

# 分析报告

——新零售无人智能售货机商务数据分析

何志锐

2019-6-7

# 目录

一、前言 .....	2
1、背景介绍 .....	2
2、数据介绍及预处理 .....	2
3、分析目标 .....	2
二、销售情况分析 .....	3
1、每单平均交易额 .....	3
2、日均订单量 .....	3
3、总交易额及环比增长率 .....	4
4、毛利润占比 .....	7
5、各类型商品的单均交易额 .....	8
6、商品出售时间 .....	10
三、商品清单分析 .....	12
1、各商品的销售量情况 .....	12
2、不同地点的消费者商品偏好 .....	12
四、业务预测 .....	14
1、原理简述 .....	14
2、预测过程 .....	14
五、总结 .....	17

# 一、前言

## 1、背景介绍

随着中国人口红利的渐渐消退和人口老龄化日益加剧，零售行业的人力成本正在变得越来越高。从行业竞争角度来看，传统零售行业急需寻找新的变革，通过融入先进科技加速行业转型，以得到更快更好的发展。由此，“新零售”应运而生。

新零售是指以互联网为依托，通过运用大数据、人工智能等先进技术手段对传统的商品生产、流通与销售过程进行升级改造的零售新模式<sup>1</sup>。无人售货机作为典型的线下零售方式，可以通过接入物联网的方式采集大量的销售数据，并运用大数据对销售过程进行分析改造使其智能化，是新零售模式的典型代表。

## 2、数据介绍及预处理

本次分析的数据来自于某地区中的无人智能售货机于 2017 年的销售记录（附件 1），该地区共投放了 5 台售货机于 A、B、C、D、E 处，共 70680 份订单，除 1 份订单的发生时间超出日历范围外其余订单记录均为有效数据。

销售记录中提供了订单号、设备 ID、应付金额、实际金额、商品（名称）、支付时间、地点、状态（是否出货及是否退款）和提现（是否提现）共 9 种数据属性。由于流水的订单号没有分析意义故作舍去处理。设备 ID 和地点呈一一映射关系，为了便于操作和分析，直接使用地点属性的标记对应设备 ID。而通过简单的统计可以发现所有订单记录中的“状态”和“提现”两种属性均只有一种状态，因此也没有分析意义故舍去。

最后将销售记录根据地点分成 5 份数据，并按照具体的分类标准（附件 2）给销售记录中的商品贴上大类（是否为饮料）及二级类（膨化食品、饼干糕点、功能饮料等）标签。

## 3、分析目标

---

<sup>1</sup> <https://baike.baidu.com/item/%E6%96%B0%E9%9B%B6%E5%94%AE/20143211>

本次分析旨在通过无人智能售货机的销售记录对其订单量、交易额及其增长率、毛利率、各类型商品的销售情况和商品出售时间及地点进行分析，得出该地区消费者的商品偏好和消费习惯，针对性地给出更优的商品供给策略以更好地满足消费者需求同时提高销售情况，并对未来一个月的销售情况进行预测。

## 二、销售情况分析

### 1、每单平均交易额

由表 1 可以发现各地区各月份的每单平均交易额的差别不大，均处于 3.09 元到 4.73 元之间，均值为 4.02 元，标准差只有 0.35。这表明处于不同地点的消费者商品价格偏好基本一致，而同一地点的消费者在一年的周期中对商品价格的偏好也不会发生明显的变化。

表 1 2017 年各地点售货机各月份每单平均交易额（元）

月份 地点	1	2	3	4	5	6	7	8	9	10	11	12
A	4.51	3.86	3.59	4.04	4.48	4.05	4.10	3.36	4.31	4.02	4.47	3.79
B	3.75	3.26	3.61	4.08	4.24	4.07	4.40	3.58	4.13	4.11	4.27	3.67
C	4.33	3.83	3.77	4.40	4.73	4.50	3.99	3.91	4.43	4.27	4.35	3.94
D	3.69	3.09	4.31	3.79	4.24	4.03	4.23	3.32	3.90	3.88	3.86	3.57
E	4.68	3.64	4.31	4.16	4.41	3.82	3.92	3.80	4.13	3.68	4.28	4.17

### 2、日均订单量

由表 2 可以发现各地区各月份的日均订单量差异显著，计算得标准差为 33.02，其中峰值可达一个月中平均每天销售 167.33 单，而最低则在一个月中平均每天只销售出 4.07 单。

在一些月份中（如 1 至 4 月），不同地点售货机之间的订单量差异并不大，而当订单量差异较大时（如 9，11，12 月），均呈现出位于地点 E 的售货机销量明显高于

其他地点的售货机销量。则可以发现处于同一地点的消费者，在一年的周期中对售货机的需求会发生明显的变化，总体而言下半年的销售情况会比上半年的销售情况会更好。结合上一节的分析可以得出结论，**不同售货机之间的总销售额差异是由订单量决定的。**

表 2 2017 年各地点售货机各月份日均订单量

月份 地点	1	2	3	4	5	6	7	8	9	10	11	12
A	10.81	4.07	8.23	14.90	24.39	55.63	15.35	21.48	34.67	50.48	38.67	64.61
B	11.81	6.61	8.55	20.10	28.03	61.87	11.13	31.65	58.17	65.35	67.70	71.29
C	12.23	7.39	8.48	24.47	25.45	62.73	24.65	40.61	55.93	71.48	64.77	76.74
D	8.35	5.04	6.19	14.77	18.19	34.67	10.23	23.06	32.77	38.26	40.33	53.65
E	11.42	9.21	11.29	29.83	41.68	86.43	26.23	57.00	137.80	89.58	167.33	104.90

### 3、总交易额及环比增长率

通过对比图 1 至图 5 的折线可以发现，各地点的售货机总交易额变化趋势的差异较小，均呈现出年初到年中逐渐增长，年中骤降，骤降后到年尾再次回弹的变化趋势，并于 11 和 12 月份达到一年总的峰值，这旁证了上一节中得出的结论，即消费者在一年的周期中对售货机的需求会发生明显的变化。

进一步考虑各地点售货机总交易额的环比增长率，均表现为先负后正再负后正的规律，其中出现的两次负增长率均分别出现在 2 月和 7 月份，这是一个很明显的特征，考虑到 2 月和 7 月正处于学校的寒暑假，可以认为**该地区中售货机的主要消费群体可能是学生，所以售货机的商品选择应该更多地考虑学生群体的需求。**

