

大型百货商场会员画像描绘

摘要

商场会员消费是商场盈利的主要来源，本文主要针对会员与非会员、活跃会员与非活跃会员等进行数据分析，旨在描绘商场会员画像，有利于商场获利最大化。

针对问题一：首先对原始数据进行预处理及标准化处理，再借助于 Python 语言、Excel 处理目标数据，按性别分析商场会员消费特征，得到男女消费总额比例为 3:17，男女人均消费分别为 14260.41 和 13430.69（元）。利用 Excel 绘制会员与非会员消费金额对比图，更加直观地说明会员群体给商场带来的价值。

针对问题二：给出购买力的定义，建立购买力数学模型，并利用 Python 语言进行数据处理，得到会员购买力数据即购买商品的平均价格，并利用对数标准化公式对会员购买力进行标准化处理，绘制会员购买力的价值直方图，对每个会员的价值进行识别。

针对问题三：计算会员的生命周期，

针对问题四：

针对问题五：通过查阅资料得知连带消费和连带率的定义，利用 Python 编程计算得出每天的连带率，且连带率在 2016 年 11 月 24 日达到最大值，在对取连带率最高的四天进行分析可得，兰芝化妆品正价瓶、素然正价件、欧舒丹化妆品瓶等商品更符合会员的喜好。

关键词：

一、问题重述

随着互联网的发展，实体店商场受到了极大的冲击，在目前我国商场行业中，会员的价值主要体现在连续不断地为零售运营商带来稳定的销售额和利润，同时也可以为零售运营商想要制定的策略提供数据支持，因此，运营商需要有针对性地实施营销策略来加强运营商与会员的良好关系。完善会员画像描绘，加强对现有会员的精细化管理，定期向其推送产品和服务，与会员建立稳定的关系是实体零售行业得以更好发展的有效途径。

请根据所给附件数据，建立数学模型解决以下问题：

(1) 分析该商场会员的消费特征，比较会员与非会员群体的差异，并说明会员群体给商场带来的价值。

(2) 针对会员的消费情况建立能够刻画每一位会员购买力的数学模型，以便能够对每个会员的价值进行识别。

(3) 作为零售行业的重要资源，会员具有生命周期(会员从入会到退出的整个过程)，会员的状态(比如活跃和非活跃)也会发生变化。试在某个时间窗口，建立会员生命周期和状态划分的数学模型，使商场管理者能够更有效地对会员进行管理。

(4) 建立数学模型计算会员生命周期中非活跃会员的激活率，即从非活跃会员转化为活跃会员的可能性，并从实际销售数据出发，确定激活率和商场促销活动之间的关系模型。

(5) 连带消费是购物中心经营的核心，如果商家将策划某次促销活动，如何根据会员的喜好和商品的连带率来策划此次促销活动？

二、问题分析

本文主要目标是通过对百货商场会员在商场消费记录及消费频率进行分析，研究如何使会员持续不断地为零售运营商带来稳定的销售额和利润，从而使商场利润最大化，鉴于此目的，针对本文具体 5 个问题，可以进行如下分析：

2.1 针对问题一的分析：

本问题主要通过处理附件一《该大型百货商场的会员信息表》与附件三《会员消费明细表》会员数据，通过附件三会员卡号的注释可以了解在该大型百货商场消费的会员消费明细也拥有其他分店的会员，进行数据处理可分离出该大型百货商场会员与其他分店会员，把附件一与附件三的数据处理之后可以分析出该大型百货商场会员的消费特征，再通过数据处理对比附件二与附件三已分离出的会员群体与非会员群体的群体差异，最后通过数据对比，从而反映出会员群体给商场带来的价值。

本题主要从以下几点考虑：

1. 通过处理附件二及附件三中的数据可以分离出会员群体与非会员群体，利用 Python 语言编程可对性别等各个方面分析会员消费特征；

2. 在分离出会员与非会员的信息之后，通过附件三中的会员消费信息，利用

Python 语言编程计算可得出会员与非会员消费总额对比，之后再次利用 Python 语言绘制出会员与非会员的人均消费金额图，两份图进行数据对比即可比较出会员与非会员的差异；

3. 由 2 中绘制的结构图可知，虽然会员人数较少，但是消费金额远远超出非会员群体，所以会员群体能给商场带来较大利益。

2.2 针对问题二的分析：

本问题针对会员消费情况，建立商品平均购买价格的购买力数学模型，并对购买力进行对数标准化处理，并且用数学模型刻画出每一位会员的购买力。根据购买力的不同，对不同会员进行购买力划分。

2.3 针对问题三的分析：

附件三中所得的数据是注册时间不同的会员在同一时期的消费情况，这些数据反映的可能是会员的不同时期的消费水平，如果可以利用附件三得到的会员消费情况就可以反映出会员在周期内各个时段的消费情况，经过数据预处理后对前五十万的原始数据进行处理，计算出各个会员的注册时间直至相应会员的最后一次消费时间的间隔时间，即为某个时间窗口，把所得到的间隔时间与一段时间内的消费情况相结合，得到的一段时间的消费情况就可以反映出会员周期内各个时段的消费情况；

将间隔时间进行数据处理分为若干合理区间，通过 Python 语言编程并计算结果，之后利用 Excel 对若干个区间点进行累加求和，得到目标散点图；在得到散点图后，再次利用 Logistic 模型进行数据拟合，最后再对得到的拟合曲线进行拟合度检验，即可建立会员生命周期和状态划分的数学模型。

