

The Analysis of Consumers' Housing Purchase Intention Based on Apriori Association Rule Mining

—A Case Study of North Kunming District

Mengdi Zhai

Yunnan University of Finance and Economics, Kunming Yunnan
Email: 540532394@qq.com

Received: Jul. 28th, 2016; accepted: Aug. 13th, 2016; published: Aug. 18th, 2016

Copyright © 2016 by author and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Apriori association rule was used to find out the key factors of consumers' housing purchase intention. By analyzing the frequent item sets, and using consumers' personal factors as the preceding paragraph, the diversity of different housing purchase intention was analyzed, which was also predicted through understanding of users' personal characteristics. Then this paper provided suggestions for developers who need product positioning, as well as methods of precision marketing for retailers.

Keywords

Consumers, Purchase Intentions, Association Rules, Personal Factors, Frequent Item Sets

基于Apriori关联规则挖掘的消费者购房意向分析

—以昆明市北市区为例

翟梦迪

云南财经大学, 云南 昆明
Email: 540532394@qq.com

摘要

利用Apriori关联规则挖掘出消费者购房意向中重点考虑的因素，并且通过分析频繁项集，以消费者个人因素为前项，分析不同群体的多元化购房意向，通过了解用户个人特征推测其购房意向，为开发商定位产品提供建议，为销售商精准营销提供方法。

关键词

消费者，购房意向，关联规则，个人因素，频繁项集

1. 研究背景及现状

近年来，开发商数量不断快速增长，同时一些问题也开始显现出来。一方面房价过高让不少国人望“房”兴叹，一方面住宅滞销也屡屡出现。越来越多的业内人士意识到不能盲目兴建楼盘，而是要将开发与消费需求联系起来。消费者需求的研究越来越被重视。研究者们也指出房地产开放商要想使目标顾客的需要得到满足，就必须从顾客的角度出发来思考问题。

向长江等(2002)通过对星汇园潜在购房者的问卷调查及其统计分析，探讨了品牌形象、本体素质、户型设计和地段等因素对潜在购房者感知价值的影响，以及户型设计、室外环境、感知价值等因素对购买意向的影响[1]；吴殿廷等(2007)以北京市和瓦房店市为例，对比了大都市和中小城市居民购房决策影响因素的异同点，指出大都市居民更重视物业和生态环境，而中小城市居民更重视商业环境[2]；于丹等(2007)以大连市为例，利用方差分析和因子分析的方法讨论了影响用户购买意向的因素[3]；蒋绪军(2014)结合客户购买决策模型和商品房客户购房实际情况，构建商品房客户购买决策一般模型[4]；李莹等(2015)以广东省的“85后”消费者作为研究对象，调查他们的购房意愿以及影响因素，在购房决策所考虑的房屋属性方面，“85后”消费者最重视房屋自身的地理位置、价格以及交通这三方面的价值，表现出较强的实用性导向[5]。

2. 问题提出

对于消费者购房影响的研究中，多是基于描述给出了单个因素对结果的影响，而并没有分析出影响因素与结果之间的关联关系，也没有刻画组合因素对消费者购房意向的影响。为更加全面的反映不同消费者群体的购买意向，本文利用 Apriori 关联规则挖掘的方法，以消费者个人因素为前项，以消费者购房需求为后项。通过住房自身特征、住房所在小区特征和小区周边环境特征三方面表征消费者的购房需求。进而挖掘出具有不同个人特征的消费者群体与不同购房需求之间的关系。为开发商的产品定位和消费者需求探索提供合理可靠的建议。

3 理论基础

1) Apriori 算法基本思想

首先找出所有的频繁项集，这些项集出现的频繁性至少和预定义的最小支持度一样，然后由频集产生强关联规则，这些规则必须满足最小支持度和最小可信度，再使用第 1 步找到的频集产生期望的规则，产生只包含集合项的所有规则，一旦这些规则被生成，那么只有那些大于用户给定的最小可信度的规则

