

新零售-无人智能售货机商务数据 分析报告

姓名：林苑铃

联系方式：18820971386

邮箱：943426758@qq.com

2019 年 8 月

目录

一、问题重述	2
二、数据预处理与分析	2
三、数据分析与可视化	4
3.1 2017 年 6 月销量前五商品柱状图	4
3.2 售货机总交易额折线图及月环比增长率柱状图	5
3.3 各地区毛利润占比饼图	7
3.4 二级类商品月交易额气泡图	8
3.5 售货机 C 在 6、7、8 月订单量的热力图	8
四、用户画像	10
4.1 生成饮料类商品标签	10
4.2 构建用户画像	10
五、业务预测	12
5.1 预测原理	12
5.2 预测实现	12
5.3 数据需求	15

一、问题重述

自动售货机以线上经营的理念，提供线下的便利服务，以小巧、自助的经营模式节省人工成本，让实惠、高品质的商品触手可及，成为当下零售经营的又一主流模式。自动售货机内商品的供给频率、种类选择、供给量、站点选择等是自动售货机运营者需要重点关注的问题。

本次无人智能售后机商务数据分析主要针对自动售货机的零售数据，对其数据进行基本清洗及可视化操作，使得数据反映的信息更加直观，即使是非专业人士也能从图表中获取数据能表达的信息。根据商品的销售数据，探讨销量、营业额和利润等指标的关系，并且构建各地区的售货机画像，根据各地区的特征为商售货机运营者提出策略建议或制定营销策略。最后，基于销售数据对业务进行预测，向售货机运营者提供预测结果作为策略制定的参考。

二、数据预处理与分析

首先对附件一中的数据进行初步审核和处理。通过函数检查文件中是否存在异常值，并且通过循环遍历比对订单实际金额与应付金额是否有差异，如有差异则输出有差异订单，程序运行结果如图 1：

```
In [13]: data.isnull().any()
Out[13]:
订单号      False
设备ID      False
应付金额    False
实际金额    False
商品        False
支付时间    False
地点        False
状态        False
提现        False
dtype: bool
```

图 1 缺失值检查

由图 1 可见，原数据不存在异常值，并且通过遍历对比可知数据不存在实付金额与应付金额有差异的订单。接下来定义支付时间的数据类型，即将其转换为时间类型数据。由于在对其类型进行转换时出现了错误，因此推测至少一行数据的格式异常。对于这种情况下的异常值，通常可以将数据按顺序排列并检查两端的数据。将支付时间设置为索引，排列并观察两端数据，运行结果如图 2：

```
2017/12/31 22:09      DD20170613035815879F3941D1762 ...
已提现
2017/12/31 22:39      DD20170613020607768E3940FA188 ...
已提现
2017/12/31 23:10      DD2017060217303716A53CCD6B185 ...
已提现
2017/2/29  3:44:00 PM  DD201708167493241554692026752 ...
已提现
```

图 2 支付日期异常值

由图 2 可见，最后一行的支付时间异常，将该行数据格式改正后，成功转换支付日期数据类型。

数据初步处理完成后，将附件一中的数据按地区分类，得到以地区分组的销售数据。再对数据分地区进行分析，计算每台售货机在 2017 年 5 月份的交易额及订单量，得到表 1：

表 1 五月份交易额及订单量

地区	五月份交易额	订单量
A	3385.1	756
B	3681.2	869
C	3729.4	789
D	2392.1	564
E	5699.0	1292
合计	18886.8	4270

