

寒假练习项目

百货商场用户画像 描绘与项目分析

[文档副标题]

豆奶

2020-3-17

目录

- 1. 背景简介..... 2
 - 1.1 行业背景..... 2
 - 1.2 项目目标..... 2
- 2. 数据说明与预处理..... 2
 - 2.1 数据介绍..... 2
 - 2.2 数据预处理..... 3
 - 2.3 数据合并..... 3
- 3. 统计分析..... 3
 - 3.1 分析会员基本信息..... 3
 - 3.2 会员消费情况分析..... 4
 - 3.3 会员不同时间段的消费偏好..... 5
- 4. 用户画像与促销策略分析..... 6
 - 4.1 构建用户标签体系（用户画像）..... 6
 - 4.2 会员用户细分及营销策略制定..... 7
 - 4.3 关于个性化商品推荐..... 8
- 5. 结语..... 9

1.背景简介

1.1 行业背景

近年来，随着新零售业的快速发展，消费者在购买商品时有了更多的对比和选择，导致超市行业的竞争日益激烈，利润空间不断压缩。超市的经验管理产生了大量数据，对这些数据进行分析，可以提升超市的竞争力，为超市的运营及经营策略调整提供重要依据。

1.2 项目目标

本文主要利用提供的商场数据，针对商品销售数据的特点：

- （1）结合业务对数据进行探索与预处理
- （2）统计分析，探究商场会员的数据并进行可视化处理
- （3）绘制用户画像
- （4）对会员进行细分并针对不同群体制定对应的营销方案，从而提升销售利润

2. 数据说明与预处理

2.1 数据介绍

数据是某商场 2018 年的经营数据，相关数据存于附件中。附件中存有两个表，表 1（cumcm2018c1.csv）是商场会员的信息，包含会员卡号、出生日期、性别、会员入会时间；表 2（cumcm2018c2.csv）是商场顾客购物的具体信息，包含会员卡号、消费产生的时间、商品编码、销售数量、商品售价、消费金额、商品名称、本次消费得到的积分、收银机号、单据号、柜组编码、柜组名等信息。

	kh	csrq	xb	djsj
0	c68b20b4	2002-11-02 00:00:00	0.0	2013-05-11 00:00:00.000
1	1ca15332	NaN	0.0	2004-11-04 16:31:52.436
2	a37cc182	1967-02-17 00:00:00	0.0	2004-12-31 21:24:34.216
3	2ab88539	1982-06-01 00:00:00	0.0	2010-11-19 00:00:00.000
4	b4c77269	1964-02-05 00:00:00	0.0	2007-12-14 00:00:00.000

图 1：表 cumcm2018c1 展示图

	kh	dtime	spbm	sl	sj	je	spmc	jf	syjh	djh	gzbm	gzmc
0	1be1e3fe	2015-01-01 00:05:41.593	f09c9303	1	290.0	270.20	兰芝化妆品正价瓶	270.20	6	25bb	8077.0	兰芝柜
1	1be1e3fe	2015-01-01 00:05:41.593	f09c9303	1	325.0	302.80	兰芝化妆品正价瓶	302.80	6	25bb	8077.0	兰芝柜
2	1be1e3fe	2015-01-01 00:05:41.593	f09c9303	1	195.0	181.80	兰芝化妆品正价瓶	181.80	6	25bb	8077.0	兰芝柜
3	1be1e3fe	2015-01-01 00:05:41.593	f09c9303	1	270.0	251.55	兰芝化妆品正价瓶	251.55	6	25bb	8077.0	兰芝柜
4	1be1e3fe	2015-01-01 00:05:41.593	f09c9303	2	245.0	456.55	兰芝化妆品正价瓶	456.55	6	25bb	8077.0	兰芝柜

图 2：表 cumcm2018c2 展示图

2.2 数据预处理

本次数据预处理主要包含以下步骤（实现代码见 task1. ipynb）：

2.2.1 表 1 会员信息表

- 1) 对会员号(kh)进行去除重复值操作得到无重复值的会员信息
- 2) 缺失值填充，这里采用向下填充
- 3) 异常值处理，如年龄不在 0~100 之间的，会员卡号重复的行等，可以进行去除处理。

