

数据科学“云实训”项目训练营

第六课：精准营销应用

讲师：高扬



—— 数据酷客 ——



数据科学人工智能

项目报告

要求：闯关模式下完成实训任务 + 实训报告

截止时间：2020/08/14 23: 59: 59

报告提交形式：登录实训平台，完成项目主、客观实训任务，生成报告，提交至“数据科学精准营销训练营”小组（组号：109）

结业考试



要求：参加在线考试

考试时间：2020/08/07 20: 00 - 21: 00

考试地点：数据酷客-大讲堂-云教室316 在线答题

考试形式：单项选择题+多项选择题，30道；每题分值为1分



证书（可选）



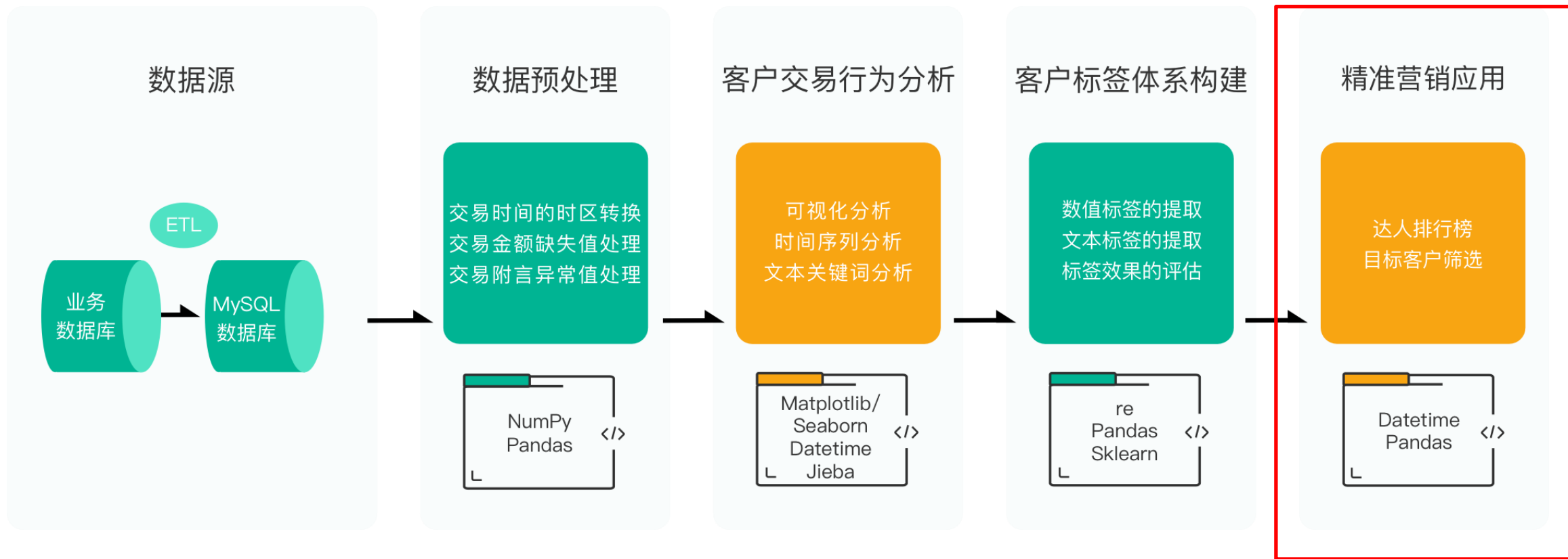
要求：参加全部6次直播 + 提交项目报告+ 结业考试

实训教学研讨群

总成绩 = 项目报告成绩 + 在线考试成绩

项目报告评审，满分70分；参与在线考试，考试满分30分；合格标准：两项成绩总和>=60分

成绩公布时间：2020/08/21；08.22日开始陆续发放证书



创建人群-标签人群

[查看建包指南](#) [查看标签说明](#)

人群包名称: *

有效时间: 2020-05-11 ~ 2020-05-14 *

人群包用途: MP消息平台 * ?