同时值得注意的是，虽然各地点售货机的总销售额的变化趋势相差不大，环比增长率也大致处于-1%到 1.5%之间，不同地点的销售总额有着很明显的差异。举例来说，售货机 E 在 10 月份中的总销售额出现了下跌的情况，而其他地点的售货机的总销售额均在增长，但售货机 E 依然创造出该月最大总销售额的记录。

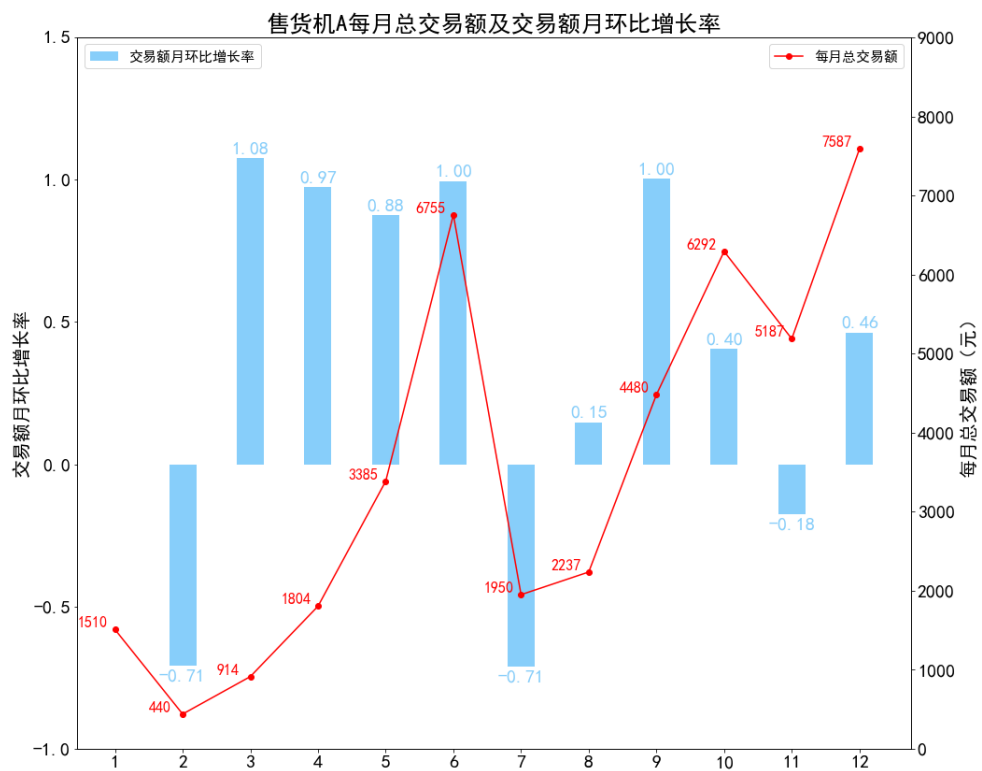


图 1 2017 年售货机 A 的各月份交易总额及环比增长率

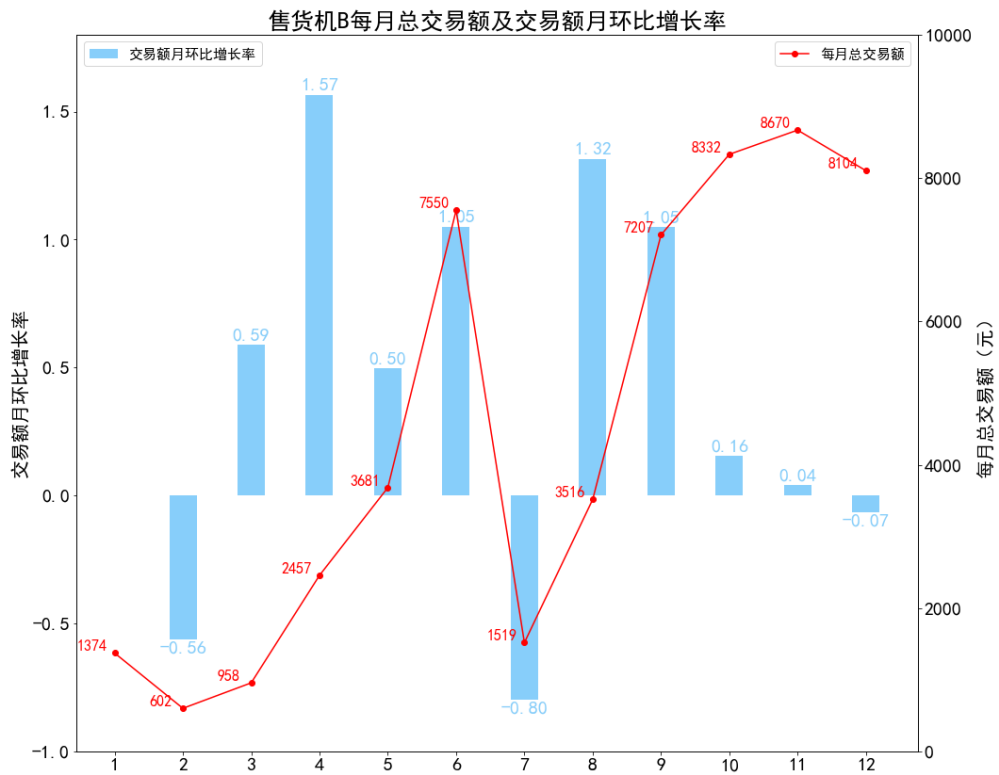


图 2 2017 年售货机 B 的各月份交易总额及环比增长率

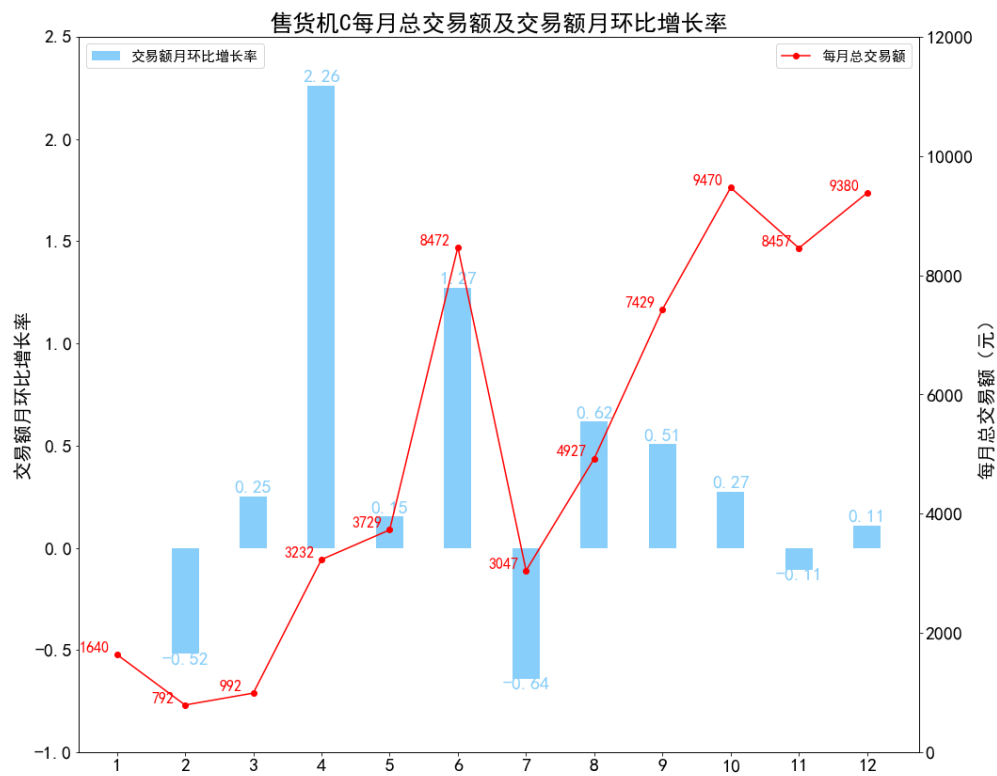


图 3 2017 年售货机 C 的各月份交易总额及环比增长率

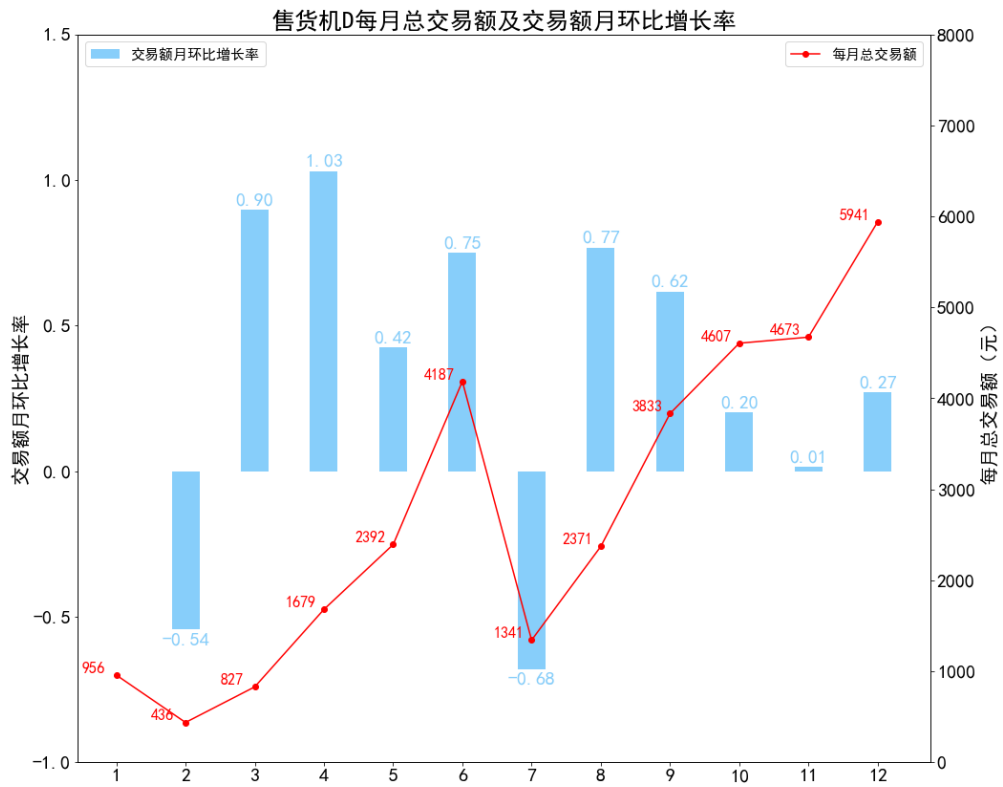


图 4 2017 年售货机 D 的各月份交易总额及环比增长率

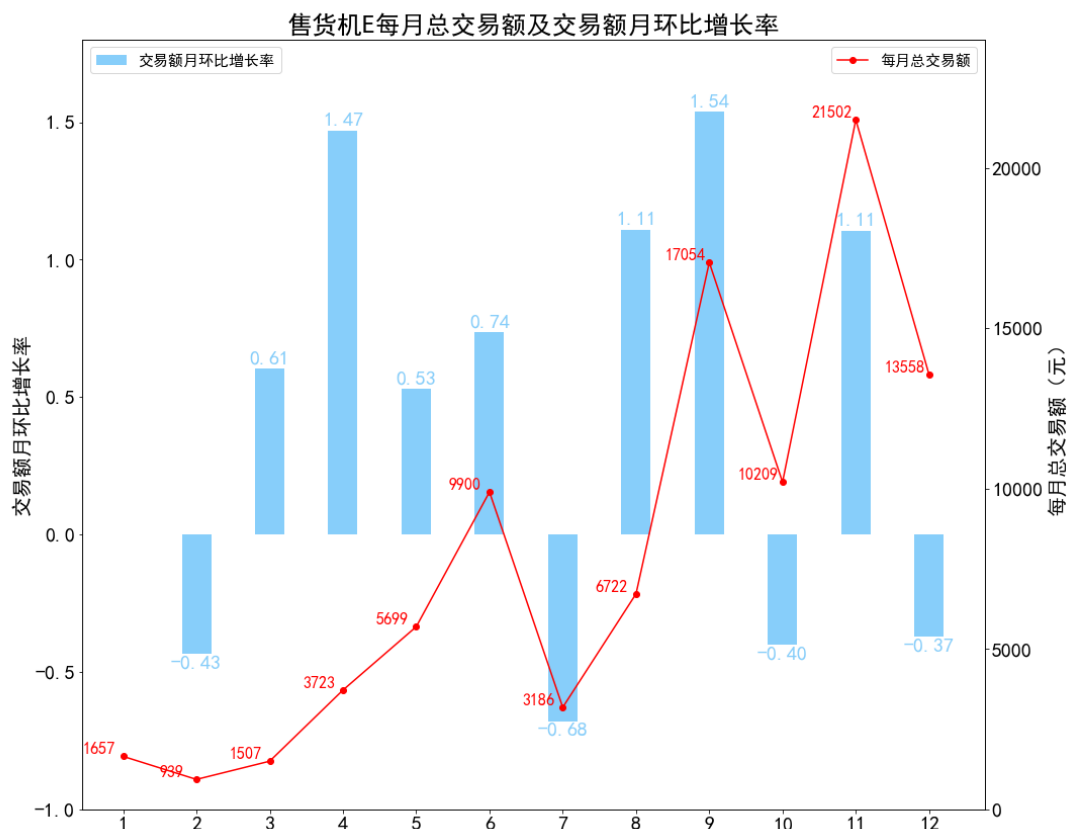