2.2.2 表 2 销售流水表

- 1) 去除表中的重复值
- 2) 缺失值填充，这里采用向下填充
- 3) 异常值处理，如订单时间为某年 2 月 29 日（该年非闰年）等。

2.3 数据合并

利用外连接，以会员卡号(kh)为键值将表 1 的会员信息表和表 2 的销售流水表合并。

3.统计分析

3.1 分析会员基本信息

统计商场会员的年龄组成，分为青年(44 岁及以下)、中年（45-59）和老年（60 以上），同时，统计商场会员的男女比例，得到如下图：

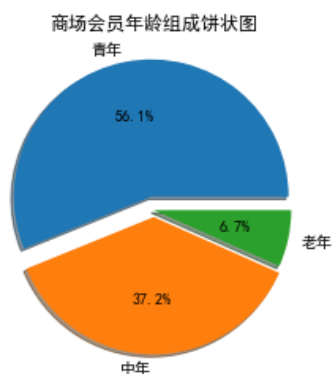


图 3：会员年龄组成图

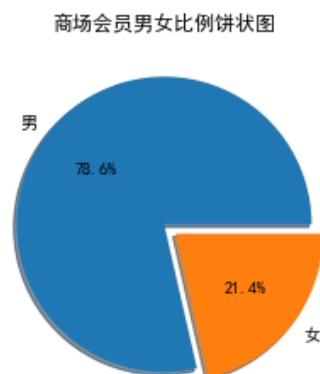


图 4：会员男女比例图

从上图可以看出，该商场的会员青年居多，而在性别比例上，男性会员占绝大部分。接下来，继续对他们的消费情况进行分析：

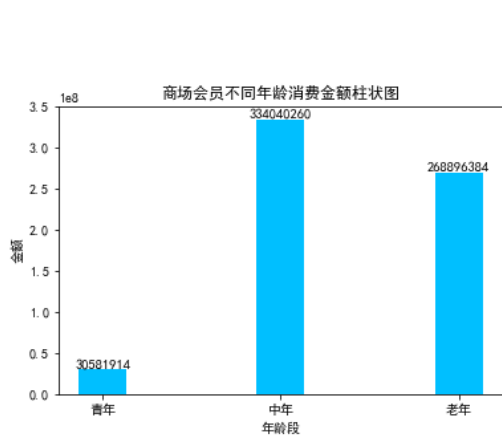


图 5: 会员不同年龄消费金额图

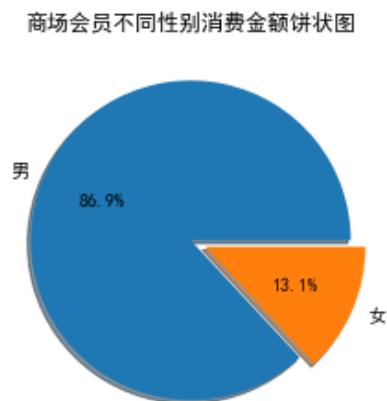


图 6: 会员性别消费比例图

从图 5 和图 6 可以看出，中年会员和老年会员的消费占据了绝大部分，而青年会员的消费则较少，这可能是因为青年比较多玩手机，更倾向于网上购物，因此为了吸引青年顾客，可以尝试开放在线购物平台，以促进消费。而继续分析会员中男女顾客的消费情况，可以看出男性消费占了很多，而女性消费比例很少。这可能是因为，男女一起出来，是男性掏钱吧。

3.2 会员消费情况分析

通过分析会员与非会员的消费情况，可以有针对的对他们提出不同的促销建议。下面通过统计单据号字段对会员和非会员的订单数进行统计，同时，也对两者的消费金额进行对比，得到下图：