计算每台售货机每单平均交易额及日均订单量得到表 2，其中括号内的数字为日均订单量。

地区 月份	A	B	C	D	E
1	4.51 (10)	3.75 (11)	4.33 (12)	3.69 (8)	4.68 (11)
2	3.86 (4)	3.26 (6)	3.83 (7)	3.09 (5)	3.64 (9)
3	3.58 (8)	3.61 (8)	3.77 (8)	4.31 (6)	4.31 (11)
4	4.04 (14)	4.08 (20)	4.40 (24)	3.79 (14)	4.16 (29)
5	4.48 (24)	4.07 (61)	4.50 (62)	4.03 (34)	3.82 (86)
6	4.05 (55)	4.07 (61)	4.50 (62)	4.03 (34)	3.82 (86)
7	4.09 (15)	4.40 (11)	3.99 (24)	4.23 (10)	3.92 (26)
8	3.36 (21)	3.58 (31)	3.91 (40)	3.32 (23)	3.80 (57)
9	4.31 (34)	4.13 (58)	4.43 (50)	3.90 (32)	4.13 (137)
10	4.03 (50)	4.11 (65)	4.27 (71)	3.90 (38)	3.68 (89)
11	4.47 (38)	4.27 (67)	4.35 (64)	3.86 (40)	4.28 (167)
12	3.79 (64)	3.67 (71)	3.94 (76)	3.57 (53)	4.17 (104)

三、数据分析与可视化

3.1 2017 年 6 月销量前五商品柱状图

读取预处理后的数据，使用聚合函数获取订单量前五的商品名称及销量，根据函数获取的结果可得图 3：

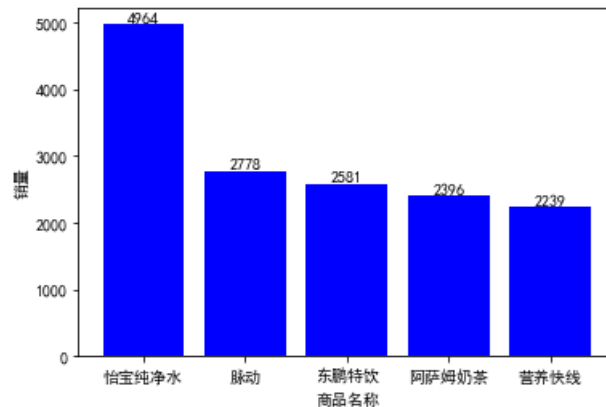


图 3 销量前五商品

由图 3 可见：2017 年 6 月销量前五的商品均为饮料类商品，其中销量最高的怡宝纯净水以 4964 的销量远远领先于其余四种饮料，其余四种饮料的销量差异相对较小。因此可以推测：人们对自动售货机种的饮料类商品的需求比对非饮料类商品的需求高，并且在饮料类商品种，纯净水的需求是最高的。可以根据该结论适当提升各地区纯净水的进货量。

3.2 售货机总交易额折线图及月环比增长率柱状图

为了获取月交易总额数据，我们将支付时间设置为索引，按月计算交易总额，

并通过公式： $\frac{M(i)-M(i-1)}{M(i-1)}$ 计算交易额月环比增长率，其中 $M(i)$ 为第 i 期的交

易额。

获取并算得了需要的数据后可以绘制出交易额折线图及月环比增长率图，见图 4 到图 8。由各地区的总交易额折线图及环比增长率图可见：各地区的交易总额变化趋势大致相同，即交易额逐步上升，但至七月骤降，后从七月起交易额逐步回升。总体来说，2017 年春季人们对自动售货机中的商品需求量较小，下半年各地区人群对售货机商品的需求量达到均达到一个较高的水平。排除季节的影响，引起这种需求量变化的可能是人们对自动售货机的了解程度及观念的改变，即在 2017 年初，各地区自动售货机在各地区部署，但是人们知之甚少，因此售货机销量普遍不高，随着时间推移越来越的人熟悉了售货机，售货机凭借其便利性吸引力越来越多的客户。