更新频率: ☒ 一次性任务 ☐ 每日更新 ?

人群包描述:

用户忠诚度

居住地所在区

工作地所在区

居住地所在商圈

工作地所在商圈

海外最活跃国家

最后一次浏览时间

租赁最活跃业务城市

所有业务最活跃业务城市

▶ 用户生命周期价值

① 数量预估功能暂不支持首次活跃距今天数、末次活跃距今天数、朝向标签，以及业务国家、业务城市取不等于的情况。即将支持，敬请期待！

人群包预估条数: 3288 条

标签组

贝壳app安装 ① 等于 已安装 x

且

工作地所在... ① 等于 北京市-惠新西街 x

拖入标签

引子：人群圈定

链接：无需填写链接，系统自动填充
人群包：[\[可用\]小贝营销0509](#)

分类：精准导购推送 任务类型：一次性任务 周期：
预计发送时间：2020-05-09 12:00:52 有效期：

*预览：
请填写手机号，每行一条
[点击预览](#) [填充上次预览手机号](#)

当前版本任务进度

创建任务
人群包2020-05-08 17:33:08

任务准备中
人群包可用
[去我的人群查看详情](#)

待发送
2020-05-09 12:01:32
预计发送2029871

历史版本执行情况

版本日期	实际发送时间	状态	预计发送	实际发送	送达	点击	商机对数
2020-05-09	2020-05-09 12:01:32	成功	2029871	1396877	1355922	6918	42

DMP									
人群洞察 我的人群 标签后台 消息推送									
C端推送									
push									
公众号消息									
自定义管理									
动态文案									
落地页									
push任务列表									
进入AB实验 导出数据 新建push任务 <input type="text" value="请输入"/>									
全部 5045 准备中 63 已完成 4739 已过期 184 已停用 83									
任务ID	任务名称	APP	任务类型	任务分类	运行状态	可用状态	创建时间	人群包	操作
9712	0511精准	贝壳	一次性任务	精准导购推送	● 准备中 (未到发送时间)	● 启用	2020-05-11 17:42:24	[可用]贝0511精	分析 编辑
9711	济南0511_复制	链家	一次性任务	无分类	● 待发送	● 启用	2020-05-11 17:33:36	[可用]C链0426	分析 编辑
9710	济南0511	贝壳	一次性任务	无分类	● 待发送	● 启用	2020-05-11 17:32:49	[可用]贝0416济	分析 编辑