2.4 针对问题四的分析：

本题若要建立关于计算会员生命周期中非活跃会员的激活率的数学模型，可以处理原始数据得到会员是否有第一次消费记录信息的数据，若在记录时间中会员无消费记录，记为生命周期中的非活跃会员，在记录时间中会员消费一笔及以上，就记为该会员是活跃会员。对于非活跃会员商家可采用促销的方式来刺激消费，吸引消费者来进行消费，在刺激消费成功后，收集数据即可建立激活率和商场促销活动之间的关系模型。

2.5 针对问题五的分析：

三、数据预处理

在进行统计描述前，有必要对数据做修复处理如下：

- 假设所给数据真实有效；
- 附件一中年龄大于等于 85 岁的会员但有消费记录者为年龄信息输入错误；
- 附件一中年龄大于 85 岁的会员且无消费记录为无效信息；
- 会员入会截止日期为 2018 年 9 月 13 日，若在之后仍有会员入会记录，则视为无效信息；
- 会员入会记录早于会员出生日期的数据视为无效信息。

四、模型假设及符号说明

为了我们更好的解决该问题，在解决问题之前我们做出以下假设

4.1 基本假设：

- 假设 1：所给数据真实有效；
- 假设 2：附件一中红色数据默认为会员购买商品后退货，若退货金额与相应购买金额相同，则退货金额与相应购买金额相抵消，视为两次无消费；
- 假设 3：会员入会截止日期为 2018 年 9 月 13 日，以后入会的视为无效数据；

4.2 符号说明：

G_i	表示第 <i>i</i> 个会员的购买力
x_{ij}	表示第 <i>i</i> 个会员第 <i>j</i> 次购买商品的总额
C_i	表示第 <i>i</i> 个会员购买商品的次数
$F(t)$	会员的累计分布函数
$Z(t)$	会员的周期函数
L_i	会员生命周期中非活跃会员的激活率
I_i	某个区间内会员的折扣
J_i	某个区间内激活的会员卡号数量
R_i	第 <i>i</i> 天的连带率
S_i	第 <i>i</i> 天的销售总数量
T_i	第 <i>i</i> 天销售小票数量

G'_i	第 <i>i</i> 人标准化之后的购买力

五、模型建立及求解

5.1 基于消费金额分析会员与非会员群体消费差异：

我们利用 Python 语言对附件一中的会员性别、消费金额及消费次数进行数据处理，再使用 Excel 合理的处理数据即可绘制得到该商场的会员消费特征的男性女性消费结构图。

利用 Python 对数据进行处理（源程序见附录 1），得到男女消费总额分别为 53205302.9 元和 296966195.09 元，男女人均消费分别为 14260.41 元和 13430.69 元，如下图：

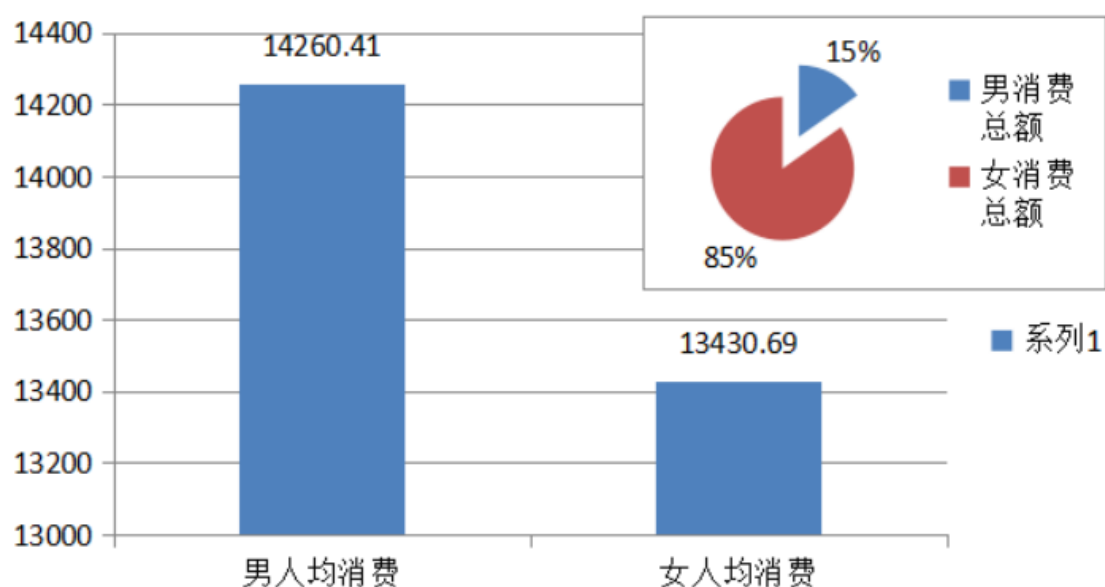


图 5.1.1-男性女性消费结构图

在上图可简单直观地看出，虽然男性消费总额占比仅为百分之十五，但是男性人均消费达到了 14260.41 元；尽管女性消费占比总消费金额百分之八十五，但是女性人均消费金额仅为 13430.69 元。由图分析可知该大型商场会员的消费特征在于虽然女性在该商场消费金额远超男性，但是男性人均消费金额却略高于女性。

