

# 自动售货机商务数据分析

## 摘 要

自动售货机为顾客提供便捷的线下服务，并为商家省下一笔不小的人工成本，其经营模式小巧、自助，并提供实惠、高品质商品，是一种备受青睐的零售经营新模式，并且这种模式越来越趋于主流。本文选取五台售货机在 2017 年一年中的销售情况作为研究对象，对数据进行分析，并给出用户画像，同时预测了 2018 年 1 月的销售情况，供售货机经营者参考，更好为客户服务。

文章首先对数据做了详实生动的分析，使用大量图标展示了各台售货机的销售数量和销售金额，并通过横向和纵向的比较找出较受欢迎的商品种类。

接着文章绘制了客户画像，使用的方法是文字挖掘方法，通过在原始数据中出现最多的地点及商品种类，找出最受客户喜爱的售货机地点及商品，勾勒出客户群体的总体形象。

最后文章预测了 2018 年 1 月各台售货机的销售金额，使用方法为灰度预测法，该方法能解决样本数据月度数据不是很多的问题，较好的预测了销售金额。

**摘要：** 用户画像 灰度预测 图表分析

# 目 录

一、背景介绍.....	3
1.1、行业背景 .....	3
1.2、分析目标 .....	3
二、数据说明.....	3
2.1 数据来源 .....	3
2.2 数据内容 .....	3
三、销售数据分析.....	4
3.1 各售货机月度销售情况 .....	4
3.2 商品销量比较 .....	6
3.3 各台售货机月交易额比较图 .....	6
3.4 交易额与商品类别气泡图 .....	8
四、自动售货机画像.....	8
4.1 自动售货机标签 .....	9
4.1 绘制自动售货机画像 .....	9
五、业绩预测.....	10
5.1 售货机各大类商品销售数量预测 .....	10
5.2 预测模型建立 .....	11
六、小结.....	13

## 一、背景介绍

### 1.1、行业背景

自动售货机为顾客提供便捷的线下服务，并为商家省下一笔不小的人工成本，其经营模式小巧、自助，并提供实惠、高品质商品，是一种备受青睐的零售经营新模式，并且这种模式越来越趋于主流。用户画像常作为勾画目标用户、产品定位的有效工具，在当今商业领域的作用越来越大。

自动售货机运营商在投放机器时需要重点关注商品的供给频率、种类选择、供给量、站点选择等问题。借助科学的商业数据分析，经营者能更好了解客户需求，是掌握经营方向的重要手段，对自动售货机这一营销模式的发展意义重大。

### 1.2、分析目标

本次选取某商场处于不同地点的五台自动售货机，对其 2017 年一年的商品销售金额及销售数量数据进行分析，解决如下问题：

1. 对销售数据和市场需求进行分析，分析月度数据；
2. 预测每台售货机 2018 年 1 月商品销量；
3. 为售货机绘制用户画像，便于货品选择决策；
4. 为售货机所销售商品贴上标签，使其能够更好展现商品特征。

## 二、数据说明

### 2.1 数据来源

数据来源：五台售货机销售记录

数据起止时间：2017 年 1 月 1 日-2017 年 12 月 31 日

### 2.2 数据内容

销售记录：70680 条记录

产品类别：316 条记录

三、销售数据分析

对售货机销售数据进行预处理与分析，可以初步分析各台售货机月度销售情况。首先利用 Excel 的数据筛选功能，将各个售货机的销售数据分别提取并分类整理，便于之后研究。同时在数据筛选阶段，发现有一条记录显示为 2017 年 2 月 29 日，存在明显错误，予以剔除。接下来对各个售货机月度销售数据进行分析。

3.1 各售货机月度销售情况

Excel 的数据透视表功能可以快速分析各台售货机销售商品总数与平均金额，下面的图一展示了售货机 A 部分天数的每单平均应付金额、每单平均实付金额、每日销售总量，图二至图六分别展示了 A、B、C、D、E 五台售货机的月均销售情况。