图 5 2017 年售货机 E 的各月份交易总额及环比增长率

## 4、毛利润占比

在假设饮料类商品的毛利率为 20%，非饮料类商品的毛利率为 20% 的前期下，对各地点售货机的销售记录进行毛利润统计可以得到如图 6 所示的结果。可得**各地点售货机对总毛利润的贡献大小为：E > C > B > A > D**，E 地点售货机的贡献最大而 D 地点售货机的贡献最小，这也和前面所作的分析结论相符。

为了提高处于 D 地点售货机的销售情况，可以针对 D 地区的消费者需求做一个需求调查，为 D 地区的消费者制定针对性的商品清单。若交易量低主要是由于该地区的人口密度相对较小或该地区的消费群体结构造成的，则可以考虑对 D 售货机进行重新选址，将其投放到对无人售货机需求更旺盛的地方。



各售货机毛利润占总毛利润的比例

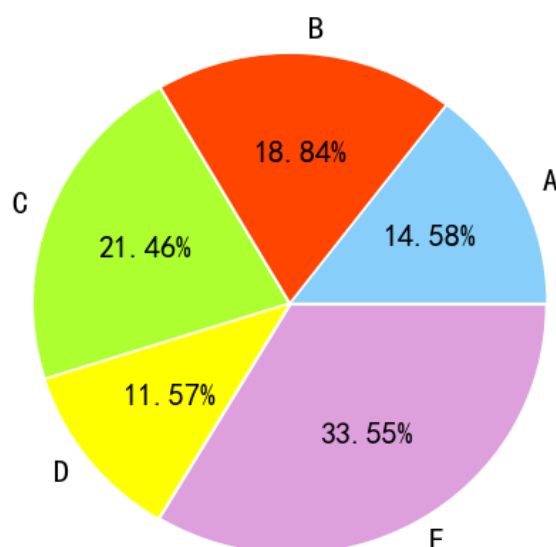


图 6 2017 年各地点售货机毛利润占总毛利润的比例

