

新零售无人智能售货机商务数据分析

目录

一、背景介绍	2
1. 行业背景	2
2. 分析目标	2
3. 数据说明	2
二、数据预处理与分析.....	3
三、数据可视化分析.....	6
1. 2017 年 6 月销量前五的商品及其销量柱状图	6
2. 各售货机每月总交易额折线图与月环比增长率柱状图	6
3. 每台售货机毛利润占总毛利润的饼图.....	8
4. 每月交易总额均值的气泡图	9
5. 售货机 C 在 2017 年 6、7、8 月份订单量的热力图	11
1. 贴标签	13
2. 生成各自动售货机画像	13
四、销售额预测	16
1. 预测未来销售额的原理.....	16
2. 关于本案例中销售额的预测	17
五、小结	19

一、背景介绍

1. 行业背景

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。因此，科学的商业数据分析能够帮助经营者了解用户需求，掌握商品需求量，为用户提供精准贴心的服务，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有着非常重要的意义。

2. 分析目标

本数据分析报告的对象是某商场在不同地点安放的编号分别为 A、B、C、D、E 的 5 台自动售货机从 2017 年 1 月 1 日至 2017 年 12 月 31 日的商品销售数据，分析的主要目标为：

➤ 根据自动售货机的经营特点，对经营指标数据、商品营销数据及市场需求进行分析，完成对销量、库存、盈利三个方面各项指标的计算，绘制对应图表，并预测每台售货机的销售额。

➤ 为每台售货机所销售的商品贴上标签，使其能够很好地展示销售商品的特征。

3. 数据说明

附件 1 提供了某商场在不同地点安放的编号分别为 A、B、C、D、E 的 5 台自动售货机从 2017 年 1 月 1 日至 2017 年 12 月 31 日的商品销售数据。

附件 2 提供了商品的分类。

二、数据预处理与分析

数据预处理可具体分为清洗数据、合并数据、转化数据和标准化数据，四者没有先后之分。本次分析报告的数据预处理主要集中在清洗数据上，包括检查与处理缺失值、检查与处理重复值、检查与处理异常值。

其中对于重复值，主要考虑记录重复，没有考虑特征重复，因为在这里特征重复可以通过有目的地忽略或者删除相关列解决。对于异常值，有一条如下记录的支付时间有异常，2017 年没有 02 月 29 日，删除该条记录。

70681	DD20170816749324155E43A6E078A04228	4	4 55g 奥利奥原味芝士饼干	2017/2/29 3:44:00 C	已出货未退已提现
-------	------------------------------------	---	-----------------	---------------------	----------

关于合并数据，转化数据和标准化数据，在本次报告中也分别有一定程度的涉及。例如，在可视化数据之前需要把清洗过后的数据合并到同一张表格里，把销售信息关联起来；在日期格式的处理中，需要把“支付时间”的类型从 `object` 转化为 `datetimes`。

数据预处理过后，我提取了每台售货机对应的销售数据，保存在 CSV 文件中。文件名分别为“task1-1A.csv”、“task1-1B.csv”、“task1-1C.csv”、“task1-1D.csv”、“task1-1E.csv”。

第二，我计算了每台售货机 2017 年 5 月份的交易额、订单量，每台售货机每月的每单平均交易额以及每月的日均订单量。制成表格，分别如下：

	2017年5月份交易额	2017年5月份订单量
A	2816.8	756
B	3223.8	869
C	2993.8	789
D	2047.9	564
E	4996.1	1292
sum	16078.4	4270

表 1 每台售货机 2017 年 5 月份的交易额、订单量

每台售货机每月的每单平均交易额					
地点 月份	A	B	C	D	E
1	3.600597	3.416667	3.646966	3.427799	3.773729
2	3.280702	3.139459	3.567633	2.975177	3.404264
3	3.177255	3.346415	3.526616	3.479167	4.134286
4	3.620134	3.502156	3.754087	3.529571	3.586927

5	3. 725926	3. 709781	3. 794423	3. 631028	3. 866950
6	3. 505333	3. 550108	3. 580553	3. 631442	3. 489780
7	3. 576261	3. 617971	3. 533246	3. 497792	3. 493358
8	3. 120270	3. 399796	3. 621922	3. 179720	3. 612960
9	3. 664423	3. 489341	3. 641776	3. 454934	3. 668191
10	3. 487987	3. 599062	3. 692554	3. 476560	3. 506986
11	3. 737759	3. 493255	3. 646629	3. 463306	3. 771972
12	3. 398952	3. 386697	3. 566877	3. 306494	3. 703690

表 2 每台售货机每月的每单平均交易额

每台售货机每月的日均订单量					
地点 月份	A	B	C	D	E
1	10. 806452	11. 806452	12. 225806	8. 354839	11. 419355
2	4. 071429	6. 607143	7. 392857	5. 035714	9. 214286
3	8. 225806	8. 548387	8. 483871	6. 193548	11. 290323
4	14. 900000	20. 100000	24. 466667	14. 766667	29. 833333
5	24. 387097	28. 032258	25. 451613	18. 193548	41. 677419
6	55. 633333	61. 866667	62. 733333	34. 666667	86. 433333
7	15. 354839	11. 129032	24. 645161	10. 225806	26. 225806
8	21. 483871	31. 645161	40. 612903	23. 064516	57. 000000
9	34. 666667	58. 166667	55. 933333	32. 766667	137. 800000
10	50. 483871	65. 354839	71. 483871	38. 258065	89. 580645
11	38. 666667	67. 700000	64. 766667	40. 333333	167. 333333
12	64. 612903	71. 290323	76. 741935	53. 645161	104. 903226