1月3日	2.83	2.83	10
1月4日	6.75	6.75	4
1月5日	3.666666667	3.666666667	3
1月6日	5.833333333	5.833333333	3
1月7日	3.5	3.5	1
1月8日	4.83030303	4.83030303	33
1月9日	4.8	4.8	75
1月10日	4.395061728	4.395061728	81
1月11日	3.547368421	3.547368421	19
1月15日	4	4	2
1月16日	6.185714286	6.185714286	7
1月17日	2	2	1
1月20日	1.5	1.5	1
1月21日	3.75	3.75	2
1月22日	4.25	4.25	2
1月23日	3	3	1
1月25日	3.5	3.5	2
1月26日	3.875	3.875	4

图 1. A 售货机部分日期销售情况

A售货机销售情况		
月份	每单平均交易额	月订单量
1月	4.507	335
2月	3.864	114
3月	3.585	255
4月	4.037	447
5月	4.478	756
6月	4.047	1669
7月	4.098	476
8月	3.359	666
9月	4.307	1040
10月	4.021	1565
11月	4.472	1160
12月	3.788	2003
总计	4.057	10486

图 2. A 售货机月均销售情况

B售货机销售情况		
月份	每单平均交易额	月订单量
1月	3.753	366
2月	3.256	185
3月	3.615	265
4月	4.075	603
5月	4.236	869
6月	4.068	1856
7月	4.402	345
8月	3.584	981
9月	4.130	1745
10月	4.112	2026
11月	4.269	2031
12月	3.667	2210
总计	4.003	13482

图 3. B 售货机月均销售情况

C售货机销售情况		
月份	每单平均交易额	月订单量
1月	4.328	379
2月	3.826	207
3月	3.770	263
4月	4.404	734
5月	4.727	789
6月	4.502	1882
7月	3.988	764
8月	3.914	1259
9月	4.427	1678
10月	4.273	2216
11月	4.352	1943
12月	3.943	2379
总计	4.248	14493

图 4. C 售货机月均销售情况

D售货机销售情况		
月份	每单平均交易额	月订单量
1月	3.693	259
2月	3.089	141
3月	4.306	192
4月	3.790	443
5月	4.241	564
6月	4.026	1040
7月	4.230	317
8月	3.317	715
9月	3.899	983
10月	3.884	1186
11月	3.862	1210
12月	3.573	1663
总计	3.815	8713

图 5. D 售货机月均销售情况

E售货机销售情况		
月份	每单平均交易额	月订单量
1月	4.680	354
2月	3.638	258
3月	4.306	350
4月	4.160	895
5月	4.411	1292
6月	3.818	2593
7月	3.919	813
8月	3.804	1767
9月	4.125	4134
10月	3.676	2777
11月	4.283	5020
12月	4.169	3252
总计	4.070	23505

图 6. E 售货机月均销售情况

如上图所示，根据提取出来的销售记录可以做出如下分析：

首先是每一天的销售平均金额和销售商品总量均不一样且相差较大，同时在同一天各个时段销售概况也不尽相同差别较大，在一年中的不同月份销售情况也有差异。

其次是各台售货机的销售对比。从图中可以看到，E 售货机的总订单量是最大的，在部分月份比其他售货机多卖出很多商品。同时从图表可以看出五台售货机销量较多的月份为六月、九月到十二月这五个月份。

接下来我们统计了每台售货机每单平均交易额与日均订单量，如下表所示：

售货机名称	每单平均交易额	日均订单量
A	4.057	28.7
B	4.003	36.9
C	4.248	39.7
D	3.815	23.9
E	4.070	64.4

表 1. 各售货机平均销售情况

从表格可以看出，虽然 E 售货机日均订单量显著多于其他售货机，但是他的每单平均交易额低于 C 地点，初步说明人们在 C 地点可能更偏好于购买更高价的商品。

### 3.2 商品销量比较

接下来我们继续比较各台售货机的商品销量,通过 Excel 的数据透视表功能,以 2017 年 6 月为例,我们提取出销量前五的商品,分别为怡宝纯净水、40g 双汇玉米热狗肠、东鹏特饮、脉动、250ml 维他柠檬茶,这五个商品在 6 月的具体销量呈现为如下柱状图。