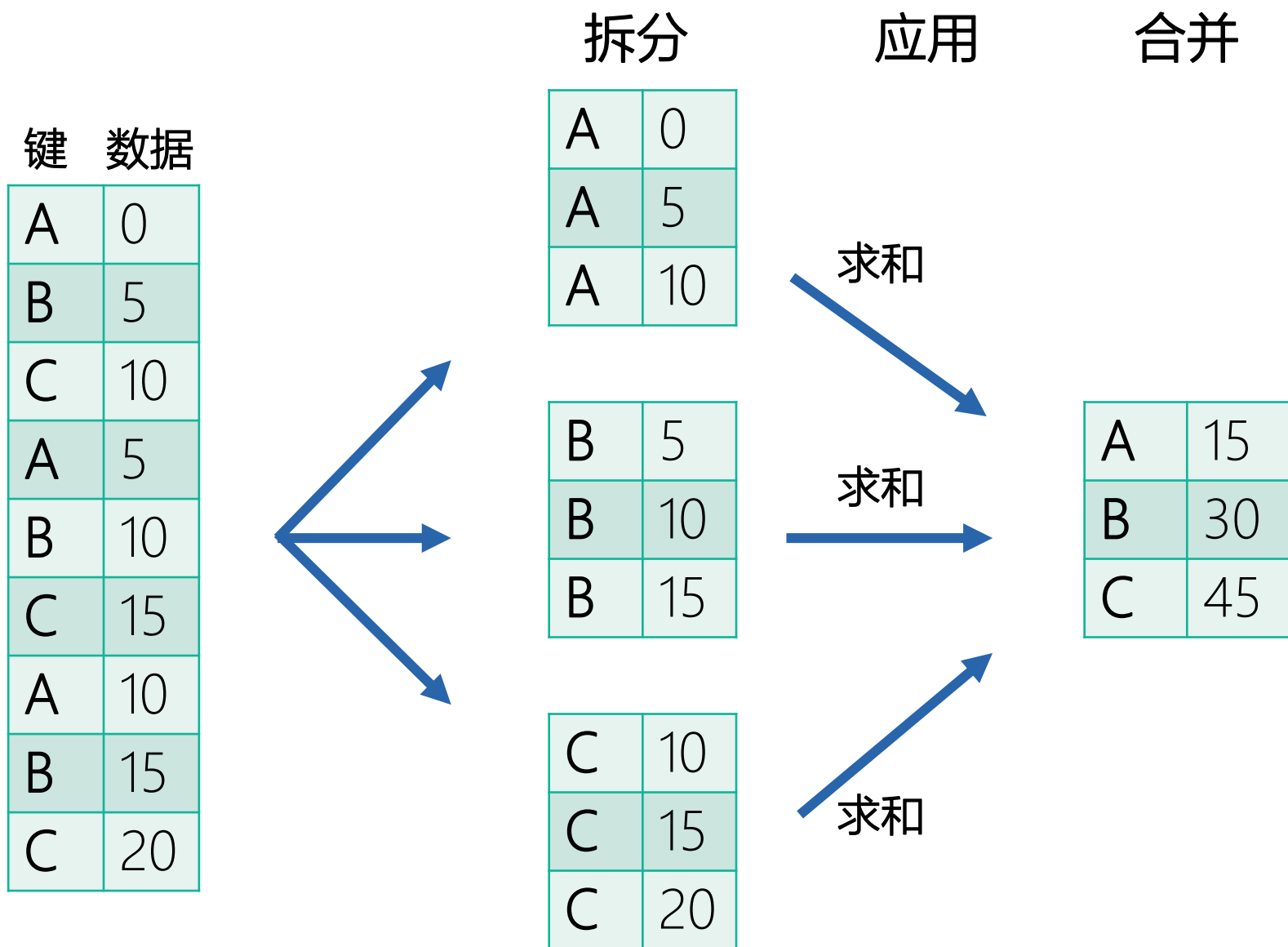
本部分主要包括两大部分：

- 以彩票类消费为例，构建客户的商品兴趣度排行榜
- 以推销信用卡为例，筛选满足固定特点的客户

user_id	time_penalty	payment_sum	tfd_df_sum	final_score
544518	0.993610	1.000000	0.487399	2.481008
22321749	1.000000	0.521733	0.600132	2.121865
17488993	0.999996	0.310164	0.750119	2.060280
19199845	0.999999	0.009039	0.978040	1.987078
15001742	0.981358	0.004678	1.000000	1.986036
12947427	0.992627	0.026958	0.956306	1.975891
11957177	1.000000	0.080297	0.881162	1.961460
11012971	1.000000	0.069979	0.880869	1.950848
3382860	1.000000	0.293317	0.654995	1.948311
8878476	0.999948	0.011284	0.933472	1.944704



- Pandas提供了灵活高效的GroupBy技术
- GroupBy技术是对数据进行分组计算并将各组计算结果合并的一项技术，包括三个过程：
 - 拆分（Splitting）：即数据分组
 - 应用（Applying）：对每组应用函数进行计算
 - 合并（Combining）：将计算结果进行数据聚合