通过附件二（大型百货商场近几年会员与非会员销售流水表）与附件三（商场本店与其他分店的会员消费总额明细表）的内容对比，及使用 Python 语言（具体源程序见附录 2）数据处理可得出该商场近几年非会员消费次数，进行深度计

算可得出会员与非会员的每次消费金额对比图。

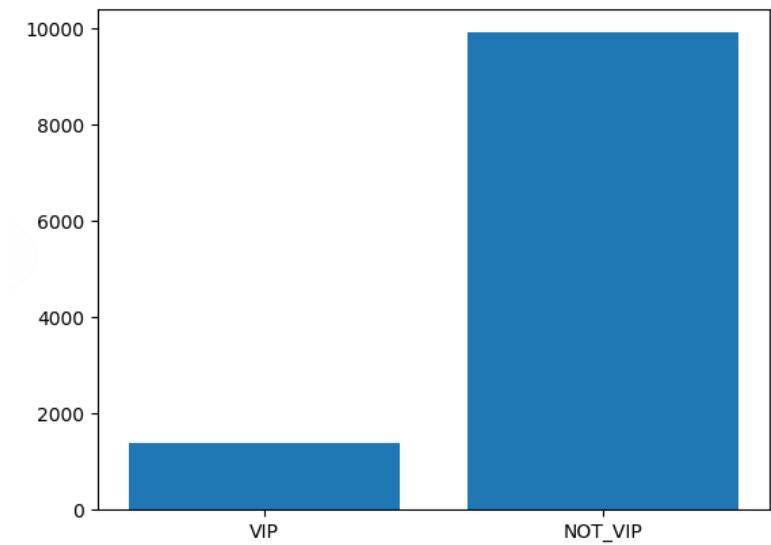


图 5.1.2-会员与非会员的均次消费金额对比柱形图

从图 5. 1. 2 中可以看出，会员与非会员平均每次的消费金额差距较大，结果发现差距较大的原因为重复计算销售数量为“-1”（即为退货）的货物，在重新计算之后，我们利用 Excel 对此进行数据分析，即可得到新的图 5. 1. 5。