## 5、各类型商品的单均交易额

将商品按照二级类的分类标准进行分类，再统计出其各个月份的销售额和交易量即可得出如图 10 所示的气泡图。图中的气泡大小主要反映出是不同类型商品的单价大小，纵向比较可见香烟、咖啡、糖果/巧克力和蜜饯/果干类的商品的价格较高，因此售出单件商品带来的利润也会更高，但这不能表明更多地在售货机中销售这 4 类商品就能带来更高的利润。

从横向来看，**大部分类型的商品单均交易额变化不大**，这可能是由同类型商品不同品牌间的售价差异较小造成的，也可能是由于该地区的消费者对某类型商品有特定品牌和质量偏好造成的。因此对于具有这种特征的商品，若同类商品的品牌间售价差异较大，调查分析出消费者的品牌偏好就显得尤为重要了。当然，售货机中同类型商品的选择较小也会导致气泡变化差异较小。

值得注意的是，纸巾、坚果炒货、果冻/龟苓膏和其他类型的商品随着月份的变化有较为显著的单均交易额差异。其他类型的商品由于商品间差异较大，单均交易额出现这种特征是正常的，而其他 3 类商品均在 2 月和 8 月突然骤降，考虑到它们在别的月份中一直保持着差异很小的特征，则这种骤降可以考虑是由于缺货造成的。

