**问题分析：**

**问题三：**

在问题三中，我们需要建立会员生命周期和状态划分的模型。通过查阅相关理论资料，我们确定了会员生命周期一般可划分为四个阶段：成长期-成熟期-休眠期-流失期。于是，可以通过使用RFMS模型，并选取其中R（即最近没有消费行为天数）和F(总消费次数),来作为划分会员活跃状态的指标。以此，来模拟会员生命周期。

**问题四：**

在问题四中，我们需要从实际销售数据出发，建立确定激活率和商场促销活动的关系模型。由于附件当中，2017年的数据比较完整，数量较多，且时效性比较高，于是，考虑分析2017年每月会员的活跃状态与月销售额之间的关系。其中，定义每个月活跃会员比例的增减为激活率，并观察假设会员的激活率与月销售额呈强线性关系。于是，通过最小二乘法进行线性拟合，最终得到激活率与月销售额的关系式。

**模型的建立与求解：**

**问题三：**

根据题目的要求，我们把问题三中，会员的生命周期定义为会员从入会到退出的整个过程。会员的状态，定义为会员在商场的消费是否活跃。由于题目的要求相对模糊，我们具体地对上述概念作如下定义：

首先，根据RF指标将会员生命周期一般划分为四个阶段：成长期-成熟期-休眠期-流失期。其中：

成长期：最近没有消费行为天数在30天及以内，消费1到10次；

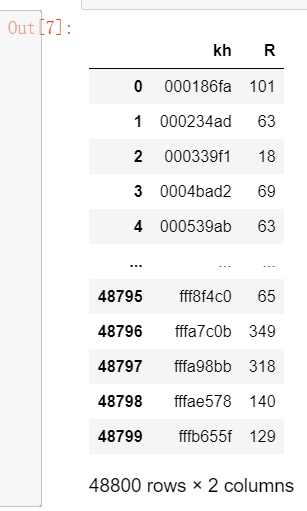
成熟期：最近没有消费行为天数在30天及以内，消费10次及以上；

休眠期：31-90 天没有消费行为；

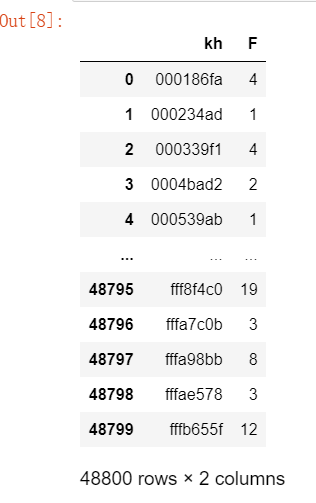
流失期：90 天以上没有消费行为。

其次，我们可以将处于成长期或成熟期的会员归类为活跃会员，并将休眠期和流失期的会员归类为非活跃会员：

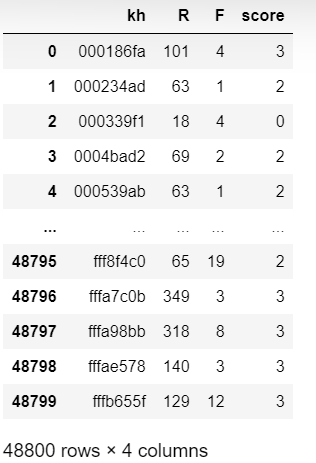
确定了模型定义后，我们提取‘附件3-会员消费明细表’中的数据，并进行预处理。考虑到15、16年的会员数据时效性较差，于是将他们删除，只考虑17，18年的客户数据。然后，将表中的dtime（即消费时间）具体到日，并与数据收集截止日期2018.1.3相减，得到每笔消费距今天数，再通过‘kh’（即会员卡号）groupby归类取最小值得到R（即最近没有消费行为天数），如下图所示：



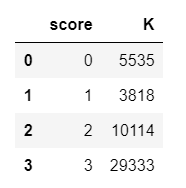
同理，我们依然使用groupby，统计每个会员的消费次数，计数以得到F：



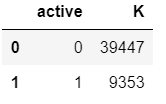
将两张表合并，并通过上述定义的规则，对每个会员的状态进行评分（成长期-0，成熟期-1，休眠期-2，流失期-4），如下图所示：



之后继续使用groupby对不同评分的会员，进行归类并统计计数，得到每种状态会员的人数，如下图所示：



同时，可以依据上文提到的规则，对处于不同状态的会员进行归类并统计计数，如下图所示：



其中，‘active’中的0表示非活跃会员，1表示活跃会员。此举也为问题四提供了重要的数据。

在上述结论的基础上，我们还可以考虑研究会员状态与入会时长的关系，考虑用户自然流失的情况。为此，我们读取了‘附件1-会员信息表’中的数据，并提取其中的‘kh’（即会员卡号）和‘djsj’（即会员注册时间）两列，以‘kh’为索引，与上表合并。

