

健身房团购数据研究报告

自律给我自由

复旦第三区交通委小组

李泽君、赵雅滢、谢炳辉、王维实、钟诚

2019.3.31



自律给我自由

目录

一、 背景介绍.....	2	3.7 有效期.....	7
二、 数据预处理.....	3	3.8 起始年份.....	7
2.1 缺失数据处理.....	3	3.9 节假日可用.....	8
2.2 哑变量处理.....	3	3.10 预约情况.....	8
2.3 数值型变量处理.....	3	3.11 是否有温馨提示.....	9
2.4 提取衍生变量.....	3	3.12 是否为体验活动.....	9
三、 变量分析.....	3	四、 模型拟合.....	10
3.1 因变量.....	3	4.1 初步拟合.....	10
3.2 评分.....	4	4.2 变量间的相互影响.....	10
3.3 评论数.....	4	4.3 模型优化.....	11
3.4 评论数 X 评分.....	5	五、 外部资料与模型的改进方向.....	12
3.5 计次方式.....	6	六、 小组分工.....	13
3.6 项目.....	6		



一、背景介绍

健身房健身已然成为了一种时尚，一种潮流。健身不仅可以使肌肉强健还能预防疾病，除此之外，健身也能促进代谢速率，增加肌肉皮肤弹性，增强身体抵抗力，改善睡眠，释放工作压力等等功效。可以说，健身百利而无一害，只要通过正确的方法，规律的安排，足够的补充及合理的休息，健身都能在各方面明显改善身体机能。越来越多的都市人选择在健身房健身锻炼。

在这样安全又时尚的活动中健身房的选择就成为一个问题，由于房的规模、环境、价格参差不齐，健身房的选择就变得尤为关键。这里我们在网站上抓取了 1581 条数据，试图分析影响健身房团购数的因素。

二、数据预处理

2.1 缺失数据处理

在我们搜集的健身房团购数据中，存在缺失的数据列有“评分”，“评论数”，“规格”，“团购价格”，“团购内容”，“有效期”，“可用时间”和“预约说明”。其中“评分”、“评论数”、“规格”、“价格”，“内容”全部缺失的数据显然是应当删去的。删去后剩余 1581 条数据。

对于“评论数”数据，我们可以发现“评论数”数据栏中并不存在值为 0 的数据，故可以推测“评论数”数据的缺失是因为评论数为 0，将其缺失值全部填补为 0。

在其他数据中，“规格”、“团购内容”、“可用内容”、“预约说明”属于类别型变量，之后可以将其化为哑变量进行处理，故先不加以考虑。其余列属于数值型变量，故将“评分”，“价格”变量先设置为其他有效数据的均值。

2.2 哑变量处理

对于健身房而言，办卡的长度和次数无疑是非常重要的选择因素之一。在“标题”，“团购内容”，“温馨提示”中提取与次数相关的关键词。发现有以下几个关键词：“月卡”、“年卡”、“季卡”、“周卡”、“节”、“次”、“份”等。而对于其他使用“小时”，“课”，“期”计数的样本，我们将其看作“一次”消费。按照这样的原则，我们可以将次数变量进行分类。

与此同时，我们还对健身项目进行了简单分类，依据“团购内容”栏中出现过的种类分为桌球、瑜伽、舞蹈、健身和其他，而对于其他没有在该栏中体现的项目，我们尝试着从其地点和店名中寻找线索。（如“复旦台球馆”标记为台球项）。

2.3 数值型变量处理

对于“有效期”列这一区间变量，我们先将缺失值设置为其他数据的众数。而对于这一变量，我们将其通过字符串处理分解为“持续时间”和“起始年份”两个变量，以研究办卡的持续长度和年份对团购数的影响。另一方面，我们对起始年份进行了归一化处理，以防止其影响截距项和系数。

2.4 提取衍生变量

至此，仍有“可用时间”、“预约说明”、“团购内容”以及“温馨提示”列未被处理，我们从中提取以下二元离散变量：是否节假日通用，是否无需预约，是否有礼品，是否仅限本人，是否有温馨提示，是否仅体验。这里我们画出数据集中的词云，观察其中出现次数较多的词语。