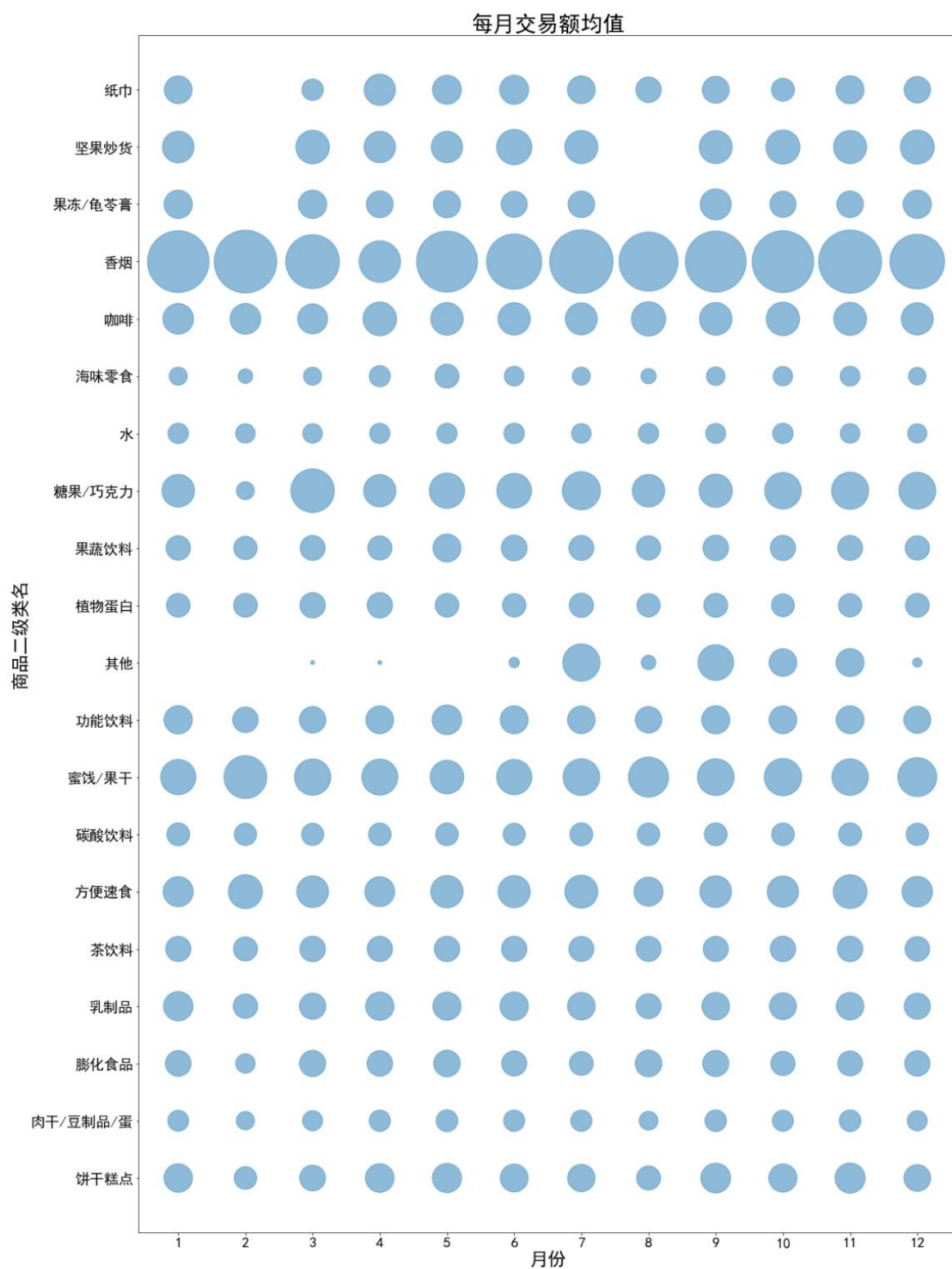


图 7 2017 年各二级类商品每月交易额均值

## 6、商品出售时间

以处于地点 C 的售货机为例，统计其在 6、7 和 8 月份的销售记录并以日期和小时为频率进行分割即可得到如图 8 至 10 的订单量热力图。首先可以直观地看出 6 月份的交易情况最优，7 月份突然暗淡下来，8 月份再次慢慢恢复，这符合在第 3 小节中得出的规律。

通过热力图也可以容易地得出该地区消费者在时间上的消费习惯，虽然在凌晨 0 点到早上 7 点间到售货机进行消费的人相对较少，但在交易量较大的月份中该段时间所带来的订单量依然是可观的，更何况在夜间依然正常营业是无人售货机的一大零售优势，因此售货机在深夜间依然存在继续营业的必要性。另外可以发现的是，在一天当中交易量最大的时刻发生在下午 4 点左右，在白天当中的其他时刻发生的订单数则较为均匀，而且峰值偏向于在后半个月出现，由此可以得出**售货机的最佳补货时间为深夜，而且在月初的时候要补足商品来应对月尾的需求高峰**。