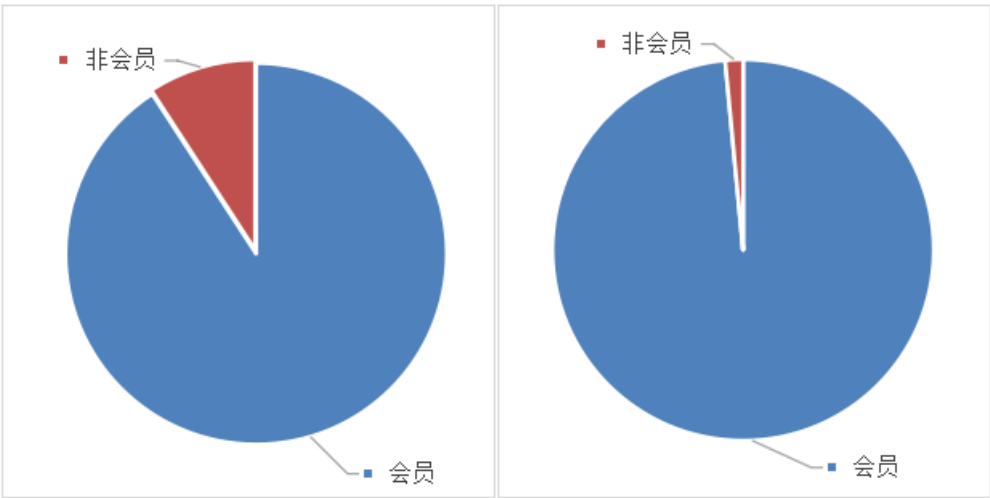


图 5.1.3-会员与非会员销售数量结构图

图 5.1.4-会员与非会员销售金额结构图

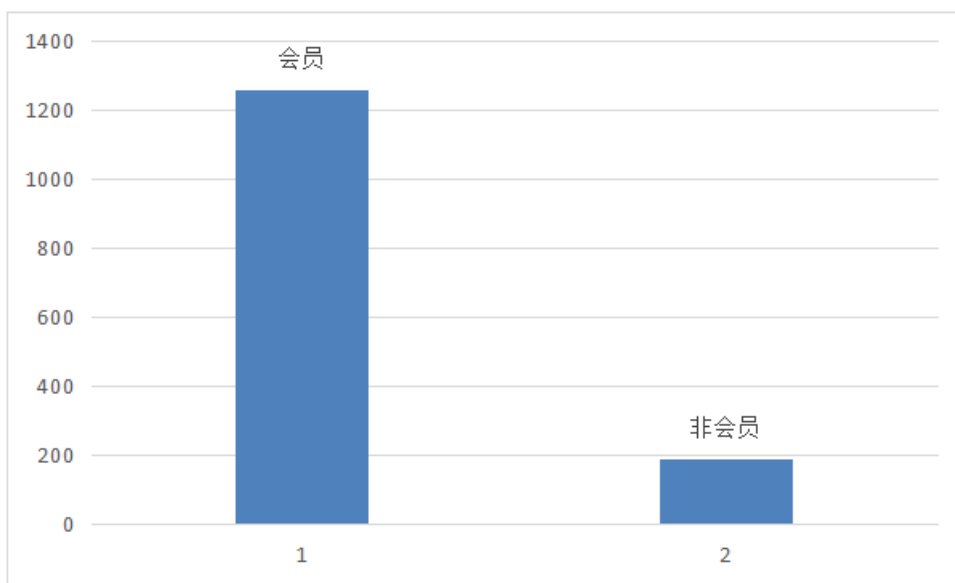


图 5.1.5-会员与非会员平均单件商品消费额度对比图

利用 Excel 表格处理数据可绘制得出图 5.1.3 与图 5.1.4，这两份结构图可以让我们观察数据更加形象直接。由图 5.1.3、图 5.1.4、图 5.1.5 这三份对比图可以直观的看出会员群体与非会员群体的消费差异。无论是在会员与非会员销售数量对比还是在会员与非会员销售金额对比图中，会员群体都是遥遥领先于非会员群体，并且在图 5.1.5 中，会员平均单件商品消费额度约为 1259.7 元，而非会员平均单件商品消费额度仅为 189.8 元左右，在直观分析的前提条件下，我们可以清晰的看出会员群体与非会员群体的消费差异。

分析结果可得，会员群体的总销售金额为 1204268449 元，非会员群体的总销售金额仅为 18331271.01 元，会员群体总消费金额约为非会员群体的 65 倍，由此可以看出会员群体给商场带来的巨大价值。

5.2 会员购买力的数学模型：

购买力是指在一定时期内用于购买商品的货币总额，因为各个会员购买商品的时间段和购买商品的次数都有所不同，所以建立购买商品时段的平均购买力模型，即从注册会员第一次购买商品到该会员最后一次购买商品总次数的平均购买额度来评价会员的购买能力，使用下列公式计算数据：

$$G_i = \frac{\sum x_{ij}}{C_i}$$

G_i ：表示第 i 个会员的购买力

X_{ij} ：表示第 i 个会员第 j 次购买商品的总额

C_i :表示第*i*个会员购买商品的次数

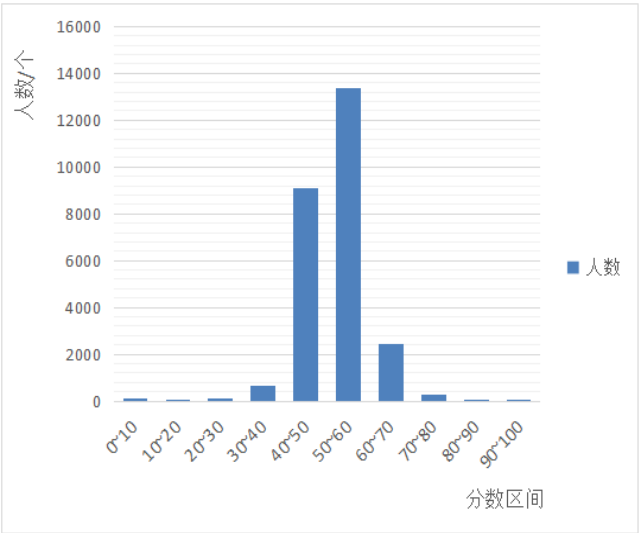
利用公式处理原始数据得目标数据（源程序见附录 3），再利用 Python 语言处理目标数据可得出不同会员的购买能力如下表所示：

会员购买商品的平均价格			
kh	ALL_SALL	Count	Mean
110640	4625	10	462.5
126951	1260	1	1260
206962	5150	10	515
.....
fffa7c0b	13835	19	728.1578947
fffb0ce	7027	3	2342.333333
fffc9664	306	1	306

我们对 G_i 进行标准化处理，得到第*i*个人标准化后的购买力数据，标准化公式如下：

$$G_i' = \frac{\ln(1 + G_i)}{\ln(1 + \max G_j)} * 100$$

对标准化之后的数据进行绘图可得：



从图中可以看出大部分会员的购买力集中在 60 左右，而取值为 0 的点由于未消费所以尚不明确这些会员的购买力。

5.3 针对问题三的模型建立及求解：

关于本题所提出的会员的生命周期（即会员从入会到退会的整个过程），附件中并未提及会员退会的信息，所以我们默认会员为从入会到最后一次消费为一位会员的生命周期。

附件三中给出的数据是注册时间不同的会员在同一时期的消费情况，此时的数据对于不同会员反映的可能是会员的不同时期的消费水平，为了更为准确的得出会员周期，就需要筛选出会员在会员周期内不同时段消费情况。

因此需要对这些原始数据进行数据处理。在经过数据预处理之后，我们取原始数据的前五十万数据进行计算，计算出各个会员的注册时间直至相应会员的最后一次消费时间的间隔时间，所得到的间隔时间与一段时间内的消费情况相结合，得到的一段时间的消费情况就可以反映出会员周期内各个时段的消费情况，计算结果及过程详情见附录。

数据处理结果为天数共 5551 天，我们将其分为 25 个区间，每个区间约 228 天，通过 Python 语言编程计算结果，之后利用 Excel 对 25 个区间点进行累加求和，得到如下散点图（代码见附录..）