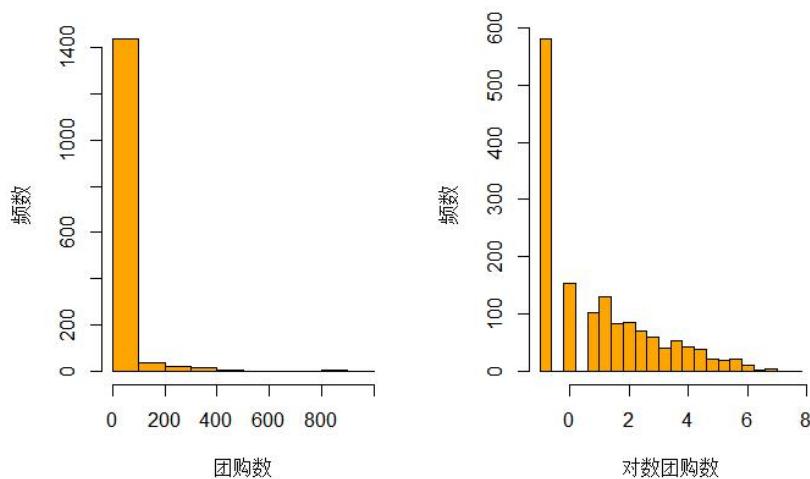


图一 数据集关键词词云

三、变量分析

3.1 因变量

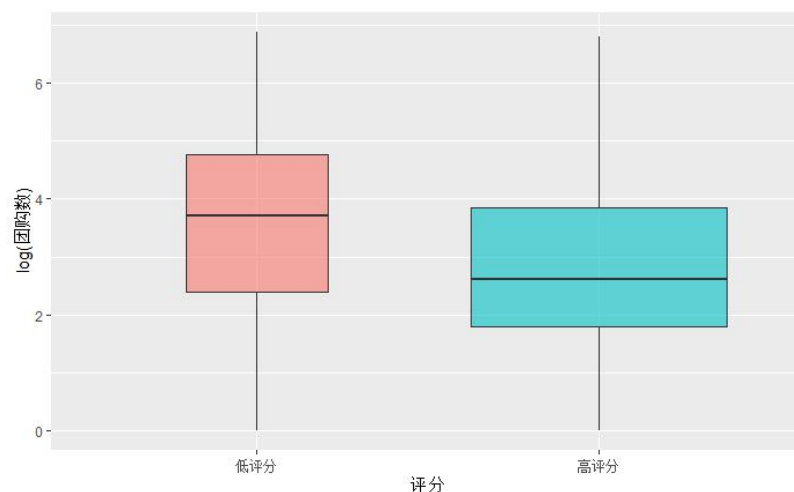
首先对因变量团购数进行分析并绘制直方图,从图中我们可以发现对团购数取对数更适合进行分析,对于团购数为0的数据,我们将其对数相应数值设置为-1 避免其影响拟合。



图二 团购数与对数团购数频数分布直方图

3.2 评分

然后再观察评分对 \log 团购数的影响，我们这里以 4.6 作为阈值区分“高评分”与“低评分”。

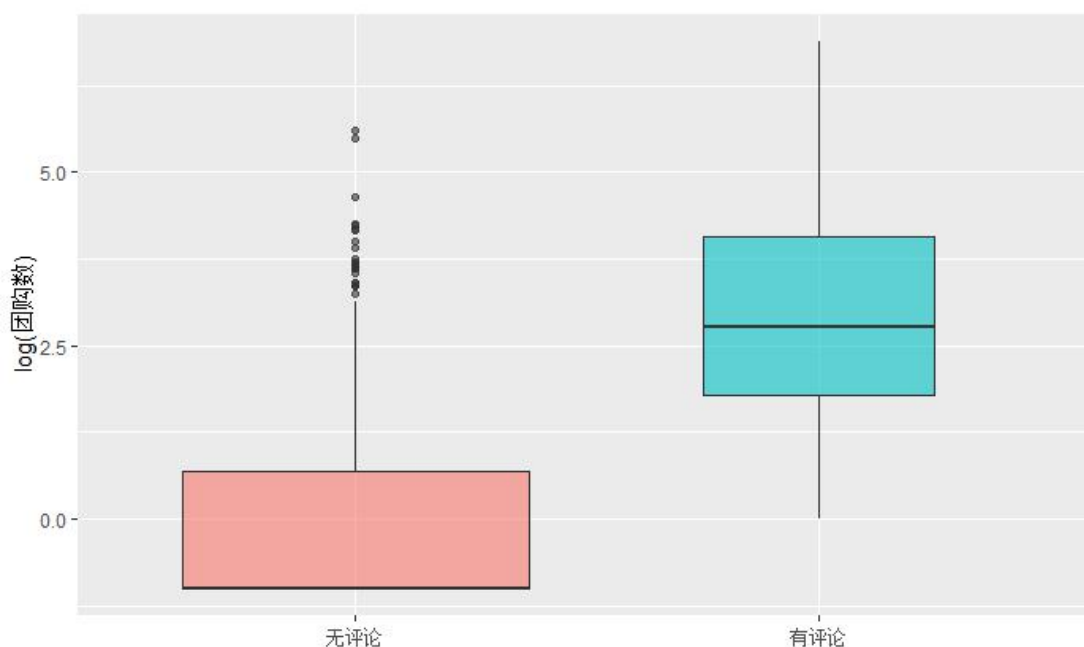


图三 评分——对数团购数箱线图

从图中可以看出，评分对对数团购数的影响并不如预期结果一般。通过观察原始数据，我们发现原始评分的中位数接近于 5，故我们推测评分本身会受到团购数的影响（如只有消费才可以进行评分），当团购数较小时，可能一两个人的高评分就会拉高总体评分；而团购数较大时，因为喜好不同导致总体评分反而不高。故虽然评分能部分反映健身项目的好坏，但与团购数本身存在复杂的相互影响。模型的线性拟合效果并不准确。