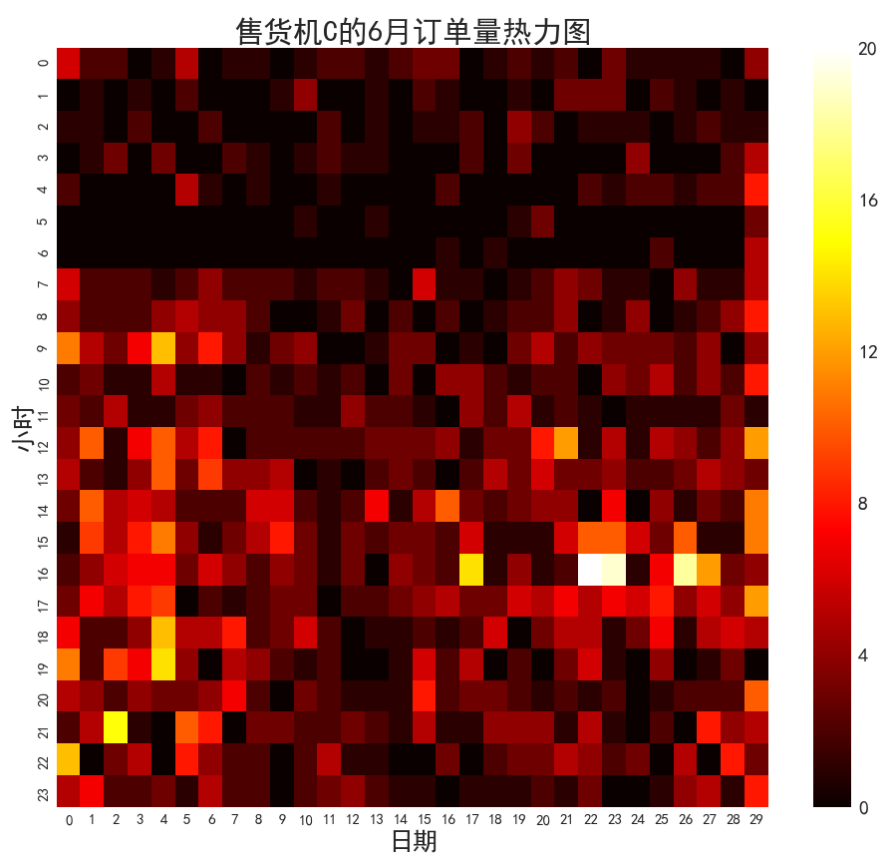


图 8 2017 年 6 月售货机 C 的订单热力图

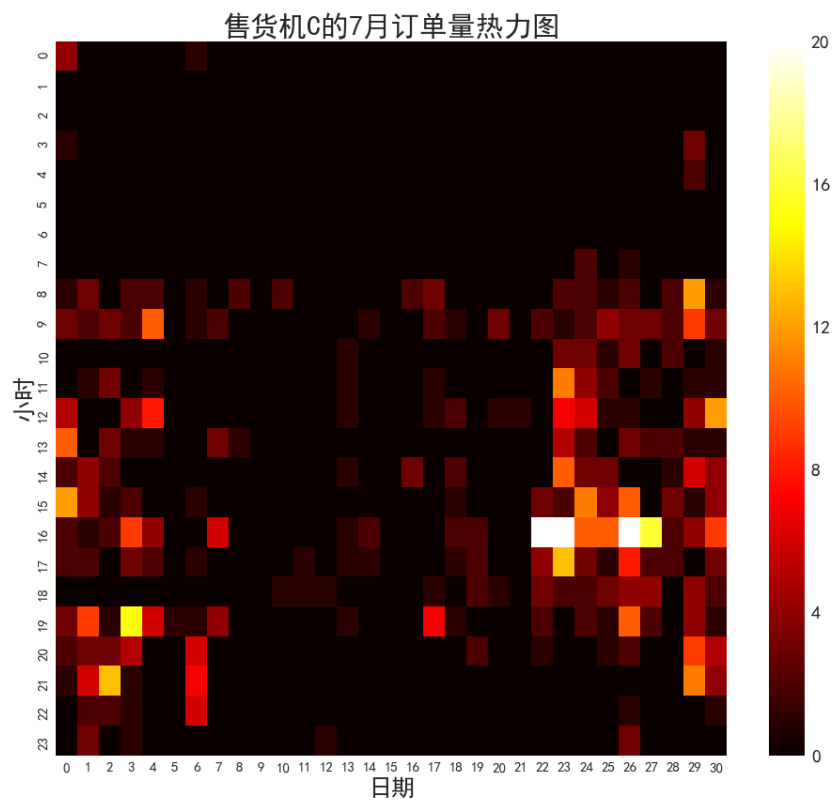


图 9 2017 年 7 月售货机 C 的订单热力图

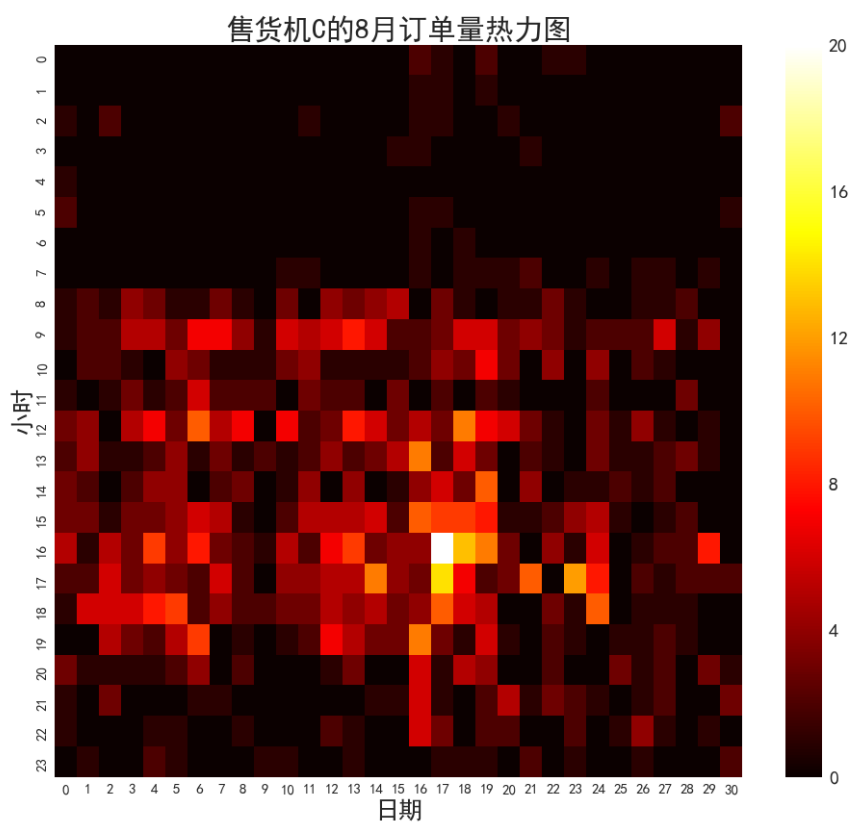


图 10 2017 年 8 月售货机 C 的订单热力图

### 三、商品清单分析

#### 1、各商品的销售量情况

由于售货机的存储空间有限，商品的数量对最终的销售利润影响很小，所以商品的选择和组合对提高交易量就显得尤为重要，于是哪些商品在消费者之间更为热捧？又是哪些商品白白占用了售货机中宝贵的空间就是首要需要关注的问题。通过对不同地点售货机的交易量进行统计，以四分位数作为划分的依据，交易量位于前 1/4 的为热销商品，位于后 1/4 的为滞销商品，其余为正常商品就可以得到具体的各个商品在不同地点的销售情况。表 3 以位于地点 E 的售货机销售情况为例，给出了其部分饮料类商品的标签，可以直接清晰地知道得知各商品的交易量情况。

表 3 2017 年售货机 E 的商品销售标签

商品名称	标签	商品名称	标签	商品名称	标签
怡宝纯净水	热销	500ml 统一阿萨姆奶茶	正常	百事可乐 330ml	滞销
脉动	热销	330ml 伊利畅意乳酸菌原味	正常	乐虎 (380ml)	滞销
营养快线	热销	600ml 可口可乐	正常	250ml 维他椰子植物蛋白饮料	滞销
阿萨姆奶茶	热销	冰糖雪梨	正常	维他柠檬茶 330ml (罐)	滞销
东鹏特饮	热销	600ml 雪碧	正常	冰红茶 (罐)	滞销
统一冰红茶	热销	可口可乐最新款	正常	经典奶茶	滞销
统一绿茶	热销	450ml 美汁源果粒橙	正常	小茗同学 (青柠红茶)	滞销
果粒橙	热销	250ml 维他奶巧克力味	正常	维他奶黑豆奶饮品	滞销
王老吉 (500ml)	热销	美年达 (罐)	正常	小茗同学冷泡茶 (乳酸菌味)	滞销
王老吉 (罐)	热销	维他奶	正常	珠江纯生啤酒	滞销

#### 2、不同地点的消费者商品偏好

通过上一节中统计的数据，可以进一步生成这些饮料类商品的词云如图 11 至 15，更直观地观察出商品的受欢迎程度。怡宝纯净水在各个地点中都是交易量最高的商品，其次是茶饮料（包括奶茶）、功能饮料和乳制品，该规律在各个地点中均无太大的变化，由此可以得出如下销售建议：

1) 适当提高茶饮料、功能饮料、乳制品尤其是纯净水在售货机中的储备量以满足消费者对这些商品相对较高的需求。2) 对于咖啡、果汁类饮料和碳酸饮料可以考虑保持当前的储备量以保证售货机能满足消费者多元的需求，这些类别的饮料在某些地点依然拥有可观的销量。3) 即使是属于热销二级类的商品也有部分特定的产品，应考虑剔除以节省售货机宝贵的存货空间。