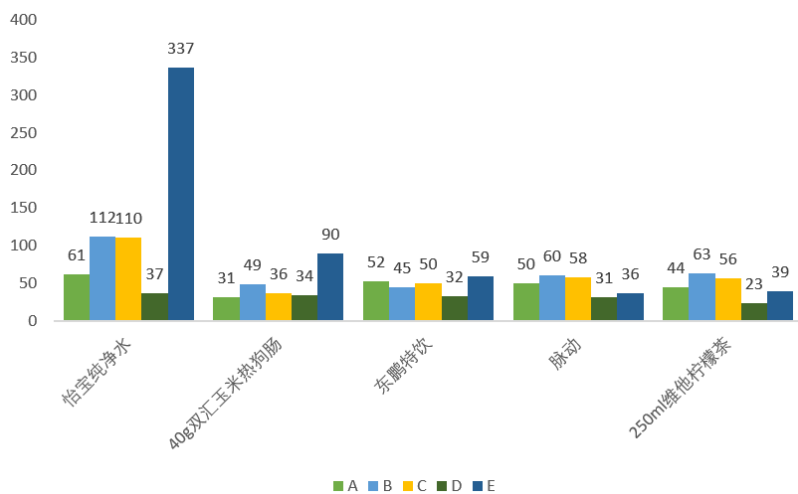


图 7. 2017 年 6 月销量前五的商品销量柱状图

从图中可以看出,这五个销量最好的商品有四个属于饮料大类,从全年看来销售最好的商品也极大多数属于饮料类,同时在饮料类中,怡宝矿泉水的销量与其他商品相比又有很大优势,其中 E 地点的怡宝矿泉水销量超过其他地点。怡宝纯净水售价不高,味道符合大众口味,且是一个常见的品牌,因此在自动售货机中会更受消费者青睐,上榜的其他饮料也是耳熟能详的热销饮料,因此会有更多人选择在自动售货机上购买他们。

### 3.3 各台售货机月交易额比较图

下图展示了每台售货机每月总交易额的折线图,反映了交易额随着月份的变动趋势。

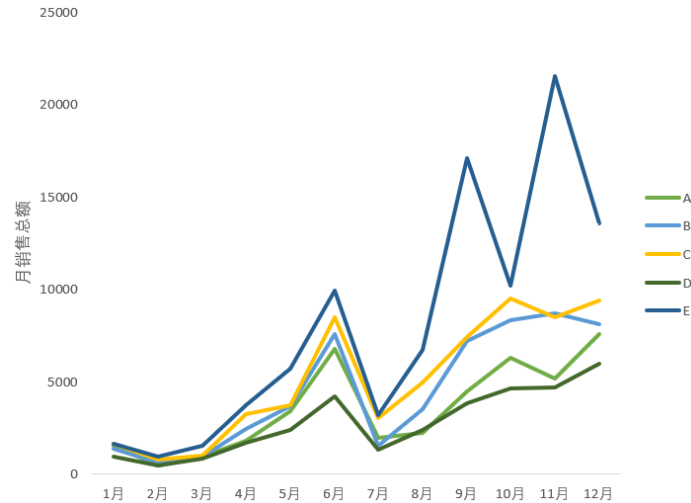


图 8. 每台售货机每月总交易额

从上图可以看出，五台售货机的总交易额变动趋势几乎相同，我们在接下来的研究中可以认为这五台机器有着相同的交易额上升或下降趋势。同时我们发现，这五台售货机在 2017 年六月、九月至十二月这五个月中销售额有明显的上涨情况，且 E 售货机在各个月份销售额都大于其他售货机，C 售货机、B 售货机、A 售货机、D 售货机分别为销售额的第二、三、四、五名。