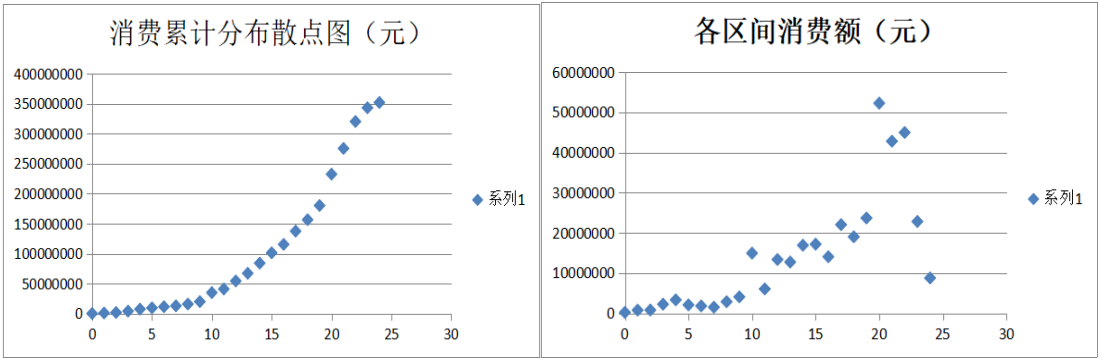


图 5.3.1

图 5.3.2

我们利用 logistic 拟合曲线进行拟合度检验可得下表 5.3.1：

	$F(t) = ae^{bt} + ce^{dt}$	备注
a	$8.413 \times 10^7 (-2.125e + 012, 2.126r + 012)$	
b	$0.2831 (-15.58, 16.15)$	
c	$-8.128 \times 10^7 (-2.126e + 012, 2.125e + 012)$	
d	$0.2844 (-15.75, 16.32)$	
	$R^2 = 0.997, Adjusted \quad R^2 = 0.9966$	

表 5.3.1

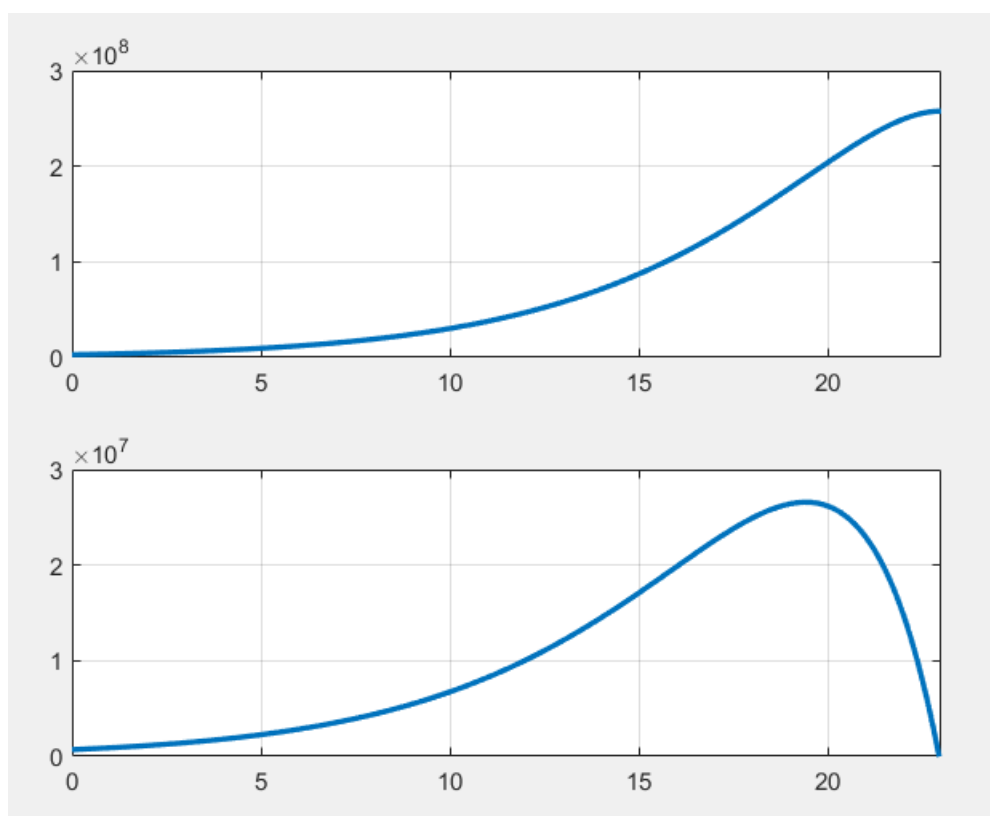
从表格的结果分析中我们可以得到，此次拟合效果较好。将表格中数据带入函数 $F(t)$ 中，可以得到会员的累计消费函数

$$F(t) = 8.413 \times 10^7 e^{0.2831t} - 8.128 \times 10^7 e^{0.2844t}$$

对累计消费函数求导可得到会员的周期函数为

$$Z(t) = F'(t) = 2.382 \times 10^7 e^{0.2831t} - 2.312 \times 10^7 e^{0.2844t}$$

在此之后，我们利用 MATLAB 拟合工具箱对消费累计分布散点图中的数据再次进行 logistic 曲线拟合，可以得到下图数据：



上图为 logistic 曲线拟合结果，我们可以明显地看出会员生命周期的状态为先上升再衰退，为了尽可能的维持峰值，商家需要尽可能的在衰退时刺激会员消费，用以增长会员周期，延长处于峰值的状态，还要不断地接收新会员使商场有充足的会员群体。

5.4 激活率和商场促销关系模型：

本题若要建立数学模型计算会员生命周期中非活跃会员的激活率，需要计算非活跃会员转化为活跃会员的可能性，所以我们需要处理附件中的原始消费数据从而得到实际销售数据。

本题使用统一计算所有会员的第一次消费时间的方式，在数据统计完毕后，若在记录时间中会员无消费记录，则视为生命周期中的非活跃会员；若在记录时间中会员消费一笔或一笔以上，则视为该会员是活跃会员。