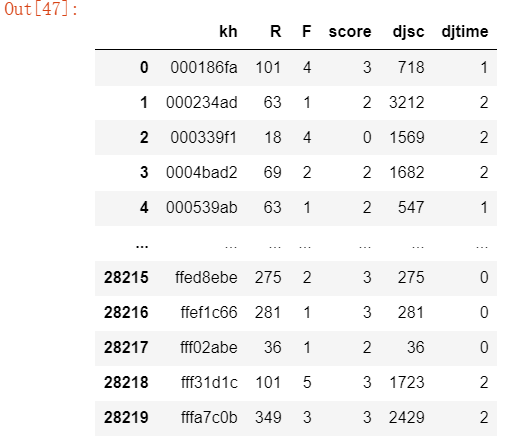
同时，再次定义：

新用户为注册时长1年以内的会员

成熟用户为注册时长1-3年的会员

老用户为注册时长3年以上的会员

通过预处理得到如下表格：



其中‘djtime’中的0表示新用户，1表示成熟用户，2表示老用户。对不同入会时长的客户归类，计算每种用户的活跃程度占比，可得到如下表格：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 成长期会员占比 | 成熟期会员占比 | 休眠期会员占比 | 流失期会员占比 |
| 新会员 | 12.20% | 5.20% | 17.51% | 65.09% |
| 成熟会员 | 10.06% | 16.34% | 21.18% | 52.42% |
| 老会员 | 11.19% | 9.71% | 21.34% | 57.76% |

此表反应了会员入会时长与活跃程度的关系，表现了会员的生命周期。

**问题四：**

依据上一问中得到的会员活跃程度，我们用如下的方式来定义会员的激活率：即本月的活跃用户占比-上月的活跃用户占比。

利用这一定义，我们选取数据样本最多，最全面也最具有时效性的2017年的会员消费数据。根据月为索引，将该月消费过的所有会员的活跃度进行统计。

可以得到以下数据：

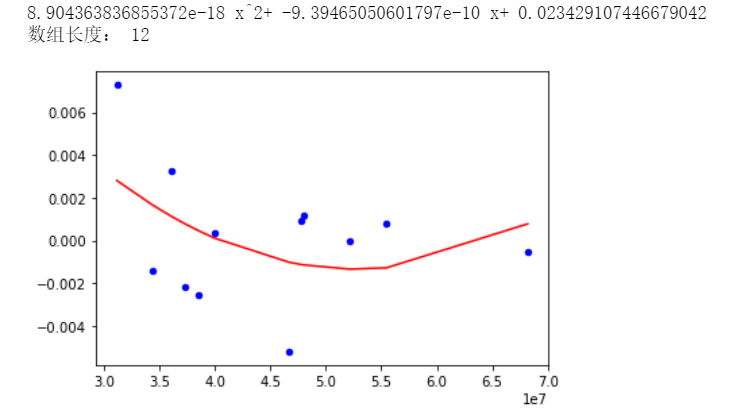
|  |  |  |
| --- | --- | --- |
| 月份 | 活跃会员占比 | 激活率 |
| 1 | 3.45% |  |
| 2 | 4.18% | 0.73% |
| 3 | 3.66% | -0.52% |
| 4 | 3.98% | 0.32% |
| 5 | 4.02% | 0.04% |
| 6 | 3.87% | -0.14% |
| 7 | 3.65% | -0.22% |
| 8 | 3.74% | 0.09% |
| 9 | 3.82% | 0.08% |
| 10 | 3.56% | -0.26% |
| 11 | 3.51% | -0.05% |
| 12 | 3.63% | 0.12% |

在此基础上，我们再将其与商场本月的销售额结合在一起：

|  |  |  |
| --- | --- | --- |
| 月份 | 销售额 | 激活率 |
| 1 | 52181562.48 |  |
| 2 | 31155181.90 | 0.73% |
| 3 | 46645196.29 | -0.52% |
| 4 | 36062148.54 | 0.32% |
| 5 | 39954229.70 | 0.04% |
| 6 | 34413449.26, | -0.14% |
| 7 | 37235751.12 | -0.22% |
| 8 | 47777679.92 | 0.09% |
| 9 | 55465081.33 | 0.08% |
| 10 | 38459753.10 | -0.26% |
| 11 | 68200820.82 | -0.05% |
| 12 | 47981021.90 | 0.12% |

观察得知，在销售额较高，即商场有促销的月份，会员的激活率就高，就越有可能表现得活跃，这两者之间有一定的线性关系，于是，我们对数据进行线性拟合：

得到如下的公式：



从而，得到了商场促销活动与激活率之间的关系。