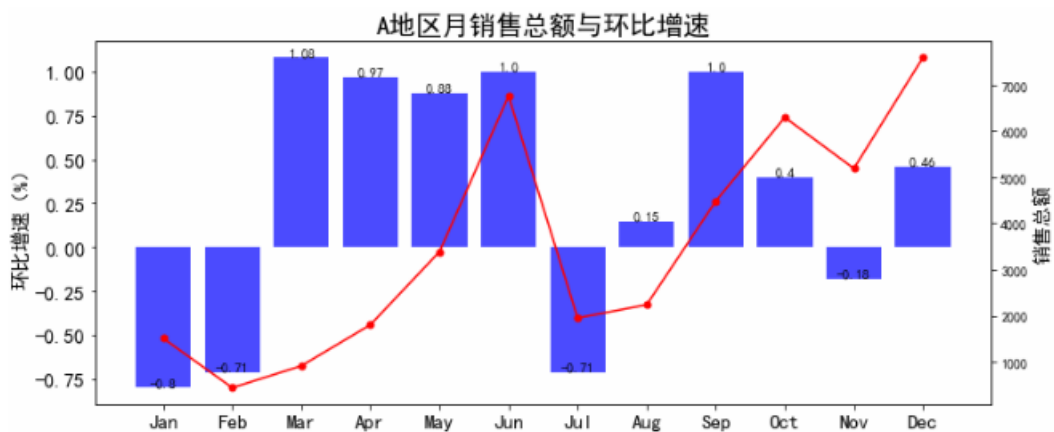


图 4 A 售货机总交易额折线图及月环比增长率柱状图

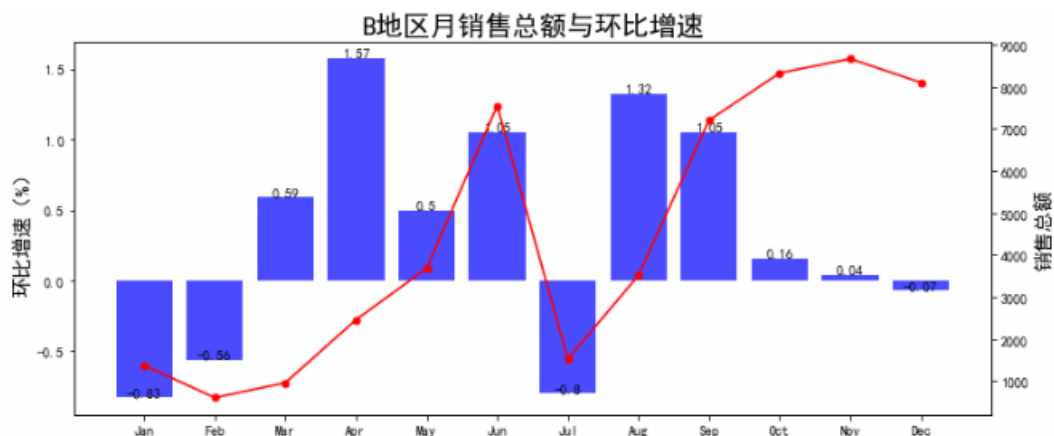


图 5 B 售货机总交易额折线图及月环比增长率柱状图

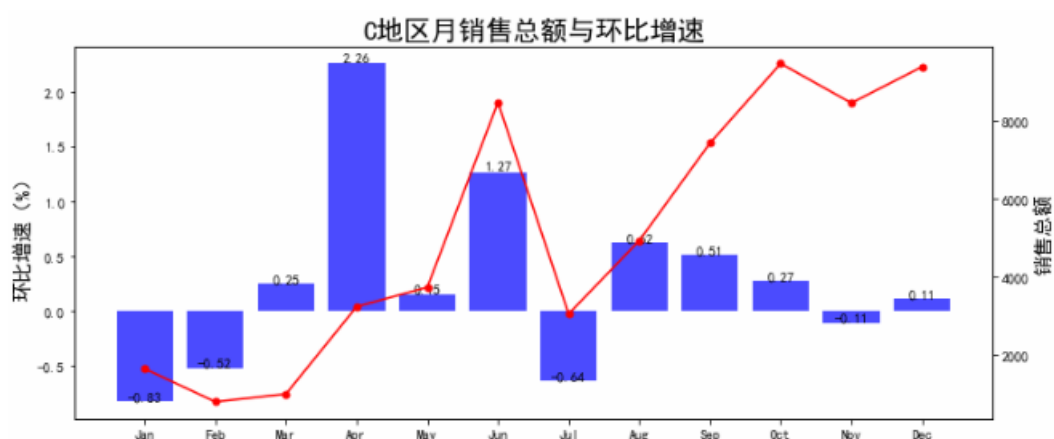


图 6 C 售货机总交易额折线图及月环比增长率柱状图

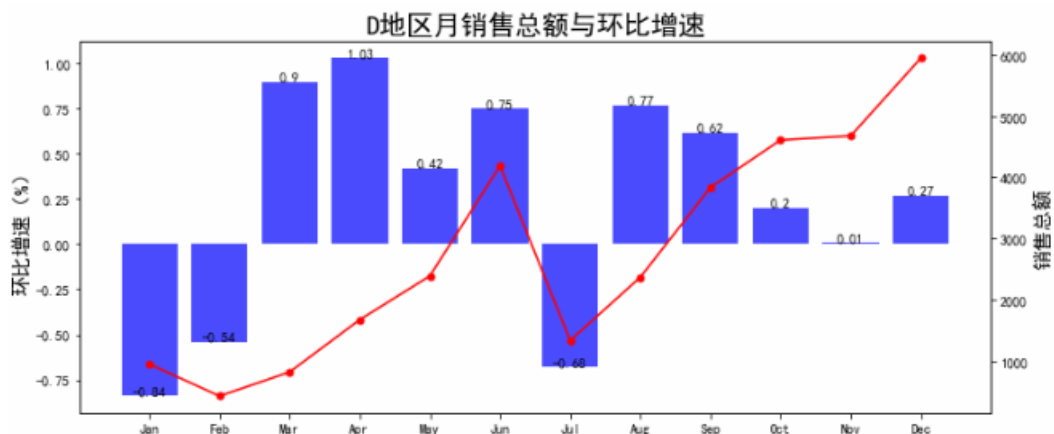


图 7 D 售货机总交易额折线图及月环比增长率柱状图

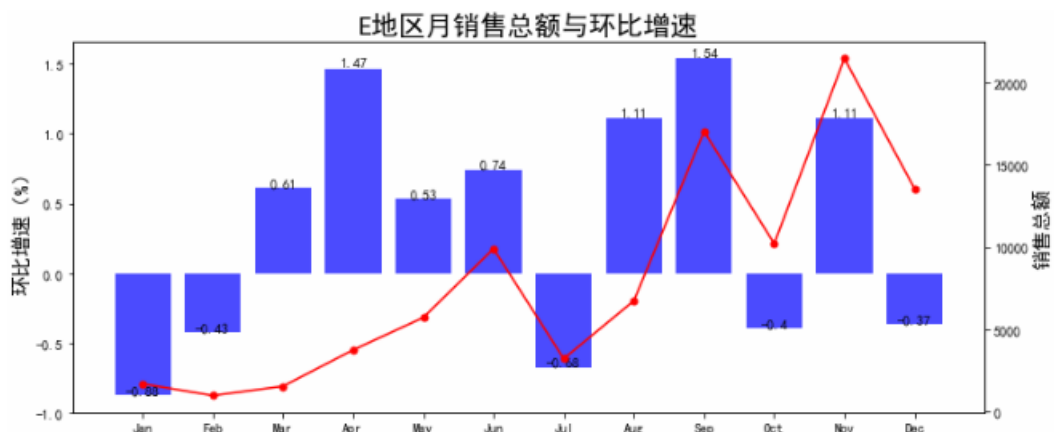


图 8 E 售货机总交易额折线图及月环比增长率柱状图

3.3 各地区毛利润占比饼图

假定饮料类商品的毛利率为 25%，非饮料类商品的毛利率为 20%，利用商品大类数据对商品进行饮料类及非饮料类的分类。计算分类后的交易总额，并带入假定的毛利率即可得到各地区毛利润占总毛利润的饼图。由图 9 可见：各地区对总毛利润的贡献值顺序为：E>C>B>A>D，E 地区毛利润对总毛利润的贡献最大，这也符合前文的结论；而 D 地区的毛利润则对总毛利润的贡献最小，由此可推测自动售货机在 D 地区的受欢迎程度不如另外四个地区。针对各地区毛利润的差异可采取差异性的营销措施，拓展低销量地区的客户稳住高销量地区的客户。



图 9 各地区毛利润占比饼图

3.4 二级类商品月交易额气泡图

将商品按二级类名称及月份分组可得到各月的各二级类商品交易额，将气泡的大小与交易额的大小关联，可以得到如图 10 的气泡图。由二级类商品交易额气泡图可见，香烟类商品的交易额显著高于其余类商品的交易额，这是由商品单价决定的。而膨化食品和纸巾等类型的商品则一直维持在较低的交易额水平，这也一定程度上反映了人们在自动售货机购买商品时，对不同类型的商品需求是不一样的。