3.3 评论数

我们先对评论数的总体分布进行分析，发现评论数的中位数是 0（1031 个样本无评论），故我们通过是否有评论对样本进行划分并绘制箱线图。



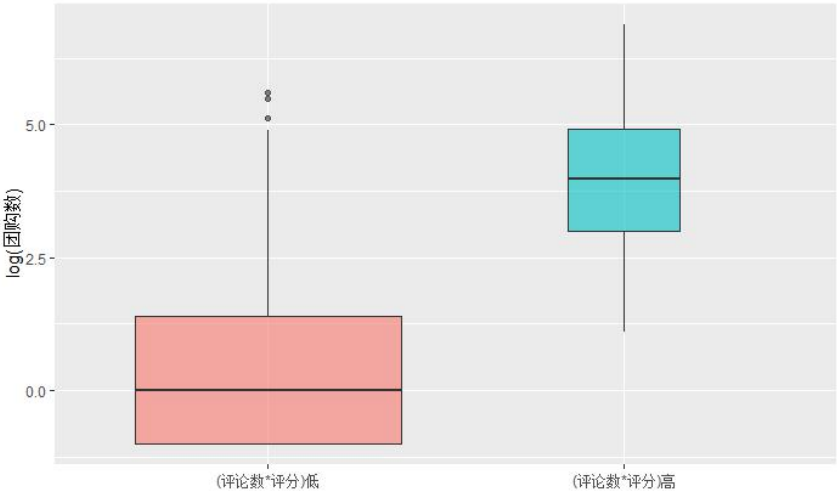
图四——有无评论-对数团购数箱线图

很明显有评论的团购项目能反映出其团购数也相对来说更多，这一点可以从两个方面解释，一是一个团购项目的购买量决定了评论基数，团购数越多说明这个项目相对来说质量

不错，也会让人更愿意去评论；二是代表了有评论的团购项目更吸引消费者，因为有之前的消费者的经验会让后来的消费者有所了解，更加放心购买产品（不考虑差评的话）。

3.4 评论数 X 评分

为了惩罚评论少且评分高的项目（去除水军或刷评论的影响），接下来我们考察评论数*评分对于团购数的影响。

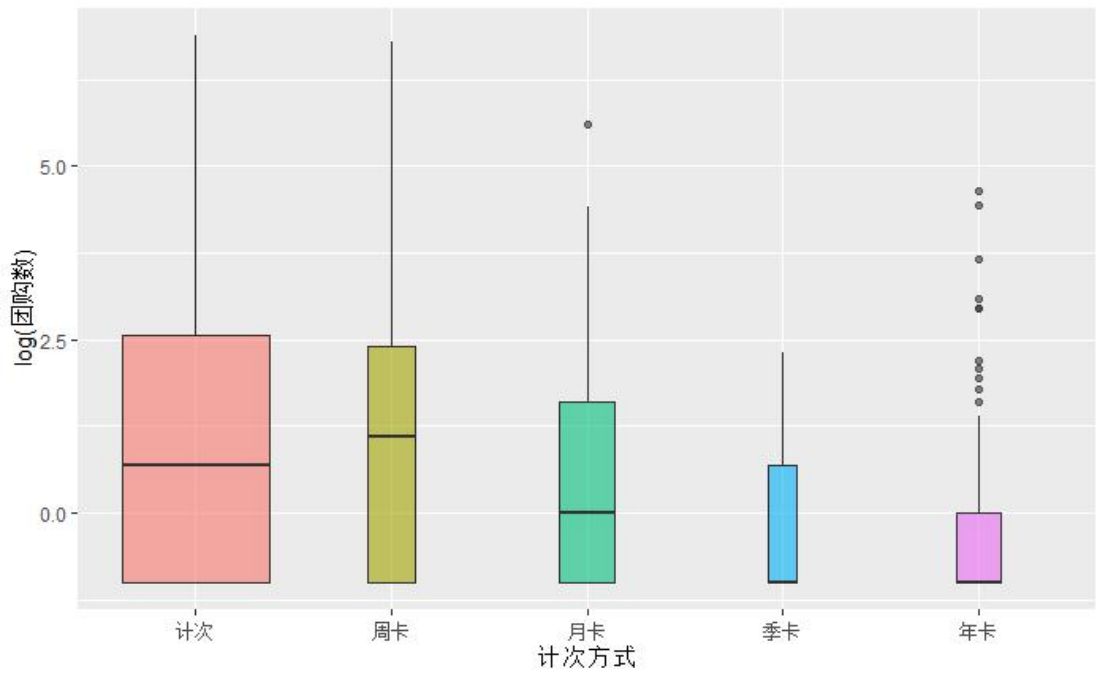


图五--评论数*评分-对数团购数箱线图

考察原始数据，我们发现，使用评论数*评分变量，对少评论数且评分高的项目进行惩罚后，明显看到（评论数*评分）高的项目的团购数远大于（评论数*评分）低的项目。说明我们使用评论数*评分的变量来描述性分析团购数有一定合理性。

3.5 计次方式

观察月卡，季卡，周卡，年卡，计次等不同的消费计次方式对对数团购数的影响。

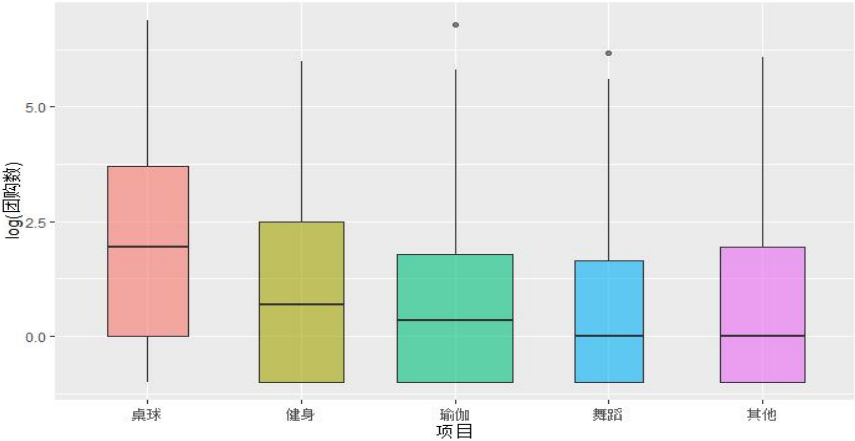


图六——计次方式--对数团购数箱线图

观察箱线图可以发现，年卡和季卡这样长时间的团购项目人数较少，说明大家更喜欢团购短期能消费完的项目，其实不光是团购，很多消费都有这样的特点，因为预定了的长期消费可能会有更多不确定因素，年卡，季卡的目标人群少且稳定的老客户，所以数量比较少，但同时方差也较小，而且月卡周卡和计次项目同一消费者可能在一段时间内多次购买，而年卡季卡不太可能。

而月卡周卡和次卡相对而言更受人群欢迎，因为时间比较短，方便大家随时购买，也会存在一人多次购买的情况，其中周卡更受欢迎，因为既免去了计次的麻烦，时限又不会太长。同时这几类中的对数团购数方差也更大（箱子高度看出），可能因为短期卡里活动会更加五花八门，差异较增大，导致因变量差异增大。