才被留下来。

2) 算法分析步骤

第一步：通过迭代的方法，找出事物数据库中的全部频繁项集，即支持度大于或等于用户设定的阈值的项集；

第二步：利用频繁项集构造出满足最小置信度的规则[6]。

4. 实例分析

4.1. 数据来源与预处理

数据来自 2015 年昆明市北市区消费者住宅项目调查问卷，问卷有效样本量为 200。在对消费者进行购房意向的关联规则挖掘时，以消费者个人特征为前项，以购房意向为后项，将各项表征列示如表 1、表 2。

4.2. 个人因素与住房自身特征之间的关联挖掘

1) 个人因素与购房类型选择之间的关系

通过分析图 1 购房类型的频繁项集可以发现样本中接近 100% 的消费者集中倾向于选择现房，而非期房或者二手房。个人因素在这个问题中影响并不显著，不过这也是目前购房的整体趋势。

2) 个人因素与可接受房价之间的关系

通过分析图 2 可以发现，可接受房价的频繁项集分别为 $A_{13} = 1$ ， $A_{13} = 2$ ， $A_{13} = 3$ ，即价格在 5000~7000 元/m²，7001~9000 元/m² 和 9001~10,000 元/m²，那么选择房价小于 10,000 元/m² 的情况较为普遍。

进行 Apriori 关联规则挖掘，以个人因素为前项，以可接受房价区间为后项，按照提升度降序排序，选择提升度最高项得到表 3。

可以看出：第一，个人特征为女、无工作、家庭收入为低收入并且选择购买价格在 5000~7000 元/m² 的房屋的所有消费者的 1%；而在上述个人特征下购买价格在 5000~7000 元/m² 的房屋的所有消费者占比 100%；这个比例是总体中购买价格在 5000~7000 元/m² 的房屋的所有消费者比例的 1.818 倍。第二，个人特征为男、无工作、家庭收入为低收入并且选择购买价格在 7001~9000 元/m² 的房屋的所有消费者的 1%；而在上述个人特征下购买价格在 7001~9000 元/m² 的房屋的所有消费者占比 100%；这个比例是总体中相应比例的 3.125 倍。通过和(1)对比可知，同样特征下，相较于女性，男性出于综合考虑会选择面积较大的房屋。第三，个人特征为青年、男、有固定工作、家庭成员小于等于 2 人、个人收入属高收入、家庭收入属高收入并且购买 9001~10,000 元/m² 的房屋的所有消费者占比 1.5%；在上述个人特征下购买 9001~10,000 元/m² 的房屋的所有消费者占比 42.8%；这个比例是总体中相应比例的 4.762 倍。

3) 个人因素与购房户型选择之间的关系

由图 3 可以看出，购房户型的频繁项集分别为 $A_{14} = 3$ ， $A_{14} = 6$ ， $A_{14} = 7$ ，所以消费者倾向选择的户型分别是两室两厅一卫、三室两厅一卫和三室两厅两卫。

进行 Apriori 关联规则挖掘，以个人因素为前项，以购房户型选择为后项，按照提升度降序排序，选择提升度最高项得到表 4。

可以看出：第一，个人特征为青年、家庭成员小于等于 2 人、个人收入为低收入、家庭收入为高收入并且选择购买户型为两室两厅一卫房屋的所有消费者的 1%；而在上述个人特征下购买两室两厅一卫房屋的所有消费者占比 28.5%；这个比例是总体中购买价格在两室两厅一卫房屋的所有消费者比例的 1.731 倍。第二，个人特征为中老年、女性、家庭成员小于等于两人并且选择购买户型为三室两厅一卫房屋的