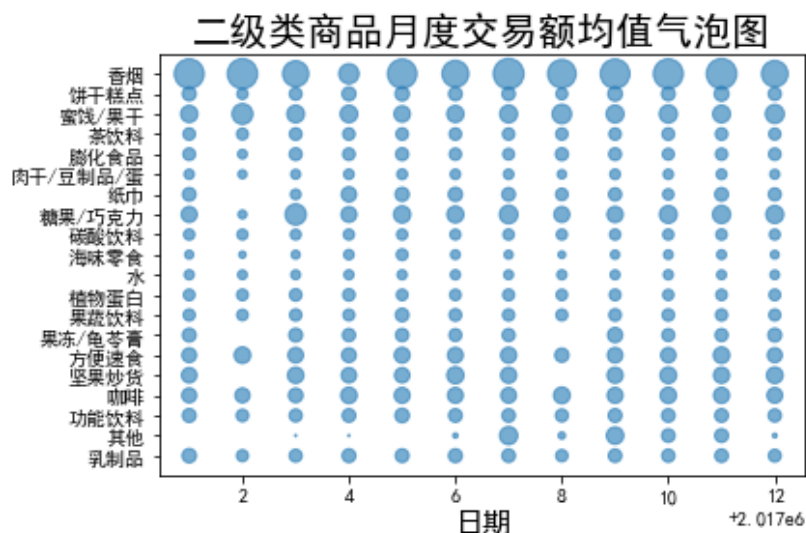


图 10 二级类商品月交易额气泡图

3.5 售货机 C 在 6、7、8 月订单量的热力图

以小时为频率获取 C 地区的售货机数据，根据日期和小时数据分组，计算各组内的订单量，生成以小时为横轴、日期为纵轴、订单量为值的矩阵。根据矩阵生成热力图得到如图 11 至图 13。通过直观地观察，在 C 地区的 6、7、8 月中，

6 月订单量相对较多且订单分布较均匀，7 月的订单量最少且大部分订单集中在月尾。凌晨 0 点到早上 7 点之间，取自动售货机购买商品的人极少，大部分的人都是在早上 9 点到晚上 7 点之间购买自动售货机的商品。这说明只有在工作时间内自动售货机售出的商品较多，一旦人们回到家中，将基本不会购买自动售货机中的商品，即自动售货机的主要受众是那些需要在外工作或游玩的人，对于家庭主妇而言，自动售货机显然不存在吸引力。