3.6 项目

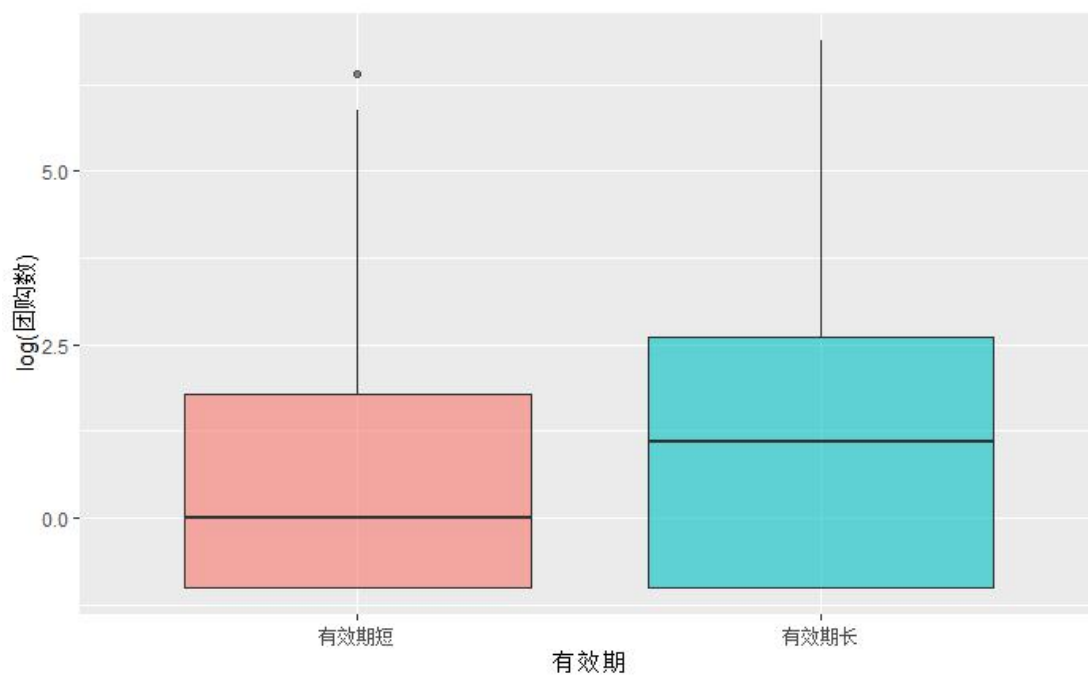


图七——不同项目对数团购数箱线图

此箱线图在一定程度上体现了不同项目的受欢迎程度，相对而言桌球要更受欢迎，团购量要更多，说明大家更喜欢桌球这样的项目（也可能是相对其他项目来说包场打台球比不团购直接去计时划算），而两大主流健身和瑜伽的团购量也相对而言稍微高于舞蹈和其他项目。