Table 1. Personal factors

表 1. 个人因素指标

指标	变量名	变量取值	变量说明
年龄	B1	0,1	0: 青年 1: 中老年
性别	B2	0,1	0: 女 1: 男
职业	B3	0,1	0: 有固定工作 1: 没有固定工作
家庭居住状态	B4	0,1	0: 家庭成员小于等于 2 人 1: 家庭成员大于 2 人
个人月收入	B5	0,1	0: 个人收入较低 1: 个人收入较高
家庭月收入	B6	0,1	0: 家庭收入较低 1: 家庭收入较高

Table 2. Index of purchase intention and its characterization

表 2. 购房意向指标及表征

	表征	变量	变量解释
购房意向	住房自身特征	A7	A7: 购房类型选择
		A13	A13: 可接受房价区间
		A14	A14: 购房户型选择
		A15	A15: 住房面积选择
	住房所在小区特征	A10	A10: 小区公众配套设施
		A11	A11: 小区物业管理需求
		A12	A12: 小区周边生活配套
	小区周边环境特征		

Table 3. The association rules table of individual factors and the accepted price range

表 3. 个人因素与可接受房价区间的关联规则表

前项	方向	后项	支持度	置信度	提升度
{B2 = 0, B3 = 0, B6 = 0}	≥	{A13 = 1}	0.010	1	1.818
{B2 = 1, B3 = 0, B6 = 0}	≥	{A13 = 2}	0.010	1	3.125
{B1 = 0, B2 = 1, B3 = 1, B4 = 1, B5 = 1, B6 = 1}	≥	{A13 = 3}	0.015	0.428	4.762

注: 对于置信度为 1 的情况是由于样本量较小。

Table 4. The association rules table of personal factors and the purchase house type choice

表 4. 个人因素与购房户型选择的关联规则表

前项	方向	后项	支持度	置信度	提升度
{B1 = 0, B4 = 0, B5 = 0, B6 = 1}	≥	{A14 = 3}	0.010	0.285	1.731
{B1 = 1, B2 = 0, B4 = 0}	≥	{A14 = 6}	0.010	1	4.545
{B1 = 1, B4 = 0, B6 = 1}	≥	{A14 = 7}	0.010	1	4.166