图 11 售货机 A 的饮料销售情况词云图



图 12 售货机 B 的饮料销售情况词云图



图 13 售货机 C 的饮料销售情况词云图



图 14 售货机 D 的饮料销售情况词云图



图 15 售货机 E 的饮料销售情况词云图

#### 四、业务预测

## 1、原理简述

对于一般的预测问题可以使用传统的回归分析或新兴的深度学习等方法对目标变量进行函数拟合从而得出在下一个可能出现的因变量约束下的预测目标值，考虑到本次所获的销售数据量和数据维度均较小而且数据拥有明显的时间序列特性，本次销售额预测采用时间序列 ARMA 模型实现。

ARMA 模型由自回归模型 AR 和移动平均模型 MA 模型相结合而成，自回归模型描述当前值与历史值之间的关系，用变量自身的历史时间数据对自身进行预测，移动平均模型关注的是自回归模型中的误差项的累加，能有效地消除预测中的随机波动。模型对时间序列数据的平稳性有较高的要求。

## 2、预测过程



首先分别统计出饮料类商品和非饮料类商品在 2017 年各月份的销售额时间序列数据，在此只展现预测饮料类商品于 2018 年 1 月的销售额的具体过程。2017 年饮料类商品的月总销售额数据有明显的增长趋势，这也可以由第二章第 3 节中的图示观察得出，由于 ARMA 对对时间序列数据的平稳性有较高的要求，需要对时间序列数据进行差分处理。其一阶差分后的结果如图 16 所示，观察得以确定对时间序列数据进行一阶差分后序列平稳，进一步绘制其二阶差分的结果也能得到差分后序列平稳的结论，故模型取一阶差分格式。

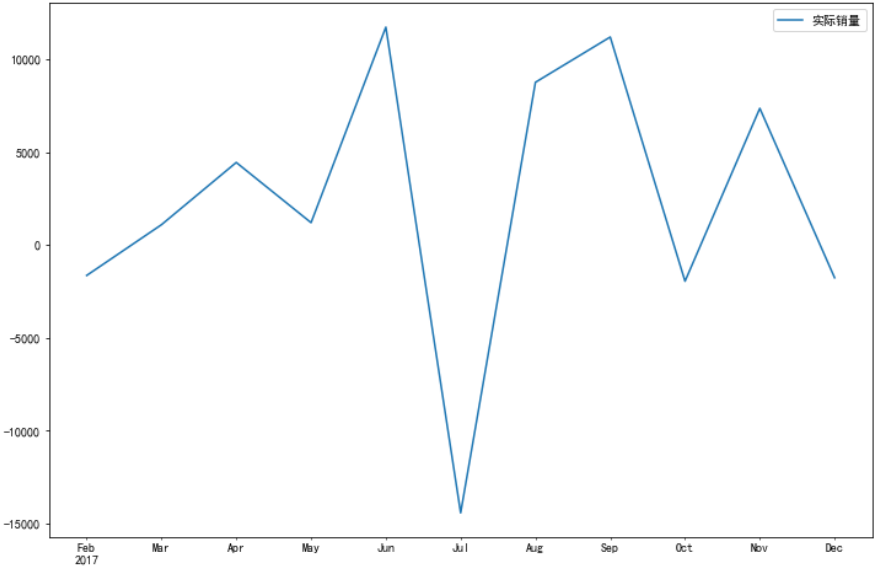


图 16 2017 年各月份销售额数据的一阶差分

为了确定模型确定 AR 模型的阶数  $p$  和 MR 模型的阶数  $q$ ，需要绘制时间序列的自相关图和偏自相关图来进行确定，可以看到图 17 中自相关系数显示出拖尾的性质而偏相关系数显示出 1 阶截尾的性质故可以初步确定模型为 ARMA(1,0)。为了更加准确地确认模型的阶数，进一步计算 ARMA(1,0)，ARMA(0,1)和 ARMA(1,1)的赤池信息量，贝叶斯信息量和汉南-奎因信息量如表 4 所示，可见 ARMA(1,0)是三者中最优的选择。