3.7 有效期

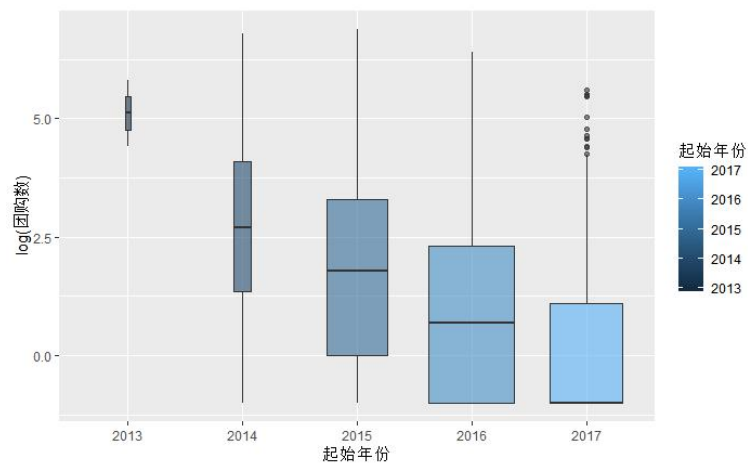
我们用有效期长短对团购数的对数画箱线图



图八——有效期长短-对数团购数箱线图

有图可得，箱线图符合我们的预期，有效期长的团购量更多，因为可以有更多的自由选择使用的时间。

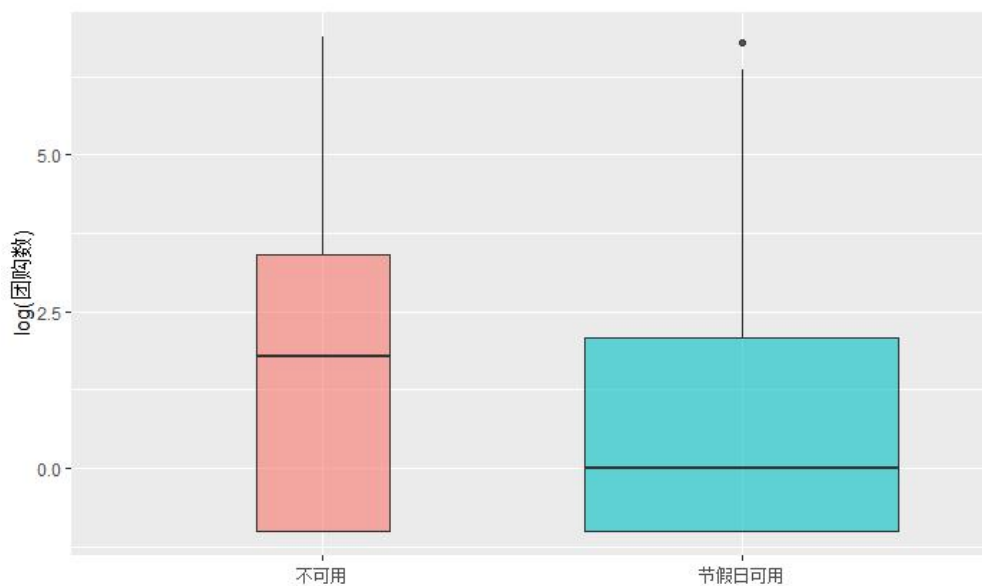
3.8 起始年份



图九——起始年份-对数团购数箱线图

出现这样现象的原因可能是时间越早的项目其在平台上出售的时间就更长，也就更可能被更多人购买，比如 2017 年很多项目其实都还没有人购买，可能是刚登陆平台的原因；

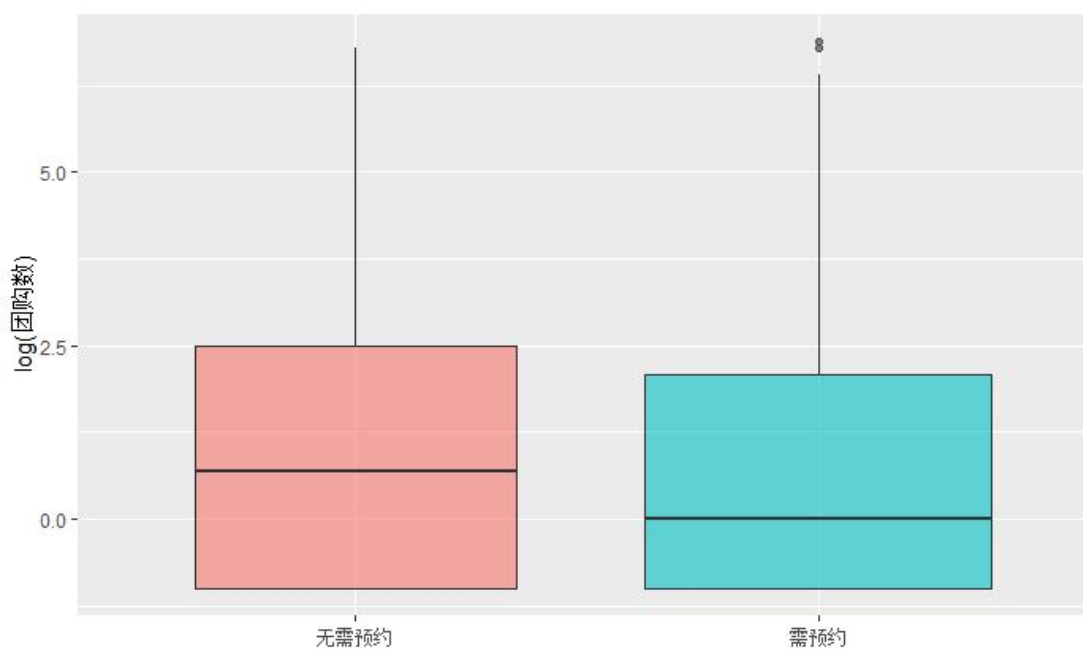
3.9 节假日可用



图十——节假日可用性-对数团购数箱线图

图中可以看出，两种可用类型的差距并不明显。出现这样的结果应该也可以用样本数的问题进行解释，因为现在很多团购项目都是节假日可用的了，所以项目多分散了消费，每个项目的团购量相对较少。

3.10 预约情况

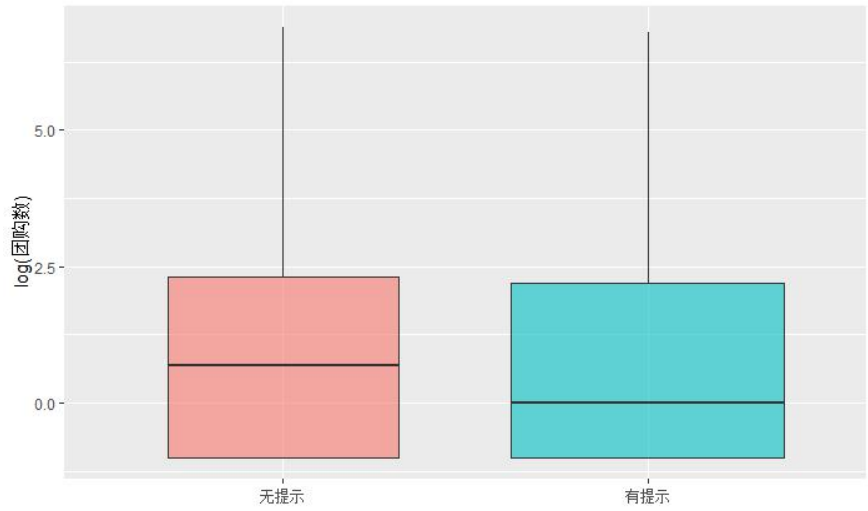


图十一——是否需要预约-对数团购数箱线图

通过观察箱子的宽度我们可以看出，在这样的划分里，两类样本相对均匀，我们得到的结果也符合我们的预期，因为无需预约直接用的团购项目更加自由，也更能吸引消费者，团购数量相对更多。

3.11 是否有温馨提示

在观察原始数据时,我们发现有许多商家会在温馨提示中加入一些宣传鼓励性的语句或提供一些赠品,例如矿泉水等,为观察顾客是否会对这类赠品感兴趣以及被这些话语鼓舞,我们以有无温馨提示对对数团购数画箱线图。

