除每个聚类中的噪音和异常点，从而提取每个聚类中的典型用户。为了避免离群点的影响，



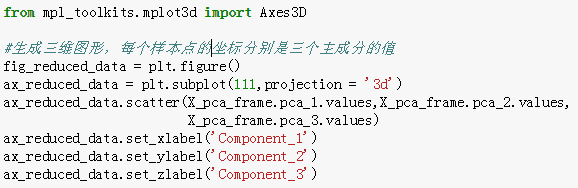


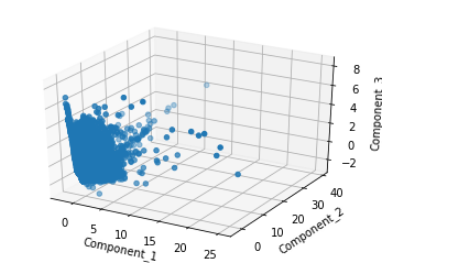




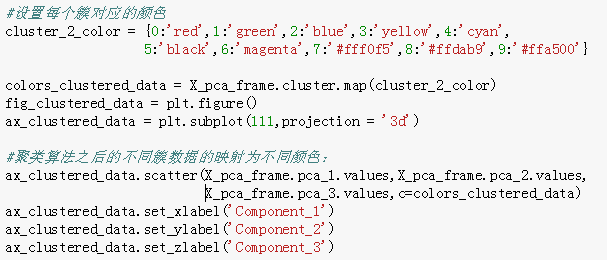
#### 查看聚类效果

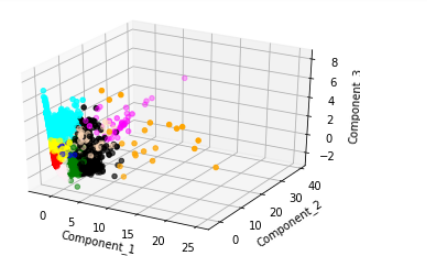
原始数据降维后的可视化



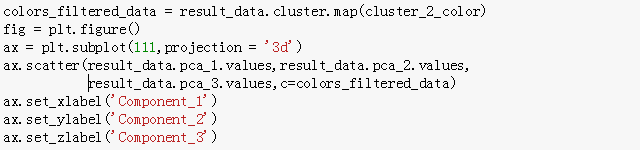


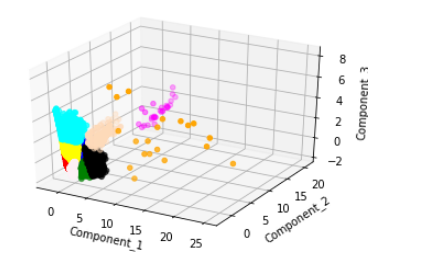
聚类算法之后的不同簇数据的可视化：





筛选后的不同簇数据聚类可视化：



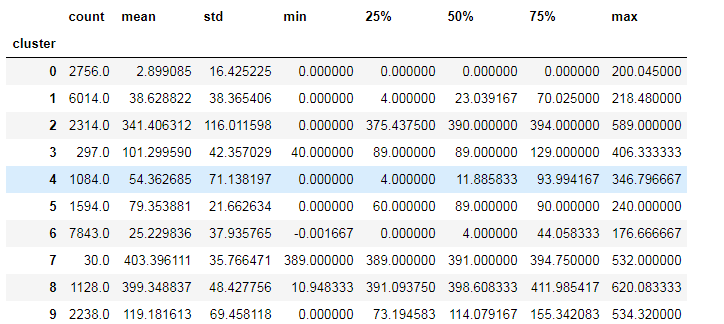


可以看到去除异常值之后的数据簇更加具有明显的分离，说明聚类效果比较好，不同的簇之间存在明显的区别。

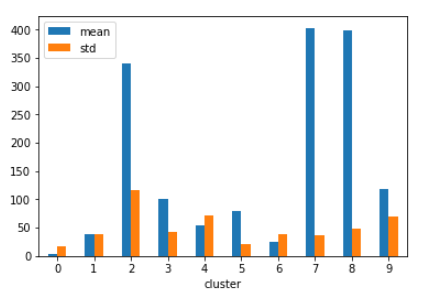