表 4 三种模型的 AIC、BIC 和 HQIC 值

	AIC	BIC	HQIC
ARMA(1,0)	254.04	255.49	253.50
ARMA(0,1)	257.31	258.77	256.78
ARMA(1,1)	337.02	338.96	336.31



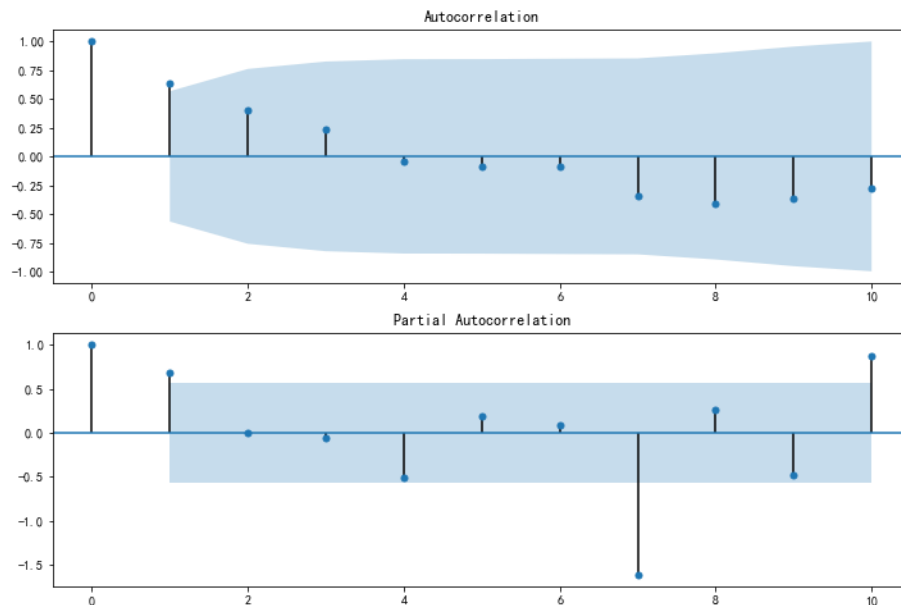


图 17 2017 年各月份销售额数据的自相关和偏相关图

建立起模型后对 2018 年 1 月和 2 月的饮料类商品销售额进行预测，得到 2018 年 1 月的预测销售额为 26750.59 元如图 18 所示。

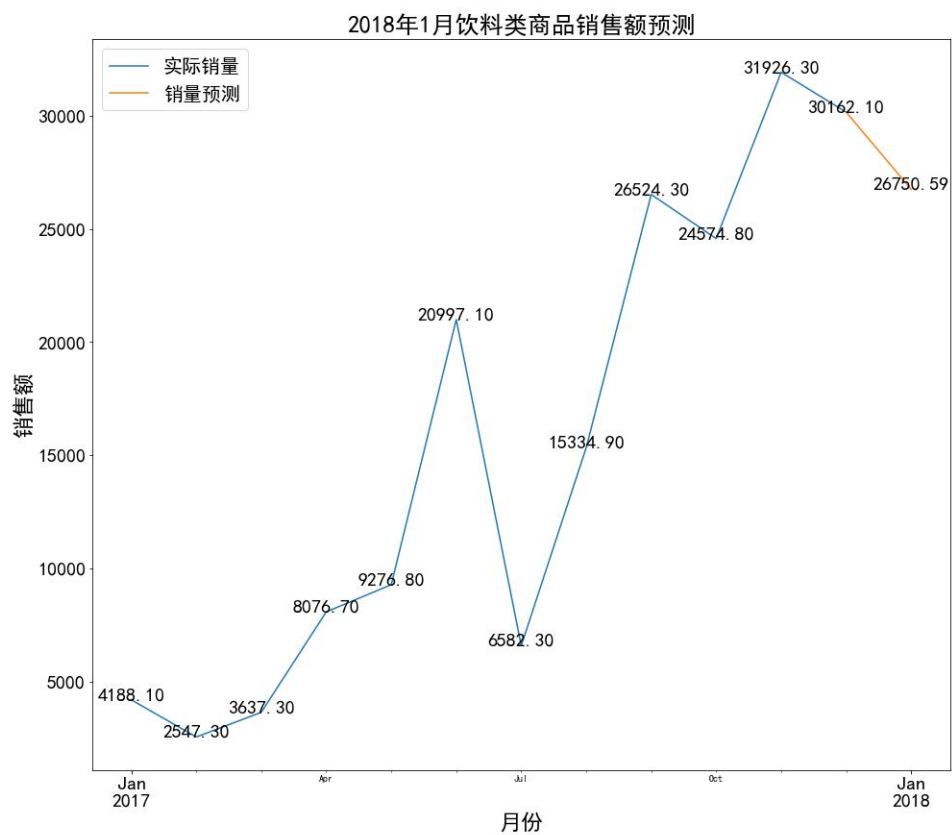


图 18 2018 年 1 月饮料类商品销售额预测

运用同样的方法再对非饮料类商品的销售额进行预测，其中确定使用的模型为 ARMA(0,1)，可以得到如图 19 所示的结果，其 2018 年 1 月的预测销售额为 9944.28 元。

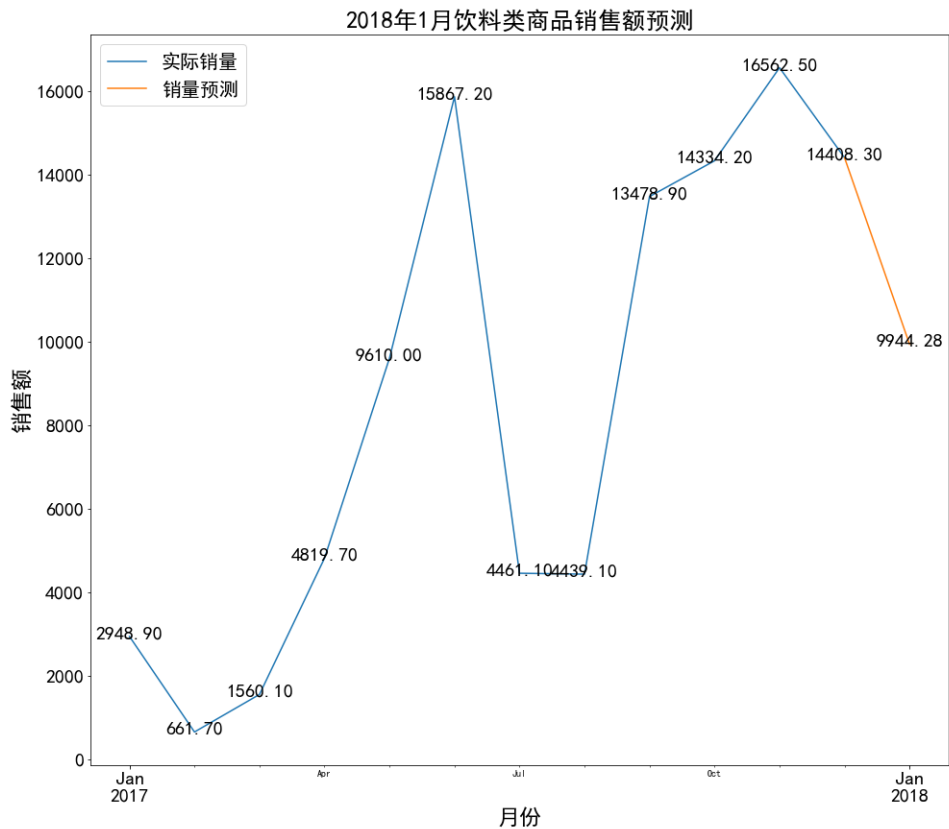


图 19 2018 年 1 月非饮料类商品销售额预测

## 五、总结

综合上述的分析该地区的无人智能售货机在下一年的销售策略可以总结为，首先应针对 D 地点消费者的需求进行调查并调整商品清单，适当采取如派送电子优惠券等的营销办法刺激消费让当地的消费者养成到自动售货机购物的情况，若情况仍然没有改善，就需要考虑对售货机进行重新选址或干脆撤掉该地点的售货机以降低业务成本。其次，通过分析得出该地区到售货机消费的人群主要为学生，应该针对该群体对商品清单进行调整，根据第三章中得出的分析结果，饮料类商品的具体调整策略为增加售货机中矿泉水、茶类饮料、功能饮料和乳制品的种类和数量。另外，对售货机的最佳补货时间应该发生在深夜的时候，并且在接近月中和月尾的时候要主要及时补满存货以免造成需求流失。最后，虽然历史数据不足以保证销售额的预测精度，但结合

第二章第三节的总销售额变化规律，在 1 月份的时候该地区的学生开始放寒假，销量发生下降的预测是合理的。