对于非活跃会员商家可采用促销的方式来刺激消费，吸引消费者来进行消费，在刺激消费成功后，收集数据建立激活率和商场促销活动之间的关系模型。

在问题三中我们建立了 0-24 总共 25 个消费区间，数据处理完毕后可得每个区间内促销折扣的价格，在计算出每个区间内的激活卡号数量后，再利用公式计算出每个区间的激活率。

由数据可得计算会员生命周期中非活跃会员的激活率的公式为：

$$L_i = \frac{J_i}{\sum_{i=0}^{24} J_i}$$

其中 L_i ：表示会员生命周期中非活跃会员的激活率

J_i ：表示某个区间内激活的会员卡号数量

利用公式可建立实际销售数据的 Excel 表格分别表示区间、折扣、激活率，在 Excel 表格中绘制消费区间与折扣、激活率之间的散点图，如下图所示：

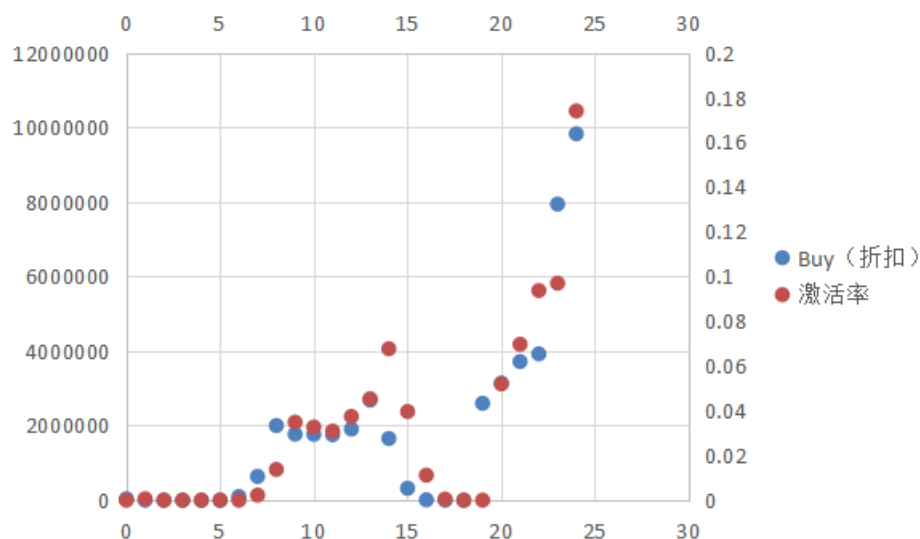


图 4.

在图中可以直观的看出在不同的区间内折扣和激活率的增长图像，我们可以看出，在刚开始的区间内，商场并未作出折扣促销活动，激活率也一直停滞不前，在 5-15 的区间内，随着商家的折扣活动的增加，非活跃会员的激活率也一直增加，而在 15-20 的区间内，由于促销活动力度下降，非会员的激活率也随之下降，从第 20 个区间之后，商家一直在大力促销，会员生命周期中非活跃会员的激活率也不断随之增加，直到最后一个区间达到峰值。

5.5 基于连带消费与连带率的促销活动处理：

连带消费是一些商场及大型超市处理经营的核心，连带消费可以使顾客产生更多的消费，从而使商场的利润在范围之内尽可能的提高，商家便可以得到更多的利益；

一般情况下，商家会针对某些节日进行商品促销活动，这些活动在刺激顾客的消费欲望之后，大部分顾客都会在商场中进行消费，从而使商场获利。

连带率是指销售的件数和交易的次数相除后的数值，反映的是顾客单次消费的产品件数。对于追求多业态混合消费的购物中心及商场来说，连带消费是购物中心经营的核心，品牌组合固定时，要提升客单价就看连带率是否能提高。

如果商家计划未来策划某次促销活动，则可以根据会员的购买记录来确定某个会员的购买次数及每次的购买件数，由于连带消费一般由价格较高的商品连带价格较低的商品，所以商家只要对会员的消费记录进行数据处理，并且统计有较高连带率的商品进行数据分析，对出现频率较高的高连带率商品进行统计（具体数据详见支撑材料）。

由连带率的定义可得公式为：

$$R_i = \frac{S_i}{T_i}$$

R_i ：第 i 天的商品连带率；

S_i ：第 i 天的销售总数量；

T_i ：第 i 天的销售小票数量；

利用 Python 语言计算出每天商场的商品的连带率（源代码见附录 7）（源代码见附录 8）。在连带率较高时，便可证明连带消费非常成功，因此我们取出连带率最高的四天里的所购部分商品与件数进行分析，得出每天商品售量的大小，如下表：

第一连带率		第二连带率	
spmc	s1	spmc	s1
兰芝化妆品正价瓶	100	DIDIER PARAKIANB. 5.	1494
欧舒丹化妆品瓶	92	LEE童装 F无	97
罗莱促销F件	76	素然正价件	27
素然正价件	73	兰芝化妆品正价瓶	25
宝明堂（农）B件	58	雅诗兰黛舒缓眼膜11ml	25
后瓶	54	DAZZLE DIAMOND A件	17
兰纯口红/唇釉/纯魅/蜜糖/	43	soho用品正价件	17
LA FEEB件	42	伊维斯正价件	17
第三连带率		第四连带率	
罗莱促销F件	117	罗莱促销F件	107
欧舒丹化妆品瓶	105	后瓶	88
ICICLE A件	102	素然正价件	78
兰芝化妆品正价瓶	99	阿玛尼丝绒/漆光唇釉.	77
得臣（AVFASON）服饰系列F	80	兰芝化妆品正价瓶	72
玛丝菲尔淑女装系列A件	78	宝姿女装B. 5件	57
厚木系列B. 5件	74	欧舒丹化妆品瓶	54
素然正价件	65	爱慕内衣正价件	53