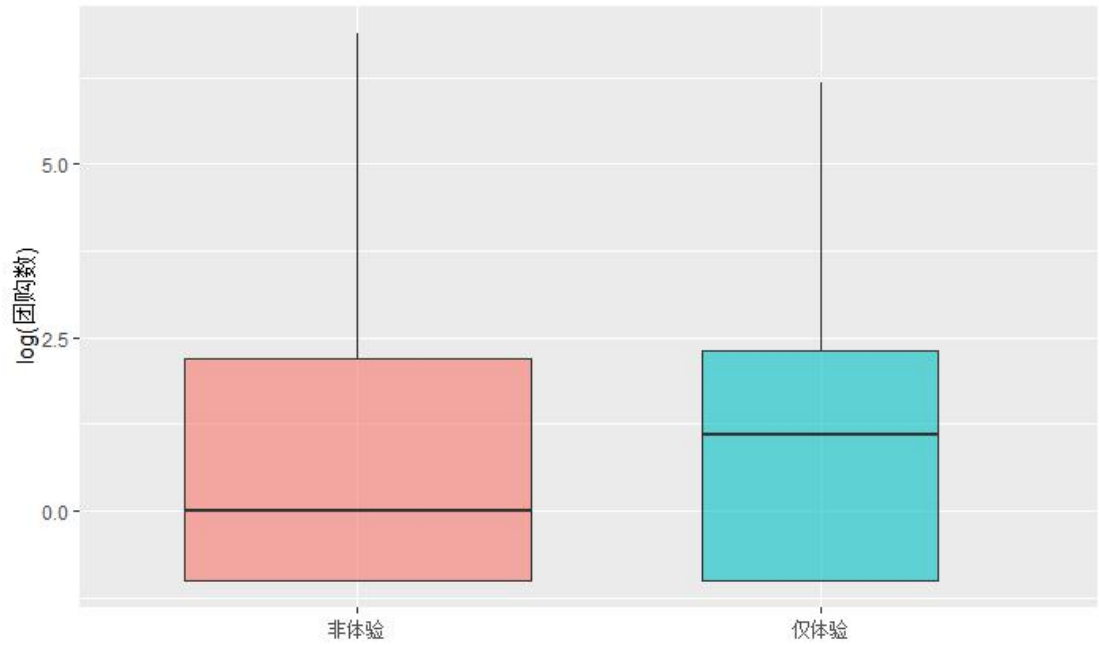


图十二——有无温馨提示-对数团购数箱线图

从箱线图中箱子的位置可以看出进行划分以后两类样本的对数团购数差不太多,说明温馨提示中的赠品对消费者的吸引力不明显,消费者更加关注于团购项目本身的内容和质量。

3.12 是否为体验活动

在观察原始数据时,还有一些店家会以“体验”作为噱头吸引新客户来进行初次体验,我们这里也尝试着通过绘制箱线图来研究“体验活动”对消费者的吸引力。



图十三——是否仅体验-对数团购数箱线图

体验活动的受欢迎程度相对较高一些,可能是因为大家团购一般都是因为想要第一次去试试,所以体验的选择很多,如果要长期参加的话可能不会选择团购而是去实体店里办卡。

四、模型拟合

4.1 初步拟合

我们先以团购数为因变量，使用上文所述各变量对其进行线性拟合。

```
Call:
lm(formula = 团购数 ~ 评分 + 评论数 + 月卡 + 年卡 + 季卡 + 周卡 +
    有效天数 + 起始年份 + 节假日通用 + 无需预约 + 有礼品 + 仅限本人 +
    是否有温馨提示 + 仅体验 + 瑜伽 + 桌球 + 舞蹈 + 健身 + 缺失数,
    data = tuangou1)

Residuals:
    Min       1Q   Median       3Q      Max
-299.14  -12.83   -5.89    2.77   419.37

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.292e+04  3.092e+03   4.180 3.08e-05 ***
评分        -4.521e+00  2.924e+00  -1.546  0.1223
评论数       8.649e+00  1.789e-01  48.345 < 2e-16 ***
月卡        -6.188e+00  3.649e+00  -1.696  0.0901 .
年卡        -9.658e+00  4.460e+00  -2.166  0.0305 *
季卡        -1.011e+01  6.611e+00  -1.530  0.1262
周卡        -8.142e-01  4.227e+00  -0.193  0.8473
有效天数    -2.506e-03  2.766e-03  -0.906  0.3650
起始年份    -6.383e+00  1.534e+00  -4.161 3.35e-05 ***
节假日通用  -1.787e+01  3.024e+00  -5.909 4.26e-09 ***
无需预约    -2.638e+00  2.318e+00  -1.138  0.2554
有礼品      -8.194e+00  4.767e+00  -1.719  0.0858 .
仅限本人     5.880e+00  5.830e+00   1.009  0.3133
是否有温馨提示 -1.697e+01  2.565e+00  -6.615 5.15e-11 ***
仅体验      -1.098e+00  2.519e+00  -0.436  0.6630
瑜伽        -5.505e+00  3.207e+00  -1.716  0.0863 .
桌球        3.533e+00  3.991e+00   0.885  0.3761
舞蹈        -1.280e+00  3.980e+00  -0.322  0.7478
健身        4.296e+00  3.511e+00   1.224  0.2212
缺失数      -1.843e+01  1.023e+00 -18.012 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

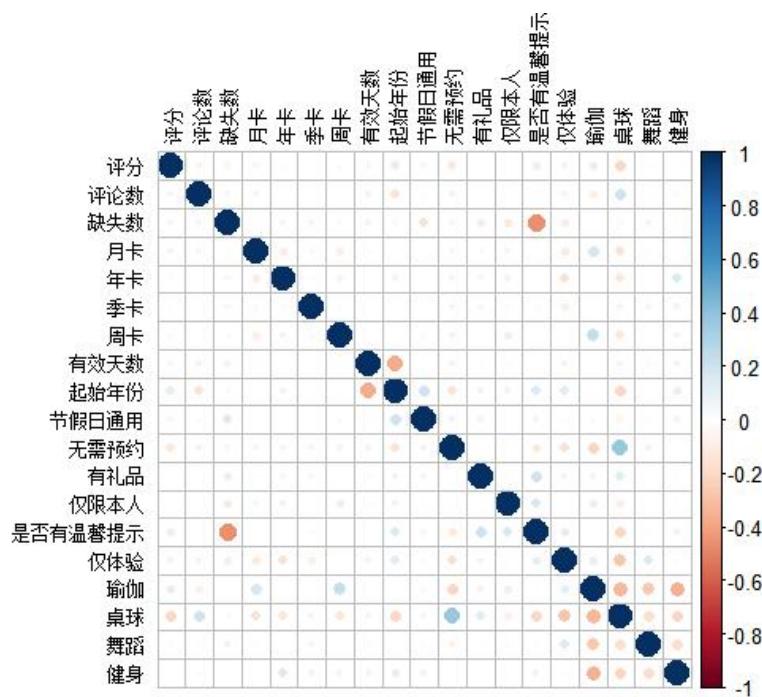
Residual standard error: 40.29 on 1497 degrees of freedom
Multiple R-squared:  0.669,    Adjusted R-squared:  0.6648
F-statistic: 159.2 on 19 and 1497 DF,  p-value: < 2.2e-16
```

图十四——初步线性拟合结果

通过初步的拟合，我们可以发现几个自变量与因变量之间似乎存在一定线性相关关系，如评论数、起始年份、是否节假日通用、温馨提示和数据的缺失值数量；整个线性模型的拟合效果不错，但其中也有许多影响不明显的变量。