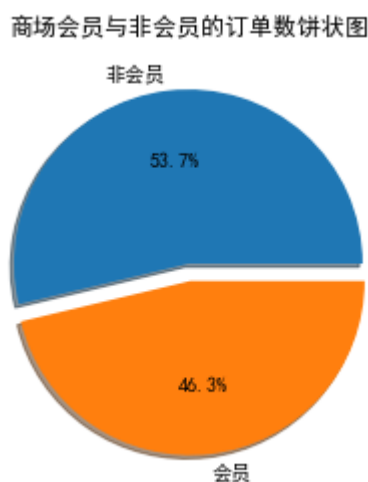


图 7: 两类顾客订单数饼状图

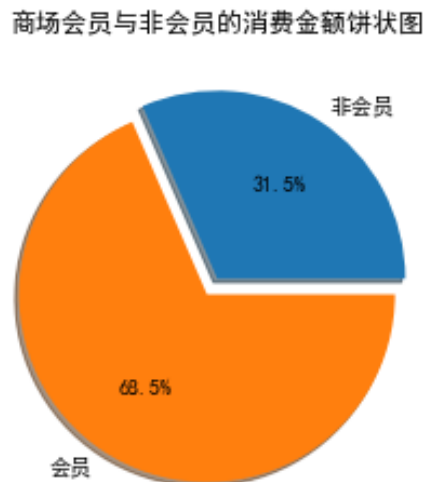


图 8: 两类顾客消费比例图

通过图 7 可以看出，会员的订单数是比较少的，因为一般来说，会员的数量应该远比非会员的少。因此，应该结合他们的消费金额（如图 8）来看，会员的消费总额是比较多的，这可能是因为会员大多是常客，而非会员可能是路过或者偶然购买。

3.3 会员不同时间段的消费偏好

分别以季度和小时为单位，分析不同时间段会员的消费时间偏好。一方面，这样可以分析会员群体在一年当中的消费情况，得出消费旺季和淡季；另一方面，分析会员群体在一天当中的消费分布，得到消费高峰期和低峰期。从而使得超市管理人员可以在消费高峰期和旺季多安排工作人员，在低峰期减少安排工作人员，以减少支出。

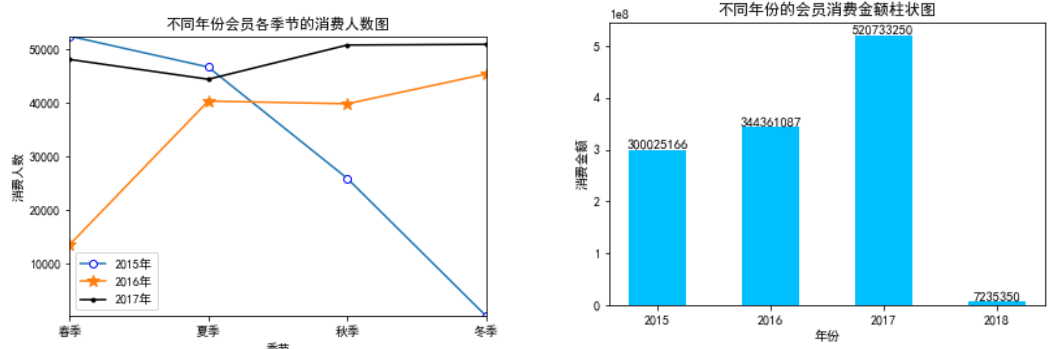


图 9：不同年份会员各季节消费人数图 图 10：不同年份消费柱状图

由图 9、图 10 可以看出，随着时间不断推移，商场每年的销售量都有所提高（2018 年数据是不全的），该商场的旺季是夏季，应该增派人手，而另外三个季节，在最近一年的销售金额也偏高，说明该商场最近一年销售稳定，生意好。