由上表可以看出，在商品的连带率较高的几个时期，可以看出当时会员的喜好物品，通过对比不难发现，有部分商品在不同时期内都会出现，如，素然正价件，兰芝化妆品正价瓶，欧舒丹化妆品瓶等，可以看出，即使在不同时期，会员的喜好大致相同，不会因为时间不同产生太大差异。故若商家想要策划某次促销活动，可以在连带率较高的部分时间内，同时通过数据处理得出让会员长时间的喜好的物品进行销售并进行促销活动。

六、模型改进及推广

6.1

6.2

6.3

七、参考文献

- [1] 姜启源 谢金星 叶俊，《数学模型》，北京：高等教育出版社，2005年；
[2] 卓金武等，《MATLAB在数学建模中的应用》（第二版），北京：北京航空航天大学

出版社，2016 年；

[3] 司守奎 孙兆亮，《数学建模算法与应用》（第二版）北京：国防工业出版社，2016 年；

[4] 盛 骤 谢式千 潘承毅，《概率论与数理统计》，北京：高等教育出版社，2015 年。

附录

注意：文件路径随着实际路径变化！

附录 1：

```
#coding:utf-8
import pandas as pd
import numpy as np
table_3=pd.read_excel(r'C:\Users\14112\Desktop\附件 3-会员消费明细表_改.xlsx')
table_3=pd.DataFrame(table_3)
table_3['Year']=table_3['Year'].values.astype(float)
table_3=table_3[table_3['Year']<85.0].to_excel(r'C:\Users\14112\Desktop\True_VIP.xlsx')    # 对于年龄是 85 岁以下并且有消费记录的附件 3 数据，是我们想要的数据

table3_g=pd.read_excel(r'C:\Users\14112\Desktop\True_VIP.xlsx')
table3_g=pd.DataFrame(table3_g)
table3_g=table3_g[table3_g['入会时间']!='Nan']          #数据清洗，拿到年龄没有 Nan 的所有数据，并写入新文件 REALL_VIP.xlsx 文件中
table3_g.to_excel(r'C:\Users\14112\Desktop\REALL_VIP.xlsx')

table_3=pd.read_excel(r'C:\Users\14112\Desktop\REALL_VIP.xlsx')
table_3=pd.DataFrame(table_3)
Man=table_3[table_3['Sex']==1]          #拿出性别是男士的所有数据
KaHao=len(set(Man['kh'].values))        #卡号去重后，真正的男士人数
print(KaHao)
#男士人数： 3731

Man_ALL=Man['sl']*Man['je']    #男士消费的总额
print(Man_ALL.sum())
#男士消费的总额： 53205602.900000006
Man_Mean=53205602.900000006/3731    #男士的人均消费
print(Man_Mean)
#男士的平均消费： 14260.413535245245

Woman=pd.DataFrame(table_3[table_3['Sex']==0])    #拿出性别是女的所有数据

KaHao=len(set(Woman['kh'].values))    #去重卡号之后，女生的真正人数
```

```

Woman_ALL=Woman['sl']*Woman['je'] #计算 女生的消费总额
print(Woman_ALL.sum())
#女生消费总额: 296966195.09
print(KaHao)
#女生的人数: 22111
Woman_Mean=296966195.09/22111 #计算女生的人均消费
print(Woman_Mean)
# 女生的平均消费: 13430.699429695625

```

附录 2:

```

#!/usr/bin/env python
# -*- coding:utf-8 -*-
'''
Author:YeYe
'''

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# 计算非会员的总消费与每条流水的平均消费
tables_2=pd.read_excel(r'C:\Users\233\Desktop\C 题附件\附件 2-销售流水表.xlsx')
tables_3=pd.read_excel(r'C:\Users\233\Desktop\C 题附件\附件 3-会员消费明细表.xlsx')
tables_3=pd.DataFrame(tables_3[['je','sl','sj']])
tables_2=tables_2[['je','sl']]
tables_2=pd.DataFrame(tables_2)
ALL=tables_2['je']*tables_2['sl']
VIP=tables_3['je']*tables_3['sl']
ALL=ALL.sum()
# print(ALL)
VIP=VIP.sum()
NOT_VIP=ALL-VIP
Soure=1015366-911702-19763*2
NOT_VIP=NOT_VIP/Soure
print(NOT_VIP)
#非会员人均消费 9916.546449972995
#非会员的总消费 2475633347.6400003

VIP_Reduce=tables_3['sl']*tables_3['sj']
VIP_Reduce=VIP_Reduce.sum()
print(VIP)
print(VIP_Reduce)