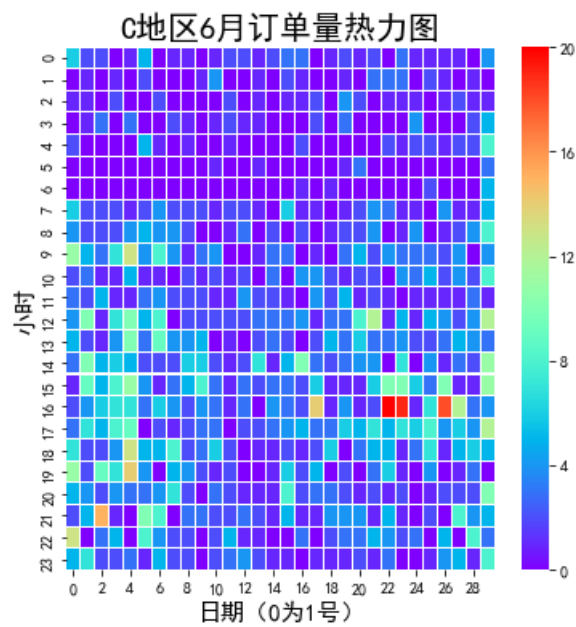


图 11 C 售货机 6 月订单量热力图

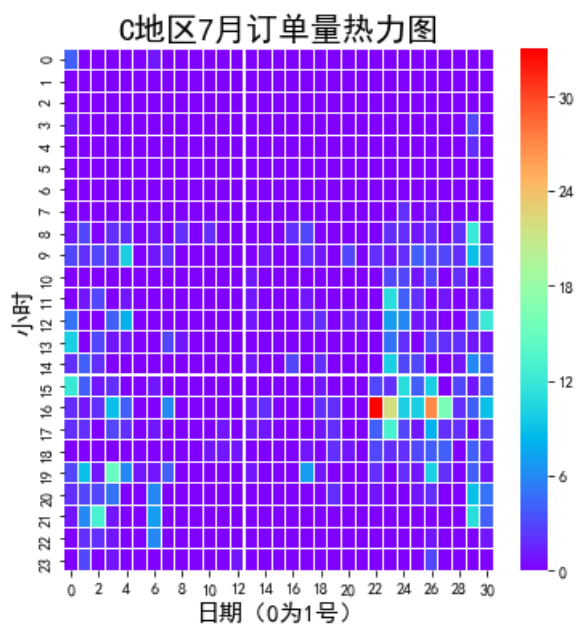


图 12 C 售货机 7 月订单量热力图

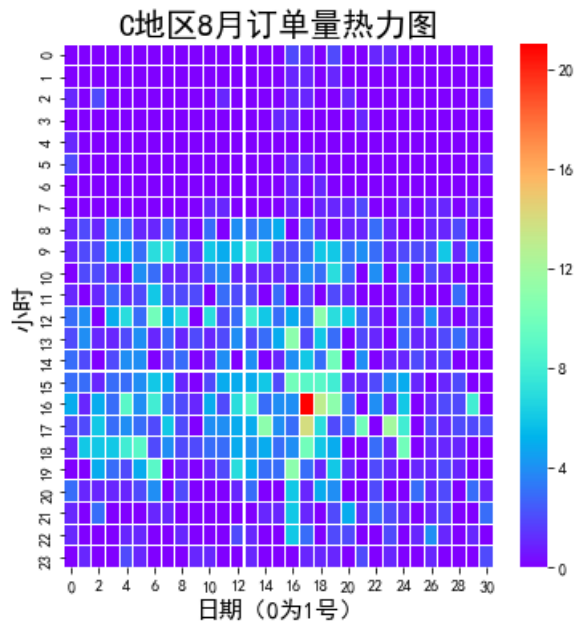


图 13 C 售货机 8 月订单量热力图

四、用户画像

4.1 生成饮料类商品标签

按地区分类，将饮料类商品数据写入文件“某地区饮料类商品数据”中，获取商品销量的描述性统计量，我们将销量大于 75%商品的归为畅销商品，销量小于 75%商品的归为滞销商品，在二者中间的则为正常商品。依据该规则可生成商品销量标签，生成标签表示例如表 3。

表 3 标签表示例

序号	商品名称	订单量	标签
1	怡宝纯净水	531	畅销
2	雪碧 330ML	67	正常
3	可口可乐 330ML(罐)	5	滞销

4.2 构建用户画像

由于时间及任务进度的原因，用户画像部分还未完成。

五、业务预测

5.1 预测原理

在计算科学蓬勃发展的今天，各种预测方法如雨后春笋般地出现，不论是传统的回归预测还是新兴的深度学习，所有的预测方法的基本原理都有共同性。几乎所有的预测方法都遵循下列三个原则：

- 惯性原理，即事物的某个属性或结构会在一个周期内具有延续性；
- 相关性原理，即许多事物之间都存在直接或间接的联系；
- 相似性原理，即根据存在共同属性的事物发展类比另一事物的发展。

5.2 预测实现问题

数据量不足，原始数据中仅有 2017 年一年的数据，无法提取数据以年为周期的特征，这种信息的损失会使预测出现较大的误差。若需要更精准的预测，则需对数据的时间长度进行拓展，例如获取历史数据或在未来持续关注并记录数据，以此不断调整模型提高精度。

在如今的信息时代，数据的重要性与日俱增，加上计算机性能及各种算法发展的加持，原本杂乱无章的数据中能获取的信息是极其有价值的。企业可以通过数据分析捕获 市场剩余，对客户采用精准营销策略，这些工作对企业具有重大意义，大数据显现的优势并不亚于石油带来的利益。可以说：掌握了数据的企业才能掌握市场的主动权。