- **商品兴趣度排行榜的计算**
- 目标客户筛选

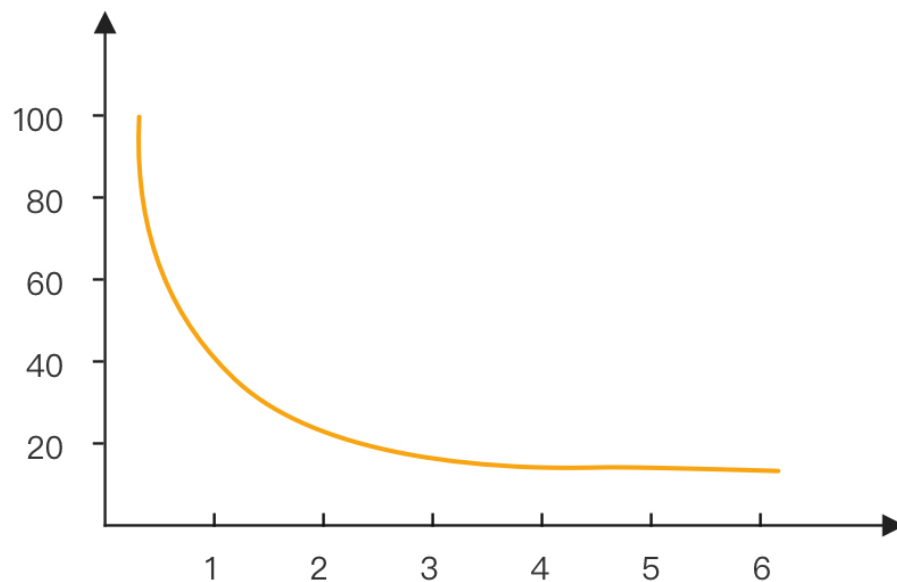
计算商品兴趣度排行榜时，参考时间衰减、**tf-idf**文本权重和消费金额三个指标：

- 在计算时间衰减兴趣度时，参考艾宾浩斯遗忘曲线，客户消费时间越近则兴趣度越高
- 在计算**tf-idf**文本权重时，参考客户文本标签的**tf-idf**值，权重越大兴趣度越高
- 在计算消费金额兴趣度时，参考客户的消费金额指标，消费数量越大兴趣度越高