接下来的图是交易额月环比增长率的柱状图，通过此图，我们可以清晰的看出这五台售货机在一年中地点交易额变动情况。



图 9. 交易额月环比增长率

如图所示，这五台销售机的环比增长态势几乎相同但存在细微差别，交易额环比增长率普遍下降的有一月和七月两个月份。

### 3.4 交易额与商品类别气泡图

利用 Excel 将附件一的销售记录与附件二的商品类别合并在一个表格，可以画出交易额与商品类别的气泡图，商品类别展示为商品的二级类目，如下图所示：

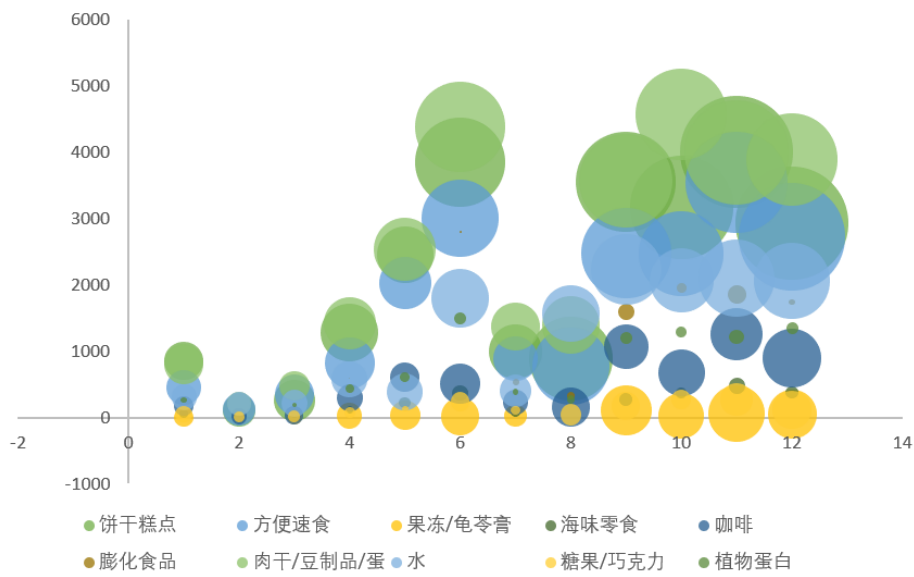


图 10. 交易额与商品类别气泡图

从图中可以看出，交易额较多的商品为饼干糕点类、肉干/肉制品/蛋类、海味零食类、植物蛋白饮料类；其次是方便素食类、水类。这几种食品都较为常见且商品体型不是很大，比较能够满足顾客的需求。

## 四、自动售货机画像

用户画像是一种勾画目标用户、联系用户诉求与设计方向的有效工具。用户画像在大数据时代背景下在电商领域得到广泛应用，他能够利用标签将用户的形象具体化，总结用户群体总体的偏好，便与为客户提供针对性服务。



## 4.1 自动售货机标签

接下来，我们为每台售货机销售商品归类，贴上热销/正常/滞销的标签。评判依据如下：热销商品定义为连续销售满 12 个月且总销量排名在前 5%的商品；滞销商品定义为总销量排名在后 5%的商品，其余商品为正常商品。

此外对售货机更进一步贴上标签，将销售价格将消费档次分为高档/中档/低档。低档定义为单价为所有商品单价的最小 25%，高档定义为单价为所有商品单价的最大 25%，其余商品为中档商品。

如上两步均可使用 Excel 完成，故不另附程序，结果见附件 task3-1 文件夹及 task3-2 文件夹。

## 4.1 绘制自动售货机画像

由于在这个自动售货机的案例中，所涉及的数据量较大，使用用户画像的方法能使产品的服务对象更加聚焦、更加的专注，售货机的服务会更加令客户满意。

接下来的流程图展示了我们绘制用户画像的过程：

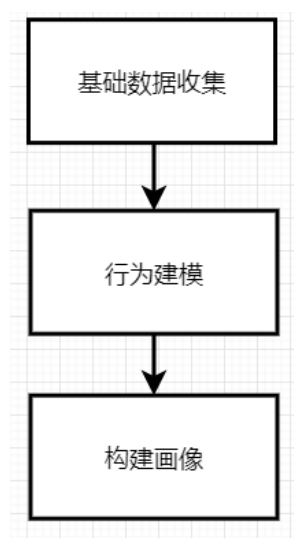


图 11. 用户画像流程图

如上图所示，第一步基础数据收集我们在之前的研究中已经完成，接下来的第二步行为建模是基于用户的基础数据，通过技术手段进行行为建模，我们采用文本挖掘的方法进行。文本挖掘是利用原始销售信息文件，找出销量最多