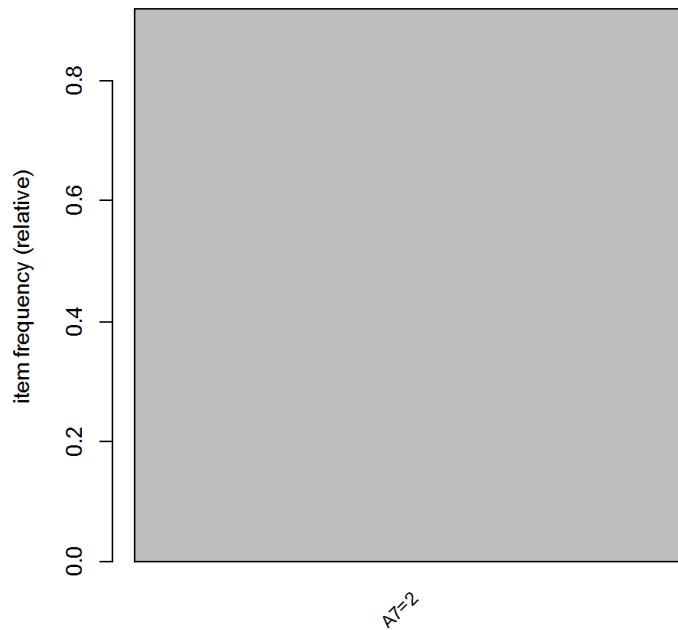


Figure 1. Frequent item sets of purchase type selection

图 1. 购房类型选择频繁项集

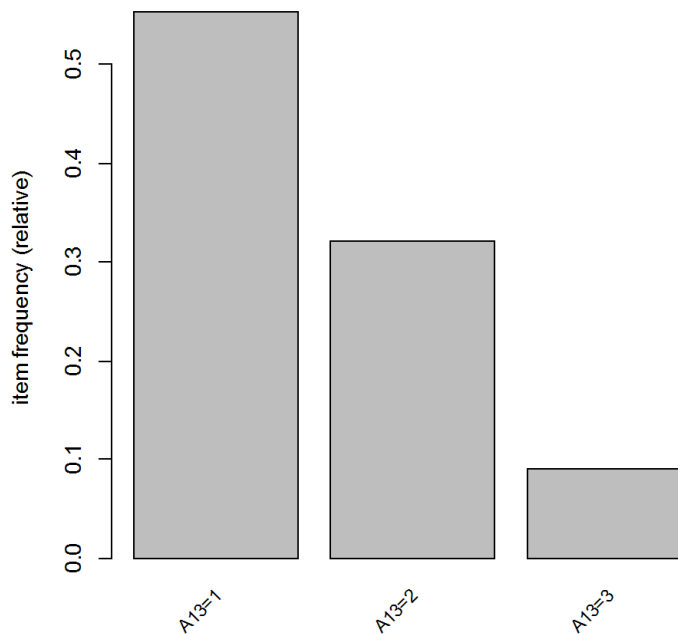


Figure 2. Frequent item sets of acceptable price

图 2. 可接受房价的频繁项集

消费者占有所有消费者的 1%；而在上述个人特征下购买三室两厅一卫房屋的消费者比例是总体中购买价格在两室两厅一卫房屋的消费者比例的 4.166 倍。第三，个人特征为中老年、家庭成员小于等于两人、家庭总收入较高并且选择购买户型为三室两厅两卫房屋的消费者占有所有消费者的 1%；而在上述个人特征下购买三室两厅两卫房屋的消费者比例是总体中购买价格在两室两厅两卫房屋的消费者比例的 4.545 倍。与(1)、(2)比较可知这类消费者多为未来购房计划，可能为三代同住做准备。

4) 个人因素与购房面积选择之间的关系

通过分析图 4 可以发现，购房面积选择的频繁项集分别为 51~70 m²、71~90 m² 和 91~120 m²，那么选择房屋面积在 51~120 m² 的情况较为普遍。

进行 Apriori 关联规则挖掘，以个人因素为前项，以购房面积选择为后项，按照提升度降序排序，选择提升度最高项得到表 5。

由表可知：第一，个人特征为青年、家庭成员小于等于 2 人、个人收入情况为较高收入并且选择 51~70 m² 的房屋的消费者群体占比 1%；在上述个人特征下购买 51~70 m² 的房屋的消费者群体占比 33.3%；这