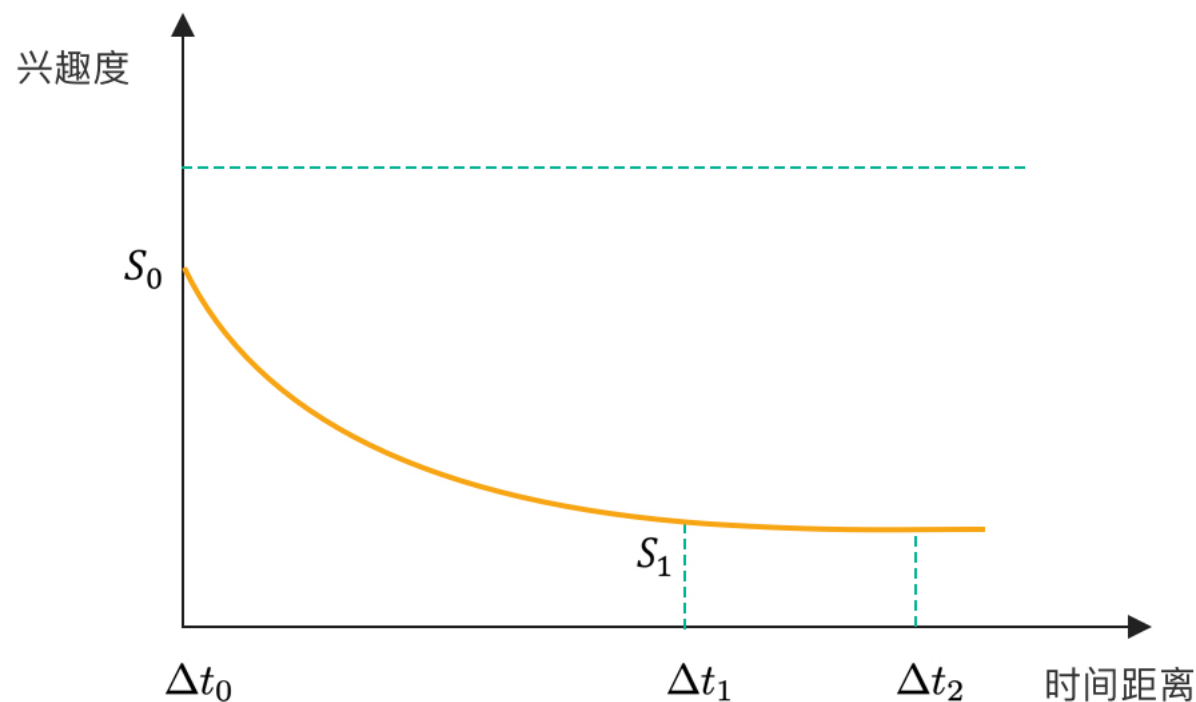
对三种兴趣度进行归一化和加权，综合计算总得分

- 艾宾浩斯遗忘曲线：在学习中的遗忘是有规律的，遗忘的进程很快，并且先快后慢
- 在分析客户对电商平台的商品兴趣偏好时，也可以借鉴遗忘曲线

记忆的数量（百分数）



如果客户在观察点之前多个时刻 (t_0, t_1, t_2) 购买某类商品, 则每个时刻与观察点的距离为 $\Delta t_0, \Delta t_1, \Delta t_2$
对于该商品的兴趣度衰变曲线如下图所示:



每个时刻兴趣度 S_t 的计算公式为: $S_t = e^{-\eta \Delta t}$

+ Δt 表示消费时刻 t 距离观察点的月数

+ η 表示衰减因子

考虑到客户会出现多次消费行为(默认每次消费的衰减因子一致), 对某商品的兴趣总和公式如下:

$$S = \sum e^{-\eta \Delta t}$$

假定此处为彩票类商品消费, 我们设定衰减因子 η 为0.5, 我们可以快速计算出每个时刻的兴趣度, 将每个时刻的兴趣度相加, 得到总的基于时间的商品兴趣度。

- 彩票类消费关键词：['停彩','大乐透','双色球','福利彩票','彩票','竞彩','追号']
- 将每个客户这些关键词的tf-idf值进行加和，即为tf-idf兴趣度