的商品及销售地点。于是我们得到了如下图所示的用户画像。



图 12. 用户画像

上图所示的用户画像勾勒出售货机顾客的基本形象。首先，用户更加偏好去 E 地点购物，其次用户更加偏好饮料类的商品。

## 五、业绩预测

接下来我们使用灰色预测方法预测 2018 年 1 月的各台售货机销售金额，销售金额与销售的货物有关，由上面的分析我们看出销量与月份可能存在一定的关系，但由于我们每个月的数据只有一个，因此在接下来的预测中暂时不考虑时间的影响，且考虑的因素为饮料类和非饮料类商品的销售数量。

### 5.1 售货机各大类商品销售数量预测

在进行灰度预测前，首先要解决的问题是能否预测未来的饮料类和非饮料类商品的销售数量。由于我们之前假定商品销量不受季节影响，因此这两个大类商品的销售数量是可以通过回归方法预测的。而如果一定要考虑季节因素对销量的

影响，尽管这样更加合情合理，但会由于每个月份销售数据样本过小，导致商品的销量无从预测。因此认为商品销量不受季节影响在本节是一个重要的前提假设。

利用 2017 年的饮料和非饮料类这两个大类每月的销量数据，我们听过 MATLAB 线性回归功能可以预测出 2018 年 1 月的销量数据，由于这是使用回归工具箱直接得到的数据，因此不另附代码，销量数据如下图所示。

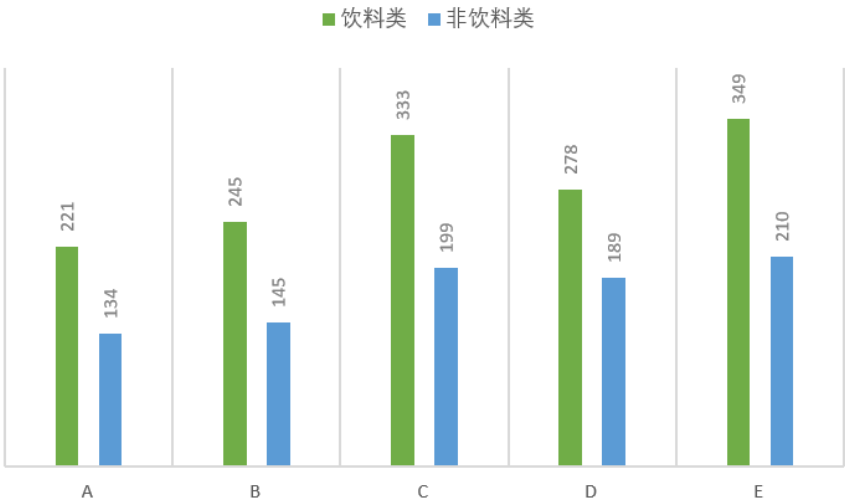


图 13. 2018 年 1 月商品销量预测

### 5.2 预测模型建立

我们建立饮料类和非饮料类商品销量的定权组合预测模型，并通过灰色关联分析的方法确定回归系数，定量描述两种商品销售数量对最后销售金额的影响。

模型建立的主要步骤如下 [3]：

(1). 选定母因素和子因素，并记录数据。

母因素是评价的主要因素，我们选取售货机的销售今晚。子因素是其他相关因素，在这里指该售货机的饮料类商品销量和非饮料类商品销量。

(2). 由于指标序列间的数值不存在运算关系，对数据进行无量纲化处理。

$$X_0(k) = \frac{S(k)}{S(1)}$$

$$X_i(k) = \frac{Y_i(k)}{Y_i(1)}, i = 1, 2, k = 1, 2, \dots, 12$$

其中，序列  $S_0(k)$  表示母因素，序列  $X_i(k)$  表示子因素  $i$ ， $Y_i(1)$  代表观测值的

初始值，即 2017 年 1 月的数据， $Y_i(k)$  为观测值在月份  $k$  的取值。采用这样“初值化”的方法处理数据得到的新序列中，各数无量纲且数值大于 0，各比较序列拥有同一个起点，有利于研究各因素之间的变化关系。