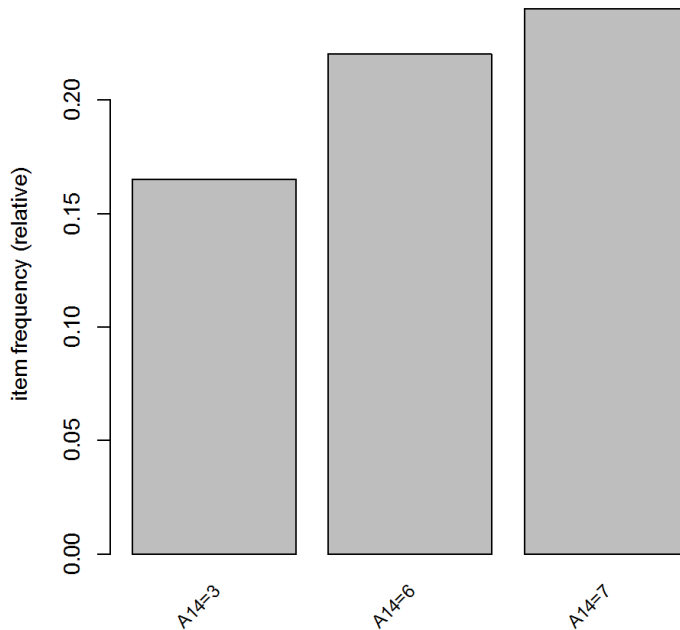


Figure 3. Frequent item sets of purchase house type choice

图 3. 购房户型选择频繁项集

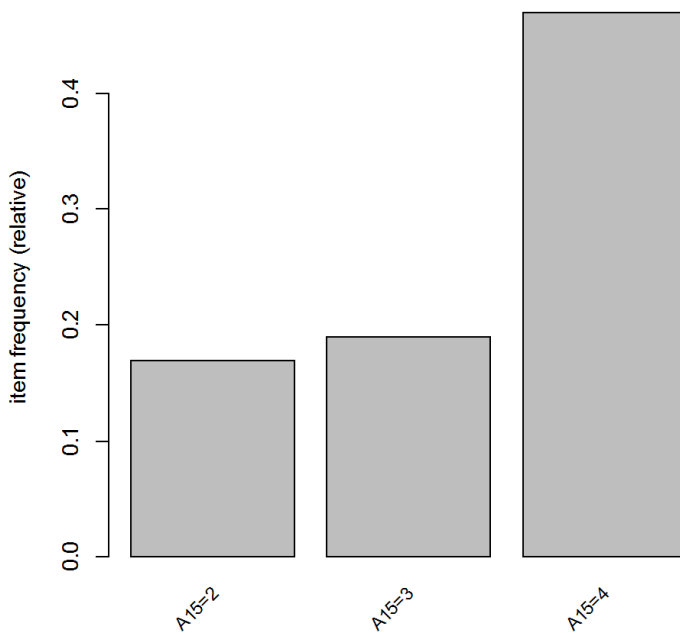


Figure 4. Frequent item sets of purchase area

图 4. 购房面积选择的频繁项集

个比例是总体相应比例的 1.961 倍。第二，个人特征为青年、性别女、有固定工作、家庭成员大于等于 2 人并且选择 71~90 m² 的房屋消费者占比 2%；在上述个人特征下购买 71~90 m² 的房屋消费者占比 66.7%；这个比例是总体相应比例的 3.508 倍。第三，个人特征为中老年、无固定工作、家庭收入较低并且选择 91~120 m² 的房屋群体占比 1%；在上述个人特征下购买 91~120 m² 房屋消费者比例是总体相应比例的 2.127 倍。从这一点可以推测，这类消费者可能是离退休人员为子女或为晚年准备住房。

4.3. 个人因素与住房所在小区特征之间的关联挖掘

1) 个人因素与小区周边公众配套设施选择之间的关系

通过分析图 5 可知，在对小区周边公众配套设施的选择中游泳池、健身设施和商场超市为频繁项集。对于前两项高频出现的情况，体现出了新时代消费者的新需求。

进行 Apriori 关联规则挖掘，以个人因素为前项，以小区配套设施为后项，按照提升度降序排序，选择提升度最高项得到表 6。

由表可知：第一，个人特征为中老年、家庭成员小于等于 2 人、家庭收入较高并且希望有游泳池的群体占比 1%；在上述个人特征下希望有游泳池的消费者例是总体相应比例的 6.25 倍。第二，个人特征为中老年、女性、家庭成员小于等于 2 人并且希望有健身设施的群体占比 1%；在上述个人特征下希望有健身设施的消费者比例是总体相应比例的 3.508 倍。从这以上点可以推测，这类消费者的物质和时间均较为宽裕，更倾向于生活享受类的设施。第三，个人特征为中老年、女性、家庭成员小于等于 2 人并且希望有商场超市的群体占比 1%；在上述个人特征下希望有商场超市的消费者比例是总体相应比例的 2.702 倍。这类消费者多负担家务，所以更倾向于实用类的公共设施。