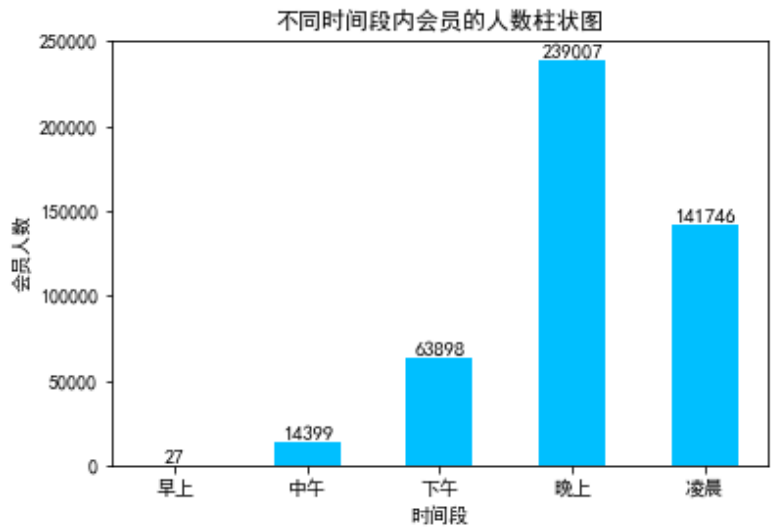


图 11 不同时间段会员消费人数图

通过分析一天内不同时间段的消费情况，可以对不同时间段的商场人手进行分配，本文定义：早上（6 点-11 点）、中午（11 点-13 点）、下午（13 点-18 点）、晚上（18 点-1 点）、凌晨（1 点-6 点）。由图 11 可以看出，晚上是高峰期，这可能是由于上班的人都下班了有时间进行购物，因此，需要在晚上的时间段加派人手，以保证服务到位。

4.用户画像与促销策略分析

4.1 构建用户标签体系（用户画像）

标签系统是用户画像的核心，标签化是对用户特征的符号表示，标签化用户画像，既方便计算机进行计算分析，又方便人们对用户画像的理解。

结合业务和数据本身，并设身处地的站在甲方的角度进行思考——用户哪方面的信息更有价值。笔者设计了两级用户标签分类体系，其中一级标签为用户基本特征、业务特征和偏好特征标签。具体如下表所示：

表 1：用户标签系统

一级标签	二级标签
用户基本特征	性别
	年龄
	入会时长
用户业务特征	会员积分
	消费水平
	消费次数
用户偏好特征	购物时间偏好
	购物季节偏好
	购物类型偏好

借助第三部分的统计分析，不难实现这一标签系统。汇聚上面三类标签，建立得到用户画像（如下图 所示）。其中，由于提供的商品数据缺乏分类，仅有商品名称，因此，购物类型偏好暂不进行。

	卡号	积分	消费水平	商品名称	性别	入会时长	年龄	购物时间偏好	购物季节偏好	消费次数
0	000186fa	5267.0	高消费	[雅诗兰黛肌初赋活原液, 雅诗兰黛密集焕白遮瑕笔, 雅诗兰黛晶透沁白淡斑精粹水, 雅...	女	4	43.0	下午	春	4
1	000234ad	11850.0	高消费	[wacoal 正价\件, 和音E, JuicyCouFureF件, 雅诗兰黛白金级奢宠紧...	女	11	45.0	下午	冬	7
2	000339f1	6141.0	高消费	[wacoal 正价\件, 植村秀眉笔4g, 植村秀汗杆/汗芯, 植村秀无色限彩妆盒四格....	女	7	32.0	下午	春	10
3	0004bad2	8964.0	高消费	[朗姿A件]	女	7	36.0	下午	秋	1
4	00075d60	32030.0	高消费	[DOUBLE SFANDARDF, 希思黎全日呵护精华乳, 希思黎焕白亮采面膜, 希...	女	4	52.0	下午	春	9

图 12：用户标签表

卡号：00075d60
积分：32030.0
消费水平：高消费
性别：0.0
入会时长：4
年龄：52.0
购物时间偏好：下午
购物季节偏好：春
消费次数：9

图 13：用户画像查询示例图

4.2 会员用户细分及营销策略制定

对会员用户进行精细划分，有助于分析不同群体带来的价值差异。由于原始数据集中的特征属性较多，从表中选取一些特征来进行分析。RFM 模型是衡量客户价值和客户创利能力的重要攻击和手段，在众多的客户关系管理的分析模式中，RFM 模型是被广泛提到的。其中 RFM 分别指代一个客户的近期购买行为(Recency)、购买的总体频率(Frequency)以及在一定时间内的总花费(Monetary)。因此，本文将从 2017 年一年的数据中提取这三项指标来描绘该客户的价值状况，即