```

```

print(VIP-VIP_Reduce)
'''
商场会员的售价总额: 1447644476.4499998
会员的实际消费总额: 1254852629.069999
会员的促销总额: 192791847.38000083
'''

Mean_VIP=1254852629.069999/911720-6875*2
print(Mean_VIP)
#会员的人均的消费 1376.3574661847924

plt.bar(['VIP','NOT_VIP'],[1376.3574661847924,9916.546449972995])
plt.show()
plt.bar(['VIP','NOT_VIP'],[1254852629.069999])

print(tables_2[tables_2['sl']==-1].count())
print(tables_3[tables_3['sl']==-1].count())
'''
je      19763
sl      19763
dtype: int64
je      6875
sl      6875
sj      6875
dtype: int64

'''
print(len(tables_2[tables_2['sl']==-1]['sl'].values))
print(len(tables_3[tables_3['sl']==-1]['sl'].values))
#19763
# 6875

NOT_VIP_MEAN=2475633347.6400003/(1015366-911702-19763*2)
print(NOT_VIP_MEAN)
# 38598.542948641996

VIP_MEAN=1254852629.069999/(911720-6875*2)
print(VIP_MEAN)
# 1397.4326860251444

```


附录 3:

```
#!/usr/bin/env python
# -*- coding:utf-8 -*-
'''
Author: YeYe
'''
import pandas as pd
import numpy as np
Data=pd.read_excel(r'C:\Users\233\Documents\Tencent Files\2500594147\FileRecv\REALL_VIP
- 副本.xlsx')

Data=pd.DataFrame(Data)
Data['ALL_SALL']=Data['sl'].astype(float)*Data['je'].astype(float) # 卡号对应的消费总金额

JieGe=pd.DataFrame(Data.groupby('kh')['ALL_SALL'].sum()) #计算每一个卡号对应的总消
费金额
Count=pd.DataFrame(Data.groupby('kh')['ALL_SALL'].count()) #计算每一个卡号消费的次数

JieGe['Count']=Count.values
JieGe['Mean']=JieGe['ALL_SALL']/JieGe['Count'] #新建一列 Mean，每一张卡号平均消
费金额
JieGe.to_excel(r'C:\Users\233\Desktop\问题 2.xlsx') #把结果写入问题 2.xlsx 文件中
```

附录 4:

附录 5:

附录 6:

附录 7:

```
#!/usr/bin/env python
# -*- coding:utf-8 -*-
'''
Author: YeYe
'''
```

```

import pandas as pd
import numpy as np
import datetime
Data=pd.read_excel(r'C:\Users\233\Desktop\附件 3-会员消费明细表_改.xlsx')
Data=pd.DataFrame(Data)
Data=Data[Data['Sex']!='Nan']

Data=Data[Data['VIP']=='T']
Data=pd.DataFrame(Data)
Data=Data[Data['Year']<=85]

#-----上述代码为数据预处理-----

Data=pd.DataFrame(Data)
Data['dtime']=Data['dtime'].astype(str)
Data['dtime']=Data['dtime'].str[:10]
li=[]

reslut2=Data.groupby('dtime')['sl'].sum()#找到每个时间对应的销售数量总和

reslut2=pd.DataFrame(reslut2)

for i in reslut2.index:
    a = len(set(Data[Data['dtime'] == i]['djh'].values))#每个日期开了多少张单据票
    li.append(a)

reslut2['djh']=li

reslut2['VALUES']=reslut2['sl']/reslut2['djh'] #新建一列 VALUES，此天连带率= 此天销售数
量/此天内开的单据票

reslut2.to_excel(r'C:\Users\233\Desktop\REALUT_4_最后.xlsx')

```

附录 8:

```

#!/usr/bin/env python
# -*- coding:utf-8 -*-
'''
Author:YeYe
'''
import pandas as pd
import numpy as np

```

```
Data=pd.read_excel(r'C:\Users\233\Desktop\附件 3-会员消费明细表_改.xlsx')
```

```
Data=pd.DataFrame(Data)
```

```
Data['Year']=Data['Year'].astype(float)
```

```
Data=Data[Data['Year']<85.0]
```

```
#-----以上为数据清洗-----
```

```
#由第五题.py 程序生成的文件 REALUT_4_最后.xlsx, 排序观察可得 VALUES 最高的 5 个时间分别为如下: 分析这些时间内卖的 spmc 和件数
```

```
'''
```

```
dtype
```

```
2016-11-24
```

```
2015-04-25
```

```
2016-11-26
```

```
2016-11-25
```

```
2015-04-26
```

```
VALUES
```

```
21.75
```

```
17.48192771
```

```
15.13539192
```

```
14.50271739
```

```
12.8136646
```

```
'''
```

```
Data=pd.DataFrame(Data)
```

```
Data['dtype']=Data['dtype'].astype(str)
```

```
Data['dtype']=Data['dtype'].str[:10]
```

```
Data=Data[Data['dtype']=='2015-04-26'] #此处可以把 2015-04-26 替换其他时间 如:  
2016-11-24 就是计算 VALUES 最高的时间
```

```
Data.to_excel(r'C:\Users\233\Desktop\第五题\第五题 Mini_4.xlsx')
```

```
#
```

```
Data=pd.read_excel(r'C:\Users\233\Desktop\第五题\第五题 Mini_4.xlsx')
```

```
Data=pd.DataFrame(Data)
```

```
Count=Data.groupby('spmc')['sl'].sum() #统计这一天卖的不同 spmc 与对应件数的和
```

```
Count=pd.DataFrame(Count)
```

```
Count.to_excel(r'C:\Users\233\Desktop\第五题\第五题 Count_4.xlsx')
```