- 彩票类消费关键词：['停彩','大乐透','双色球','福利彩票','彩票','竞彩','追号']
- 匹配出每个客户的彩票类消费记录，计算消费总金额

对时间衰减兴趣度采用Sigmoid归一化，对tf-idf兴趣度和消费金额兴趣度采用Min-max归一化

Sigmoid归一化

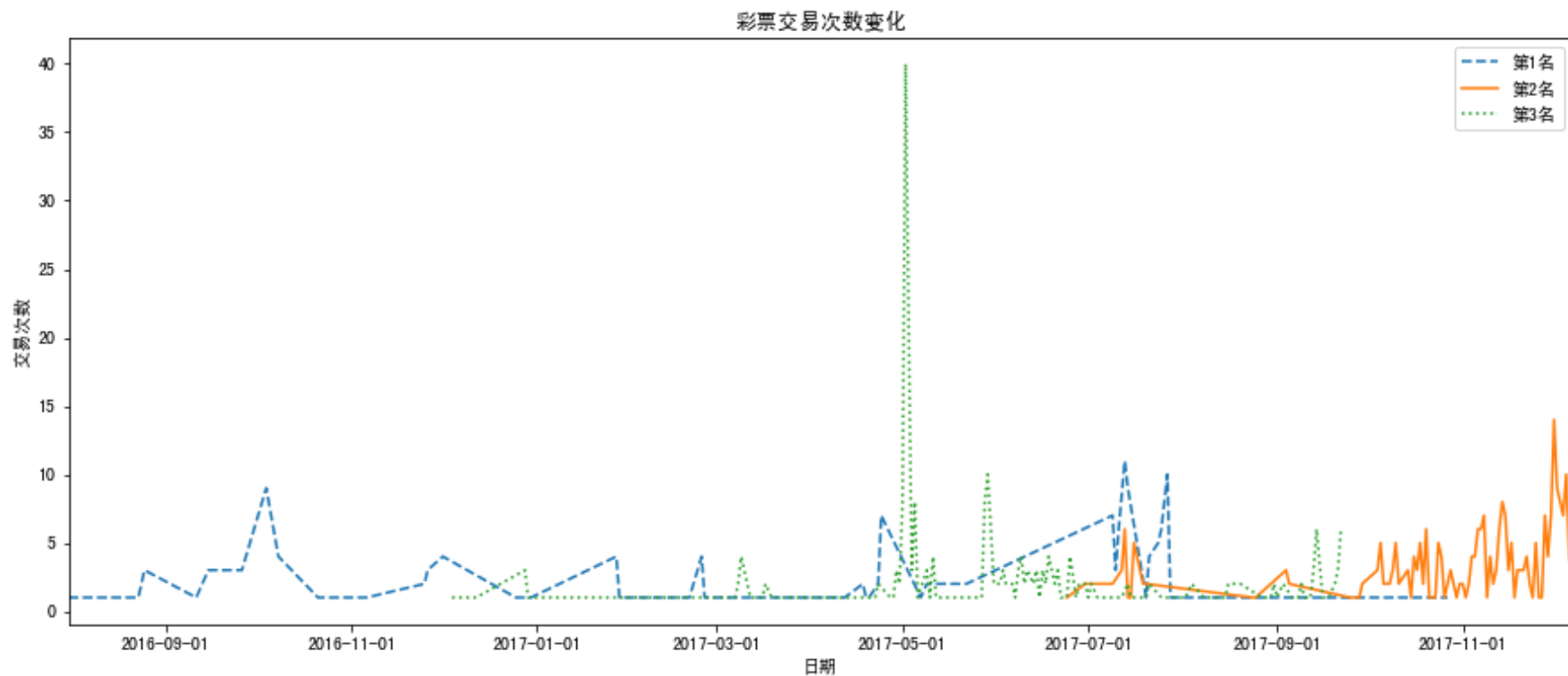
$$s(x) = \frac{1}{1+e^{-x}}$$

Min-max归一化

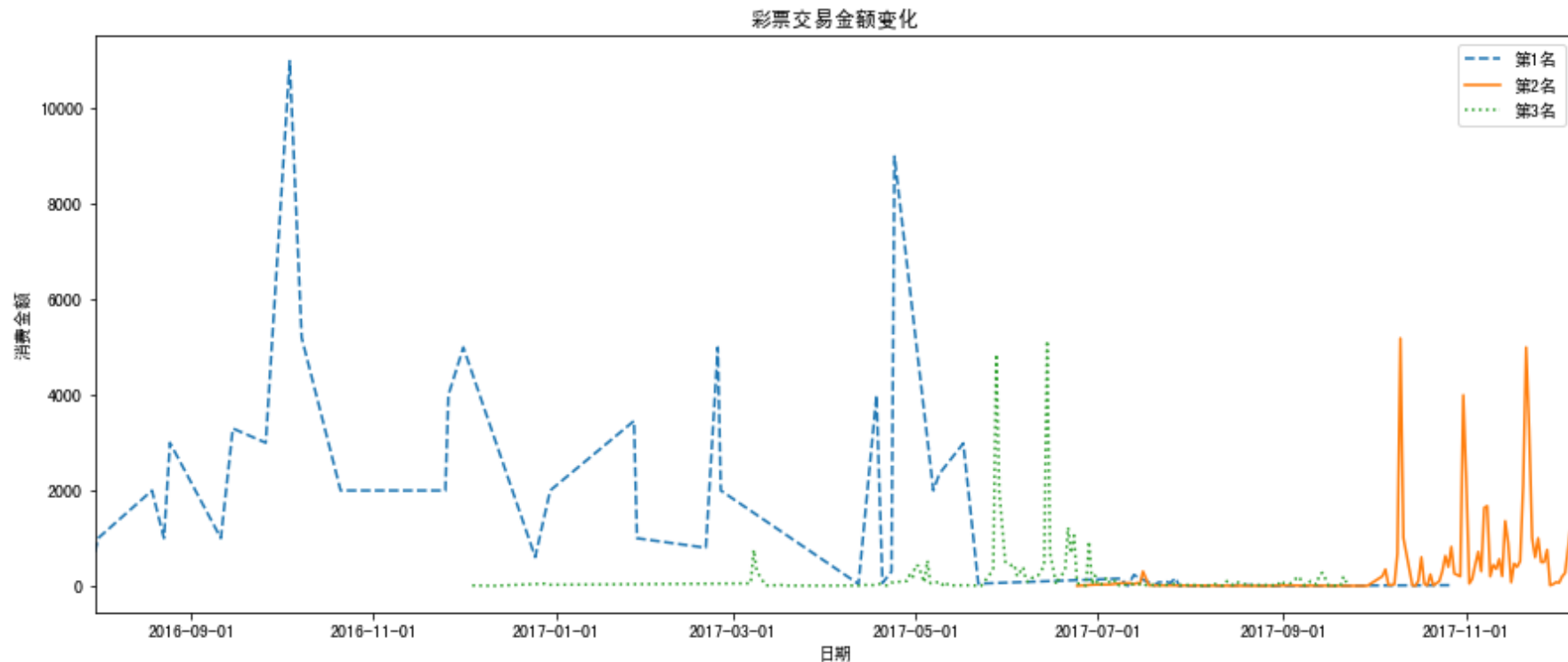
$$s(x) = \frac{x-x.min()}{x.max()-x.min()}$$