- 最近一次消费(Recency)：这里将最近一次购买时间与提数日做减法（本文选择 2017 年 1 月 1 日为提数日）。
- 消费频率(Frequency)：即顾客在统计周期内（一年内）购买商品的次数。
- 消费金额(Monetary)：即顾客在统计周期（一年）内的消费总金额，体现的是顾客为商场创利的多少。

然而，RFM 三个值之间存在量级之间的差距，无法直接通过加减来衡量用户价值。因此这里采用算法来为三个值进行评分，评分区间为 1-5 分，对于 F 和 M 都是越大分值越高，需要注意的是 R 值的大小和用户价值呈现反比，因此应该是 R 值越低，分数越高。得到每位顾客的 RFM 评分后，将每一项的平均值作为阈值进行判断，如若三项都高于平均值，则为重要价值客户。如下图所示：

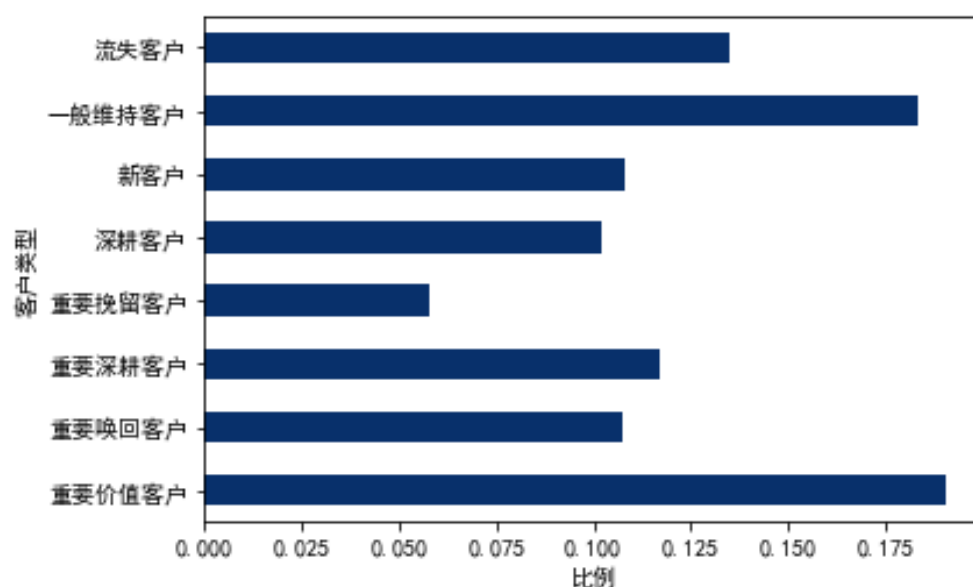


图 14：客户类型比例图

在得到每个顾客所属类别后，便可针对不同的类别制定不同的营销方案，对优质客户指定个性化的沟通和营销服务；为更多的营销决策提供有力支持。如对重要价值客户，可以赠送优惠券进行反馈，同时，还可以结合购买历史，进行个性化商品推荐，免费提供送货上门等服务。而对于重要挽留客户可以提供打折促销，优惠券赠送等吸引客人。此外，对于一般维持客户，可以开展一元购，捆绑销售等来促进消费。具体如下表 2 所示：

表 2：客户类型与营销方案制定

客户类型	营销方案
重要价值客户（RFM 三个值都高）	免费送货上门服务、个性化推荐、VIP 服务（如免排队）
重要发展客户（F 低）	个性化推荐，优惠券赠送、更新商品以吸引再次消费
重要保持客户（R 低）	个性化推荐、主动联系推荐（如短信）等
重要挽回客户（R、F 低）	促销活动、进行调查寻找流失原因并进行推荐
一般价值客户（M 低）	促销活动、一元购、优惠券等
一般发展客户、一般保持客户	提供积分制、优惠活动和打折等
一般挽留客户（RFM 都低）	放弃吧，节省点开支

