

背景

由于半个月后运营方将会推广一个大型的补贴活动，希望对用户进行针对性地补贴激励，做好精准营销。

-**核心目标：**基于订单数据给用户打标签构建用户画像标签体系，进一步对用户进行价值分层，并对各层用户制定对应的运营方案。

-**数据口径：**20200401-20200731 所有订单数据

核心成果

1. 用户画像标签体系表

已经成功构建用户画像标签体系表，目前可直接筛选表单数据来圈定目标用户。后续与 CRM 系统打通后，运营方可通过前台筛选标签类别来挑选对应的后台用户进行发券。

2. 五类价值用户分群

基于目前全量用户的 **R**（最近一次消费时间距今时长）、**F**（消费频次）、**M**（消费金额）三个指标结合 Kmeans 聚类算法可以得到 5 类价值用户：

用户类别	R	F	M
一般挽留用户	2 个月	1-2 次	33 元
一般发展用户	半个月	2-3 次	29 元
重要发展用户	一周	13 次	231 元
重要保持用户	7 天	34 次	694 元
重要价值用户	4 天内	56 次	3143 元

- **第一类：R 最大，F 最小，M 偏小，定义为一般挽留用户。**此类用户已经 2 个月没有来消费了，并且历史平均消费频次只有 1-2 次，历史平均消费金额 33 元。
- **第二类：R 偏大，F 偏小，M 最小，定义为一般发展用户。**此类用户已经半个月没有来消费了，并且历史平均消费频次只有 2-3 次，历史平均消费金额 29 元。
- **第三类：R 中等，F 中等，M 中等，定义为重要发展用户。**此类用户已经一周没有来消费了，并且历史平均消费频次有 13 次，历史平均消费金额 231 元。
- **第四类：R 偏小，F 偏大，M 偏大，定义为重要保持用户。**此类用户 7 天内刚消费过，并且历史平均消费频次有 34 次，历史平均消费金额 694 元。
- **第五类：R 最小，F 最大，M 最大，定义为重要价值用户。**此类用户 4 天内刚消费过，并且历史平均消费频次有 56 次，历史平均消费金额 3143 元。

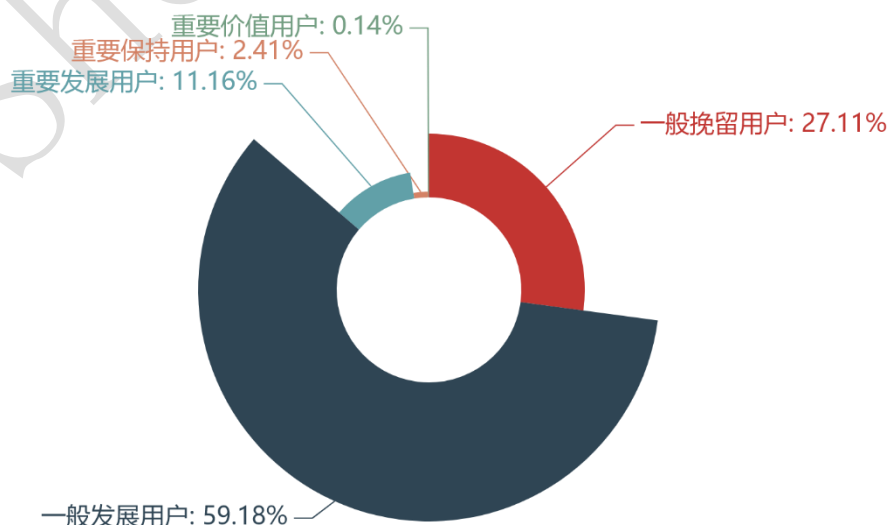


图 1 五类用户比例分布图

建议

用户类别	建议
一般挽留用户	流失风险大，急需流失召回
一般发展用户	挖掘需求，提升频次
重要发展用户	提升频次，重在建立忠诚度
重要保持用户	用户回流，培养成重要价值客户
重要价值用户	提供差异化商品，可保持现状

数据分析推理过程

1. 取数

hive SQL 查询 20200401-20200731 订单数据

2. 打标签

从用户基本特征、历史消费特征、分时间段消费特征三个维度给用户打标签，最终形成包含 27 个字段的用户画像标签体系表。后期建立 hive 底层表供 CRM 系统调用。

3. 提取标签表中的 R、F、M 数据进行 Kmeans 聚类模型构建

- 首先查看 R、F、M 的数据偏斜程度，利用四分位差判断离群点，发现不存在离群值。
- 然后依次迭代 $K=1, 2, 3\cdots, 10$ 情况下模型的 SSE 和轮廓系数，基于手肘法和轮廓系数越大越好的判断标准选取出最合理的 K

值。

- 最后计算被分为 K 类后的各类的 R、F、M 值以及每一类所占的比例。

4. 给予业务解释及可视化

针对上一步中分为 K 类的数据结合实际业务情况赋予合理的业务解释，并且针对每一类用户价值特征提出对应的运营策略。利用表格展示每类价值用户的特征及建议业务方采取的措施，以饼图展示每类价值用户的分布比例。