2) 个人因素与小区物业管理选择之间的关系

通过分析图 6 可知，在对小区物业管理选择中家政服务、水电通讯服务、保安服务为频繁项集。

进行 Apriori 关联规则挖掘，以个人因素为前项，以小区配套设施为后项，按照提升度降序排序，选择提升度最高项得到表 7。

由表可知：第一，个人特征为青年、男性、个人收入较低、家庭收入较高并且希望有家政服务的群体占比 2.5%；在上述个人特征下希望有游泳池的消费者占比总体相应比例的 2.739 倍。这类人群工作较

Table 5. The association rules table of personal factors and the purchase area choice

表 5. 个人因素与购房面积选择的关联规则表

前项	方向	后项	支持度	置信度	提升度
{B1 = 0, B4 = 0, B5 = 1}	≥	{A15 = 2}	0.010	0.333	1.961
{B1 = 1, B2 = 0, B3 = 1, B4 = 1}	≥	{A15 = 3}	0.020	0.667	3.508
{B1 = 1, B3 = 0, B6 = 0}	≥	{A15 = 4}	0.010	1	2.127

注：对于置信度为 1 的情况是由于样本量较小。

Table 6. The association rules table of personal factors and community facilities

表 6. 个人因素与小区配套设施的关联规则表

前项	方向	后项	支持度	置信度	提升度
{B1 = 1, B4 = 0, B6 = 1}	≥	{A10 = 1}	0.010	1	6.25
{B1 = 1, B2 = 0, B4 = 0}	≥	{A10 = 2}	0.010	1	3.508
{B1 = 1, B2 = 0, B3 = 0}	≥	{A10 = 3}	0.010	1	2.702