4.2 变量间的相互影响

于此同时，我们考虑到不同的变量之间可能存在相互影响，如评分与评论数，是否节假日通用与是否需要提前预约等，故我们对自变量相关性两两作图，得到图表如下。



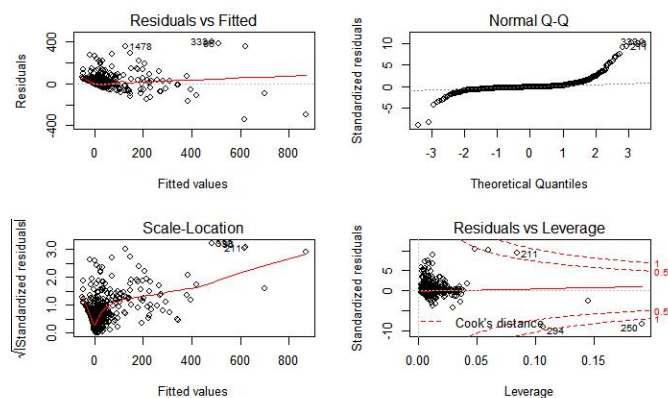
图十五——变量间相关性图

对变量的相关性进行绘图后，我们发现：

- 1) 瑜伽、桌球、舞蹈、健身项目间都是负相关的，显然是因为团购项目只能有一个项目的属性。同理，月卡、年卡、季卡、周卡间也是负相关的，可以看到相关系数都是红色。
- 2) 起始年月与有效天数呈比较明显的负相关，起始年份越早的、有效天数越多，这可能是由于我们在捕捉数据的时刻（2017 年 8 月），所有项目都是可以购买的状态，因而起始年份早的项目就自然有效天数不会短。
- 3) 是否有温馨提示与缺失值有明显的负相关，有温馨提示的缺失值较少。说明愿意写温馨提示的勤快商家倾向于展示更多公开的数据。
- 4) 桌球和预约有较明显的正相关。说明桌球项目一般比较自由，打桌球一般无需预约（这里将无需预约设置为“1”变量）。

4.3 模型优化

在得到了相关信息之后，我们先去除一些无关变量，之后添加评论数*评分这一变量，去除异常点后以 AIC 准则对模型进行优化，得到新的拟合结果。



图十六——异常值检测，由图可得需要去除 211，250，294，233 三个异常点

```
Call:
lm(formula = 团购数 ~ 评论数 + 月卡 + 年卡 + 季卡 + 起始年份 +
    节假日通用 + 有礼品 + 是否有温馨提示 + 瑜伽 + log(评分 *
    评论数) + 缺失数, data = tuangou3)

Residuals:
    Min       1Q   Median       3Q      Max
-342.86  -11.63   -6.02    1.31   400.74

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    84.0156     7.2791  11.542 < 2e-16 ***
评论数         9.9847     0.2582  38.670 < 2e-16 ***
月卡          -5.5329     3.3952  -1.630  0.10339
年卡          -8.0350     4.0978  -1.961  0.05008 .
季卡          -9.9263     6.3239  -1.570  0.11671
起始年份      -5.6077     1.3767  -4.073 4.88e-05 ***
节假日通用    -14.8699     2.9853  -4.981 7.05e-07 ***
有礼品        -6.8700     4.6145  -1.489  0.13675
是否有温馨提示 -12.2343     2.5068  -4.880 1.17e-06 ***
瑜伽          -6.6556     2.2030  -3.021  0.00256 **
log(评分 * 评论数) -15.7698    2.2958  -6.869 9.45e-12 ***
缺失数       -14.6599     1.1250 -13.031 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 39.69 on 1501 degrees of freedom
Multiple R-squared:  0.6773,    Adjusted R-squared:  0.6749
F-statistic: 286.4 on 11 and 1501 DF, p-value: < 2.2e-16
```

图十七——模型优化后的结果

相比较之后可以发现，优化后的模型变量减少后拟合程度维持稳定，且变量的相关性均有不同程度的提高，最后得出与团购数相关的变脸有：评论数与评分、起始年份、节假日通用、是否有礼品赠送，是否有温馨提示，种类为桌球，店面介绍的缺失信息数。

我们仍对各变量的系数进行了分析，大部分的变量系数都符合现实生活中的情况，对于评论数，评论数较高的样本团购数也相应较高。而相对于次卡而言，月卡，年卡，季卡对消费者的吸引力逐渐下降，这也反映了消费者在健身上更倾向于短期消费而不一定有毅力坚持长期锻炼。在起始年份上，也反映出越晚上线的项目团购数越低的特征。对缺失数也是如此，在资料上缺失越多的商家对消费者的吸引力也越低。而与我们预期不符的是，温馨提示、小礼品的赠送和在节假日是否可以使用卡反而会降低团购数，这种违背现实情况的取值可能是数据数不够多所致。而当我们对评论数乘以评分并取对数后系数由正转负，也反映了在健身房数据中“评论数高的项目评分低”这一有趣的现象。

另一方面，我们也对团购数的对数进行的类似的处理，结果如图，可以发现使用对数进行线性建模的效果并不如直接建模。


```

Call:
lm(formula = log团购数 ~ 评分 + 评论数 + 月卡 + 年卡 + 季卡 +
    周卡 + 有效天数 + 起始年份 + 节假日通用 + 无需预约 + 有礼品 +
    仅限本人 + 是否有温馨提示 + 仅体验 + 瑜伽 + 桌球 + 舞蹈 +
    健身 + 缺失数, data = tuangou5)

Residuals:
    Min       1Q   Median       3Q      Max
-3.8757 -0.7698 -0.0873  0.7092  4.1433

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.8804505   0.4852164   10.058 < 2e-16 ***
评分        -0.0439279   0.0821294   -0.535 0.592903
评论数       0.0792782   0.0049892   15.890 < 2e-16 ***
月卡        -0.3656582   0.1461882   -2.501 0.012585 *
年卡        -0.9233714   0.2386846   -3.869 0.000119 ***
季卡        -1.3528547   0.3436337   -3.937 9.01e-05 ***
周卡        -0.0305166   0.1542783   -0.198 0.843253
有效天数    -0.0002038   0.0001051   -1.939 0.052883 .
起始年份    -0.2243239   0.0593219   -3.781 0.000168 ***
节假日通用  -0.6910642   0.1066850   -6.478 1.68e-10 ***
无需预约    -0.0976089   0.0901118   -1.083 0.279066
有礼品      -0.3217671   0.2071707   -1.553 0.120805
仅限本人    -0.1436436   0.1969802   -0.729 0.466087
是否有温馨提示 -0.2314757   0.0954170   -2.426 0.015502 *
仅体验      -0.1442505   0.0968878   -1.489 0.136946
瑜伽        -0.4957725   0.1360236   -3.645 0.000286 ***
桌球        -0.0484317   0.1500613   -0.323 0.746976
舞蹈        -0.3118365   0.1707053   -1.827 0.068131 .
健身        0.0973688   0.1422095    0.685 0.493752
缺失数      -0.3841488   0.0325324  -11.808 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.089 on 757 degrees of freedom
Multiple R-squared:  0.4413,    Adjusted R-squared:  0.4273
F-statistic: 31.47 on 19 and 757 DF,  p-value: < 2.2e-16