4.3 关于个性化商品推荐

基于上述分析，我们建立了基于会员用户的标签系统，借助该系统，我们可以直观的了解会员的购买偏好与信息，以此为基准进行个性化推荐或捆绑销售。以卡号 00075d60 为例(如图 13 所示)，我们可以知道该会员的性别为女，入会时长 4 年，同时也知道了她的购物时间偏好和商品偏好，因此，我们可以在春季为其推荐加个较高的化妆品。此外，通过分析，也知道了该商场的淡季为春季，因此，可以在该季节进行一些促销。同时，也能基于旺季和淡季，来合理安排商场人员，节省开支。更进一步的，若有商品的类别信息，还能构建基于商品的标签系统，从而基于会员和基于商品的协同过滤算法，构建个性化商品推荐。

针对上面的 RFM 模型，也可进行相关改进，如增加指标数目，并利用聚类算法对会员进行分类营销，如在 RFM 的基础上，增加入会时长、消费次数、平均消费金额三个指标，并利用聚类算法得到聚集点后可得到如下雷达图：

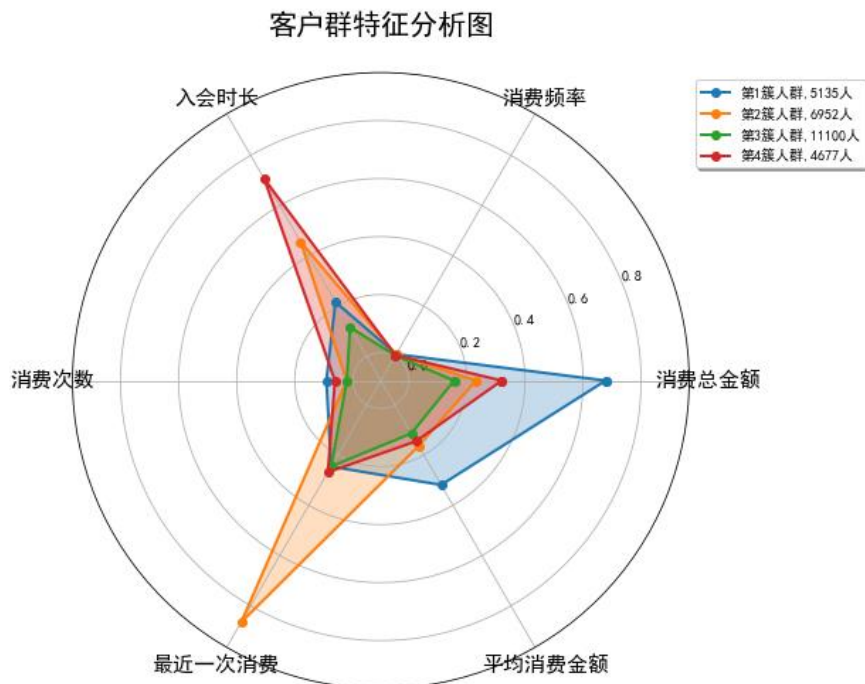


图 15: 客户群特征分析图

5. 结语

通过这次百货商场数据分析，增加了我对数据预处理的熟悉，同时，也让我对 RFM 模型有了一定的认识与实操，此外，也让我对聚类分析和雷达图的绘制有了实操经验。更让懂得了在数据分析的路上，我还有很多不会的知识，还需要不断的进行学习。

在完成这一项目的过程中，对于这一项目，还有很多地方是我没有完善好的，比如，模型预处理，RFM 模型等很多地方可以建立函数、类来完成，实现面向对象编程，而现在只是针对具体问题来编程，可拓展性不高。对于这一点，可以进一步的简化代码、模块化代码，以提高代码的可重用性。

此外，我也查阅了不少资料，让我对 RFM 模型的学习与应用大开眼界，如对 RFM 模型的指标选取，使用什么软件来完成（python、R、Excel、Tableau 等都可以），其中，使用 Tableau 软件来绘制图很丰富，而且操作很简单。对于 Tableau 这一软件，虽然之前也学过，但是并没有很多的实操经验，只是在完成作业的时候使用了，却没有进行更多的实操，因此，在之后的时间中，希望能找时间来使用 Tableau 进行 RFM 建模。