将三个商品兴趣度相加计算总兴趣度

user_id	time_penalty	payment_sum	tfidf_sum	final_score
544518	0.993610	1.000000	0.487399	2.481008
22321749	1.000000	0.521733	0.600132	2.121865
17488993	0.999996	0.310164	0.750119	2.060280
19199845	0.999999	0.009039	0.978040	1.987078
15001742	0.981358	0.004678	1.000000	1.986036
12947427	0.992627	0.026958	0.956306	1.975891
11957177	1.000000	0.080297	0.881162	1.961460
11012971	1.000000	0.069979	0.880869	1.950848
3382860	1.000000	0.293317	0.654995	1.948311
8878476	0.999948	0.011284	0.933472	1.944704



- 排名第一的客户彩票类消费时间主要集中在2016年9月-2017年9月，之后的消费出现停滞
- 排名第二的客户彩票类消费距离观察点很近
- 排名第三的客户在2017年5月出现一个交易次数的峰值



- 第一名在时间维度和消费次数上都不占优势，但是明显可以看出该客户的消费金额远远高于第二和第三名在2016年10月的一次彩票类消费超过了10000元，可以看出这是一名狂热的彩票爱好者
- 第二名彩票类的消费次数远少于第三名，且消费金额基本持平，但在时间维度上占据很大的优势，综合得分上超过第三名


- 商品兴趣度排行榜的计算
- 目标客户筛选

营销人员根据实际的业务需求，针对性的选取几个客户的关键特点，从数据库中精准查询到符合需求的目标客户

如我们要推销某出行app联名的分期信用卡，为目标客户推送广告，根据人为经验锁定目标的特征为：

- 非休眠客户
- 有商旅消费的记录
- 有信用卡的使用记录
- 有消费分期的需求(使用记录)

```
SELECT user_id  
FROM user_features  
WHERE business_travel_cnt > 0  
      AND credit_card_repay_cnt > 0  
      AND is_installment = 1  
      AND sleep_customers = 0  
ORDER BY credit_card_repay_cnt DESC  
LIMIT 200
```



	user_id	business_travel_cnt	credit_card_repay_cnt
0	10877064	1	306
1	25628740	3	296
2	32121673	9	271
3	18775289	2	259
4	11950066	14	245
5	28024923	17	239
6	24884096	17	238
7	23585613	12	236
8	31561502	28	231
9	12943621	33	207



数据酷客



数据科学人工智能



加入数据酷客交流群