```

图十八——对团购数的对数建模的结果

然而，我们是否可以仅凭 R-square 的大小去决定模型的好坏呢？为验证这一点，我们进行了一个有趣的实验。

```

Call:
lm(formula = y1 ~ x1 + x2)

Residuals:
    Min       1Q   Median       3Q      Max
-3.0685 -0.6652  0.0178  0.6384  3.5258

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.58310     5.30555  -1.052   0.293
x1          -0.96801     0.01276  -75.862 <2e-16 ***
x2           1.98185     0.01297  152.823 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

Residual standard error: 1.019 on 997 degrees of freedom
Multiple R-squared:  0.9591,    Adjusted R-squared:  0.959
F-statistic: 1.169e+04 on 2 and 997 DF,  p-value: < 2.2e-16

Call:
lm(formula = y2 ~ x1 + x2)

Residuals:
    Min       1Q   Median       3Q      Max
-9.0967 -2.1663  0.0206  2.0703  9.4785

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.32937     15.96963  -0.083   0.934
x1          -0.99051     0.03841  -25.789 <2e-16 ***
x2           1.99335     0.03903   51.067 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

Residual standard error: 3.068 on 997 degrees of freedom
Multiple R-squared:  0.7238,    Adjusted R-squared:  0.7232
F-statistic: 1306 on 2 and 997 DF,  p-value: < 2.2e-16

n <- 1000
z1 <- rnorm(n,200,2)
z2 <- rnorm(n,200,2)
x1 <- z1 + z2
x2 <- z2 + rnorm(n,200,2)
y1 <- -1*x1 + 2*x2 + rnorm(n,0,1)
y2 <- -1*x1 + 2*x2 + 3*rnorm(n,0,1)
fit1 <- lm(y1 ~ x1+x2)
fit2 <- lm(y2 ~ x1+x2)

```

图十九——探究模型扰动对于 Adjust R-square 的影响

两个回归的差别仅仅在于，第二个方程的扰动项的方差是第一个的 3 倍，导致 R^2 从 95.8% 下降到了 73.7%，于是我们可以得到一个结论：扰动值的差别会对 R^2 有比较大的影

响。另一方面我们从 y_1 减去一个单位的 x_1 ，发现回归方程几乎不发生变化。

```
Call:
lm(formula = y3 ~ x1 + x2)

Residuals:
    Min       1Q   Median       3Q      Max
-3.0685 -0.6652  0.0178  0.6384  3.5258

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.58310    5.30555   -1.052    0.293
x1          -1.96801    0.01276  -154.230 <2e-16 ***
x2           1.98185    0.01297   152.823 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.019 on 997 degrees of freedom
Multiple R-squared:  0.9699,    Adjusted R-squared:  0.9698
F-statistic: 1.605e+04 on 2 and 997 DF,  p-value: < 2.2e-16

Call:
lm(formula = y1 ~ x1 + x2)

Residuals:
    Min       1Q   Median       3Q      Max
-3.0685 -0.6652  0.0178  0.6384  3.5258

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.58310    5.30555   -1.052    0.293
x1          -0.96801    0.01276   -75.862 <2e-16 ***
x2           1.98185    0.01297   152.823 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.019 on 997 degrees of freedom
Multiple R-squared:  0.9591,    Adjusted R-squared:  0.959
F-statistic: 1.169e+04 on 2 and 997 DF,  p-value: < 2.2e-16

y3 <- y1-x1
fit3 <- lm(y3~x1+x2)
summary(fit3)
summary(fit1)
```

图二十一——探究因变量变化对于 Adjust R-square 的影响

故此，我们认为单纯凭借 Adjust R-square 的大小去评价模型的优劣是不恰当的，对我们的模型也同此理。

五、 外部资料与模型的改进方向

在对已知数据集的给定数据集进行处理时，我们还对复旦大学附近健身房的情况在美团网上进行了搜索，主要体现为评分普遍偏高，评论数较低且评论数往往与评分呈现负相关。在搜集资料的同时，我们发现在团购网站上有“只看有图片的评论”这一选项，当消费者愿意付出更多的时间撰写评论时，往往也证明了这一份评论是更有价值的，所以在之后的模型优化和数据抓取中可以考虑加入这一因素。

另一方面，对“评分”、“价格”变量的缺失值我们采用了取平均值的处理方式，这一点是否合理仍有待商榷，另对于健身这一略显小众的消费项目，如何处理评分人数较少情况下的评分虚高也是模型接下来的改进方向之一。以及对一些负系数与现实情况不符的情况，目前还没有找到好的解决方法。

六、 小组分工

数据第一次预处理： 李泽君
报告第一稿：赵雅滢、李泽君
数据第二次画图：谢炳辉、赵雅滢
撰写报告最终稿：赵雅滢、钟诚
演讲：谢炳辉

数据第一次画图：李泽君 钟诚
数据第二次与预处理：谢炳辉
收集资料：王维实、钟诚
制作 PPT 和演讲提纲：王维实