#### 用户画像（用户行为分析）

查看各族中的每月话费情况：



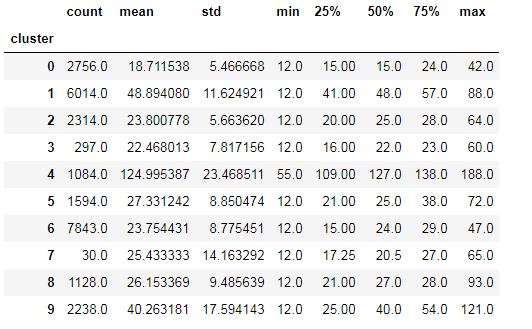




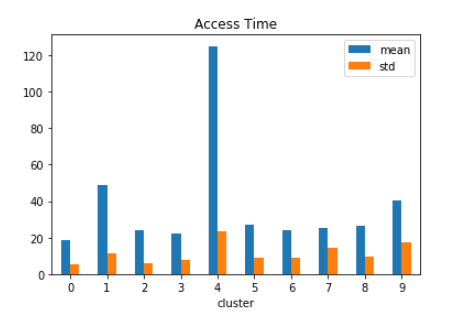


查看各族中的入网时间情况：



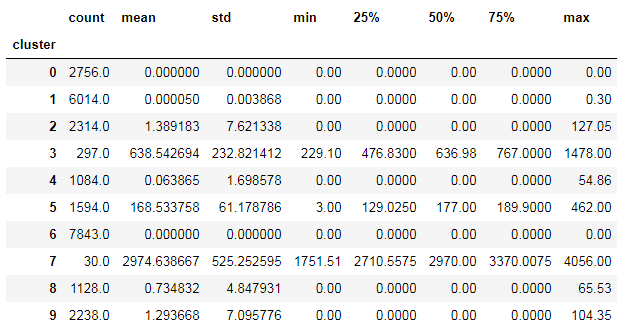




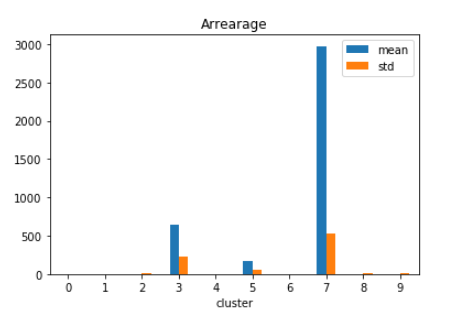


查看各族中的欠费金额情况：

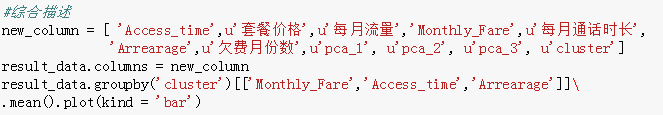


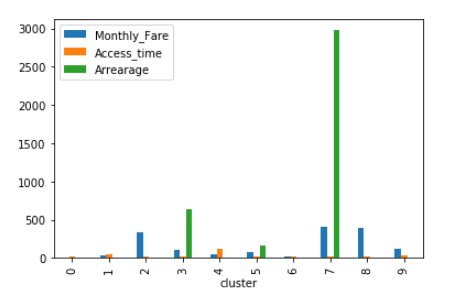






综合描述：





## 实验小结

本案例使用使用PCA降维和Kmeans聚类算法对电信用户进行分群，电信用户的通信行为数据集，对电信用户进行分群并进行行为分析。

最后通过不同簇的统计分析分析群体的行为特点。以后可以邀请行业专家进一步对不同群体数据进行更加细致的分析和打标签，从而实现客户群体画像，推广到实际营销或者客服等商业活动中。