(3). 灰色关联系数求值。

$$\xi_i(k) = \frac{\Delta_{min} + \rho \Delta_{max}}{\Delta_i(k) + \rho \Delta_{max}}, i = 1, 2$$

其中， $\xi_i(k)$  为第  $i$  个因素与母因素在时刻  $k$  的关联系数； $\Delta_{0i}(k)$  指  $k$  时刻两个序列差的绝对值，即  $\Delta_i(k) = |X_0(k) - X_i(k)|$ ； $\Delta_{min}$  与  $\Delta_{max}$  指  $\Delta_{0i}(k)$  中的最大值与最小值； $\rho$  指分辨系数，它的取值决定灰色关联系数的分布状况，我们根据分辨系数的经验值取  $\rho = 0.5$ ，此时关联度的变化幅度适中，对评价过程的稳定控制能力较强。

(4). 求解灰色关联度。

灰色关联度是灰色关联系数的平均值，即

$$r_i = \frac{1}{T} \sum_{k=1}^T \xi_i(k), i = 1, 2$$

由此得到灰色关联度矩阵  $[r_1, r_2]$ ，从中可以看出各个因素对母因素的影响大小。

灰色关联度通过比较各因素的关联度大小，从总体上综合地体现各因素对系统行为影响的差异，因此被分析对象间的差异体现得越明显越好， $r_i$  需要满足以下原则：

$$\Delta = \max\{r_i\} - \min\{r_i\}, \text{尽可能大}$$

$$\sigma = \frac{1}{n} \sqrt{\sum_{i=1}^n (r_i - \bar{r})^2}, \text{尽可能大}$$

(5). 计算各个因素在预测模型中的权重大小。

利用灰色关联度矩阵  $[r_1, r_2]$  计算因素  $i$  在线性预测模型的权重：

$$K_i = \frac{r_i}{\sum_{i=1}^2 r_i}, i = 1, 2$$

(6). 建立单因素一元非线性回归影响影响方程。

售货机销售金额与各个相关因素有一定的联系，通过分析数据，发现商品销

售金额与商品销量之间存在非线性关系，即在时刻  $t + 1$ ，销量变量  $i$  和总体月销售金额预测值  $\hat{S}$  的关系可以表示为：

$$\hat{S}_i(t + 1) = F_i[Y_i(t + 1)], i = 1, 2$$

(7). 利用因素权重构造定权组合预测模型。

$$\hat{S}(t + 1) = \sum_{i=1}^2 K_i \hat{S}_i(t + 1)$$

综合以上分析， 我们得到利用售货机各大类商品月销量预测售货机月销售金额的灰关联定权组合模型。利用上面的过程以及上一节预测的商品销量，我们得到了 2018 年 1 月各台售货机的预测销售金额：

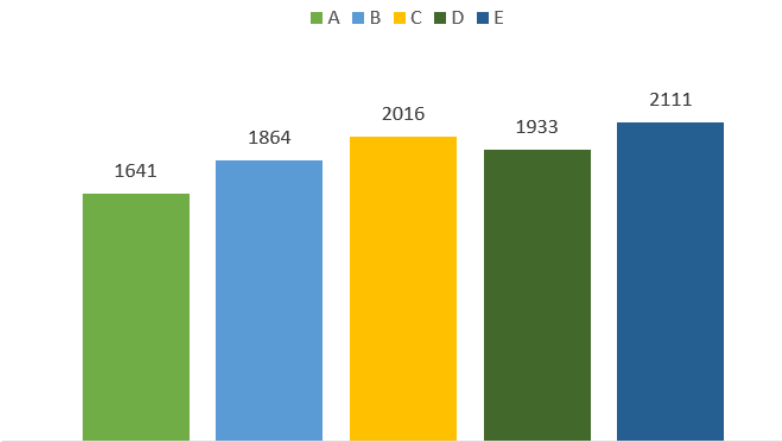


图 14. 2018 年 1 月售货机销售额预测

## 六、小结

本文运用大量图表展现数据分析结果，生动详实，但可以在数据分析方法上做更多的改进，研究发现，顾客对 E 地点的售货机更为偏好，且顾客对于饮料类商品比较偏好。