注：对于置信度为 1 的情况是由于样本量较小。

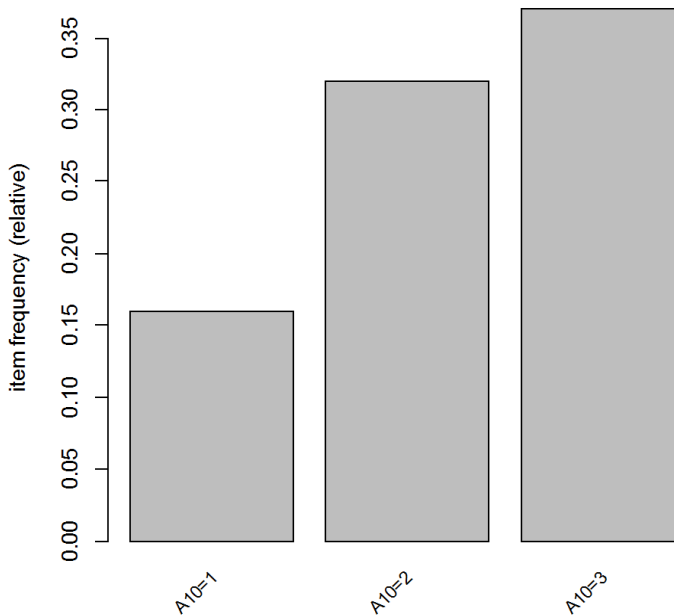


Figure 5. Frequent items of residential facilities

图 5. 小区配套设施频繁项

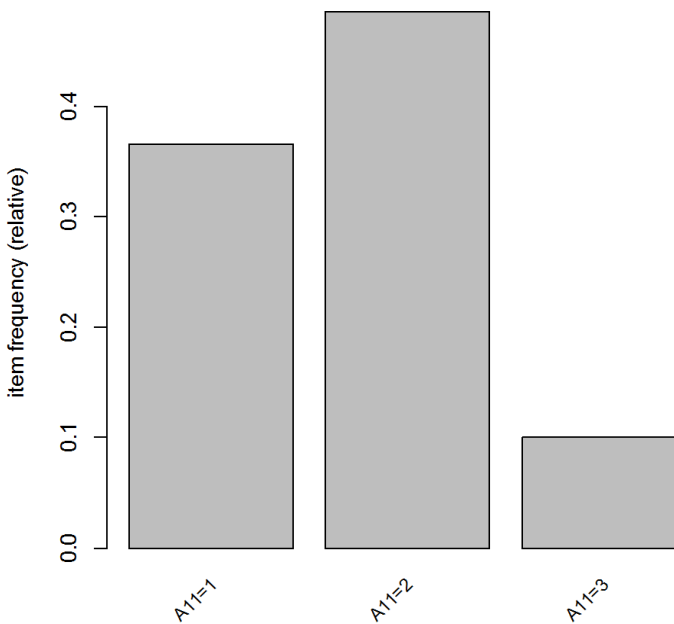


Figure 6. Frequent items of residential property management

图 6. 小区物业管理频繁项

Table 7. The association rules table of personal factors and community facilities

表 7. 个人因素与小区配套设施的关联规则表

前项	方向	后项	支持度	置信度	提升度
{B1 = 0, B2 = 1, B5 = 0, B6 = 1}	≥	{A11 = 1}	0.025	1	2.739
{B1 = 1, B2 = 1, B3 = 0}	≥	{A11 = 2}	0.015	1	2.061
{B3 = 0, B4 = 0}	≥	{A11 = 3}	0.010	1	10

注：对于置信度为 1 的情况是由于样本量较小。

忙，并且经济有支撑，所以更加倾向于家政服务。第二，个人特征为中老年、男性、无工作并且希望有水电通讯服务的群体占比 1.5%；在上述个人特征下希望有游泳池的消费者占比是总体相应比例的 2.061 倍。第三，个人特征为无工作、家庭成员小于等于 2 人并且希望有保安服务的群体占比 1%；在上述个人特征下希望有保安服务的消费者占比是总体相应比例 10 倍。可推测这类人群社会经验较少，对安全问题较为重视。

4.4. 个人因素与小区周边环境特征之间的关联挖掘

通过分析图 7 可知，在对小区周边环境的选择中银行、医院、学校为频繁项集。前两项均为生活必需设施，接下来对学校一项进行全面分析。

进行 Apriori 关联规则挖掘，以个人因素为前项，以小区配套设施为后项，按照提升度降序排序，选择提升度最高项得到表 8。

由表可知，个人特征为青年、有工作、家庭成员小于等于 2 人并且倾向于选择学校的消费者比例为 1%；在上述个人特征下希望有学校的消费者占比是总体相应比例 18.18 倍。从此可以看出，青年中暂时没有子女的消费者群体在选房时尤其注重学校因素。

5. 结论和建议

1) 首先，从整体看来，从住房自身特征来看，现房、房价小于等于 10,000 元/m²、房屋面积在 51~120 m² 的房屋更受消费者欢迎；从住房所在小区特征来看，小区配套公众设施有游泳池、健身设施和商场超

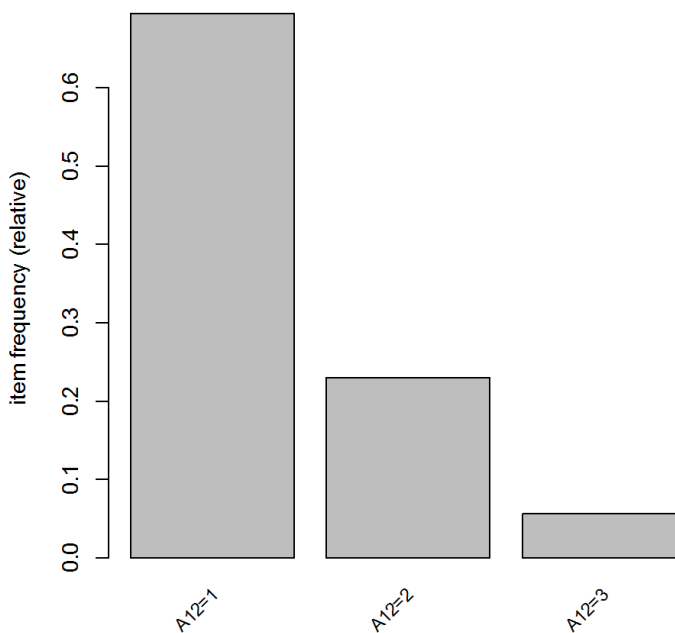


Figure 7. Frequent items of area surrounding environment

图 7. 小区周边环境频繁项

Table 8. The association rules table of the personal factors and the characteristics of the surrounding environment of community
表 8. 个人因素与小区周边环境特征的关联规则表

前项	方向	后项	支持度	置信度	提升度
{B1 = 0, B3 = 1, B4 = 0, B6 = 1}	≥	{A12 = 3}	0.010	1	18.18

注：对于置信度为 1 的情况是由于样本量较小。

市、小区物业管理包括家政服务、水电通讯服务和保安服务的小区更被关注，这些体现出现代消费者对健康和便捷的需求，开发商可以依此为卖点来吸引消费者；从小区周边环境特征来看，除传统的医院和银行外，消费者对学校的需求也十分高，那么学区房的可以保证。

2) 其次，从开发商的角度来说，顾客群的划分也十分重要，从分析可知中老年人群相较于青年选择大面积、高房价的房子的趋势更明显。年龄、收入和家庭结构会在一定程度上影响消费者选择，且购房者均多考虑未来生活安排，所以需要挖掘其可能的潜在需求。若开发商计划针对某一群体进行楼盘售卖，那么便可以推出各方面条件适合该群体的楼盘；若开发商计划开发多元化的楼盘，那么便需要各种类型取交集或互补开发以吸引客户。如针对青年、低收入、家庭成员小于等于两人(即单身、同居或结婚暂时无子女)的群体可以开发小户型、安保设施齐全、有家政服务并且距离学校较近的楼盘。从销售商的角度来说，通过了解有购房需求的消费者的个人和家庭特征，通过关联规则挖掘的结果将其购房意向具体化，有针对性的进行推荐，有助于提高房屋售卖率。如消费者特征为中老年、家庭成员小于等于 2 人，那么便可以为推荐 91~120 m²、三室两厅两卫的房屋。当然，这并不是绝对的规律，还要根据购房者具体意愿进行分析。

3) 最后，本研究发现，仅通过某个单独的个人特征来对消费者购房意向进行定位和推荐并不全面，多种因素组合才能还原出消费者的真实面貌，进而为开发商提供决策建议。通过关联规则挖掘可以通过划分出特征的细微不同，进而实现较为购房意向精准的预测和推荐，从而提高房屋出售率。需要说明的是，本次调查问卷样本量较小，可能出现个别指标与实际有一定差距，但并不影响总体特征的反映。

参考文献 (References)

- [1] 向长江, 谢荣锋, 韩小芸. 星汇园潜在购房者感知价值及购买意向调查[J]. 广州大学学报: 社会科学版, 2002, 1(7): 71-75.
- [2] 吴殿廷, 钟景胤, 江周祥, 等. 大都市与中小城市居民家庭购房影响因素对比分析——以北京市与瓦房店市为例[J]. 社会科学辑刊, 2007(2): 100-105.
- [3] 于丹, 董大海, 刘瑞明, 等. 消费者购房需求及个人因素的影响研究[J]. 管理评论, 2007, 19(9): 16-21.
- [4] 蒋绪军. 商品房客户购房需求预测应用研究[D]: [硕士学位论文]. 南宁: 广西大学, 2014.
- [5] 李莹, 林功成. “85 后”消费者的购房意向及其影响因素研究——以广东省为例[J]. 广告大观: 理论版, 2015(2): 12-17.
- [6] 蒋盛益. 数据挖掘原理与实践[M]. 北京: 电子工业出版社, 2011.

期刊投稿者将享受如下服务:

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击: <http://www.hanspub.org/Submission.aspx>