表 3 每台售货机每月的日均订单量

三、数据可视化分析

我把之前清洗数据得到的每台售货机的销售数据纵向合并在同一张表格里，并进行了以下的可视化分析：

1. 2017 年 6 月销量前五的商品及其销量柱状图

由图 1 可知，2017 年 6 月销量前五的商品主要是饮料类，唯一一种非饮料类是 40 克双汇玉米热狗肠。其中怡宝矿泉水的销量约为其他商品销量的 2.5 倍，其他四种商品销量不相上下。这与天气热时户外活动的人们倾向于饮水解渴的现象相符。

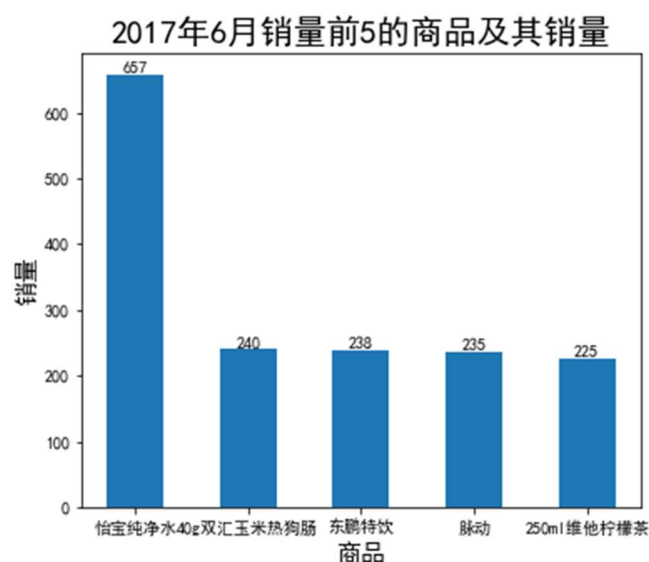


图 1 2017 年 6 月销量前五的商品及其销量柱状图

2. 各售货机每月总交易额折线图与月环比增长率柱状图

由图 2 可知，每台售货机每月总交易额的走向基本一致，2 月份是所有售货机的交易额的最低峰，这可能与中国的传统节日春节有关，人们外出旅游或者家里好吃的食物较多，很少有顾客光顾停放在

固定地点的售货机。3、4、5、6 月份的交易额逐渐升高，但到了 7 月份天气酷热，人们很少户外活动，导致 7 月份交易额剧减。之后交易额逐渐增加。有一个有趣的现象，A、B、C、D 处的售货机在 11 月份的交易额均减少，而 E 处售货机的交易额陡增，达到 E 处售货机交易额的最高峰，甚至是所有售货机月交易额中的最大值。这可能是因为在 11 月份出现某种热闹的活动，把主要的人流量都吸引到了 E 处。在图 3 中，通过每台售货机交易额月环比增长率的柱状图和上面标记的数据更能科学有力地比较每台售货机月交易额的情况。

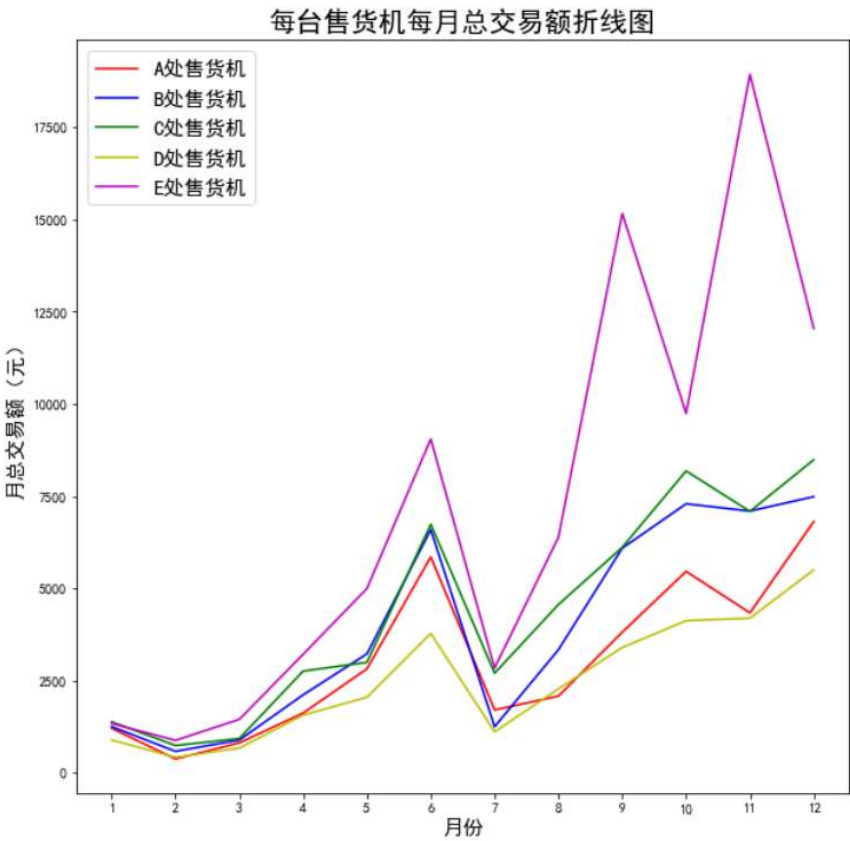


图 2 各售货机每月总交易额折线图

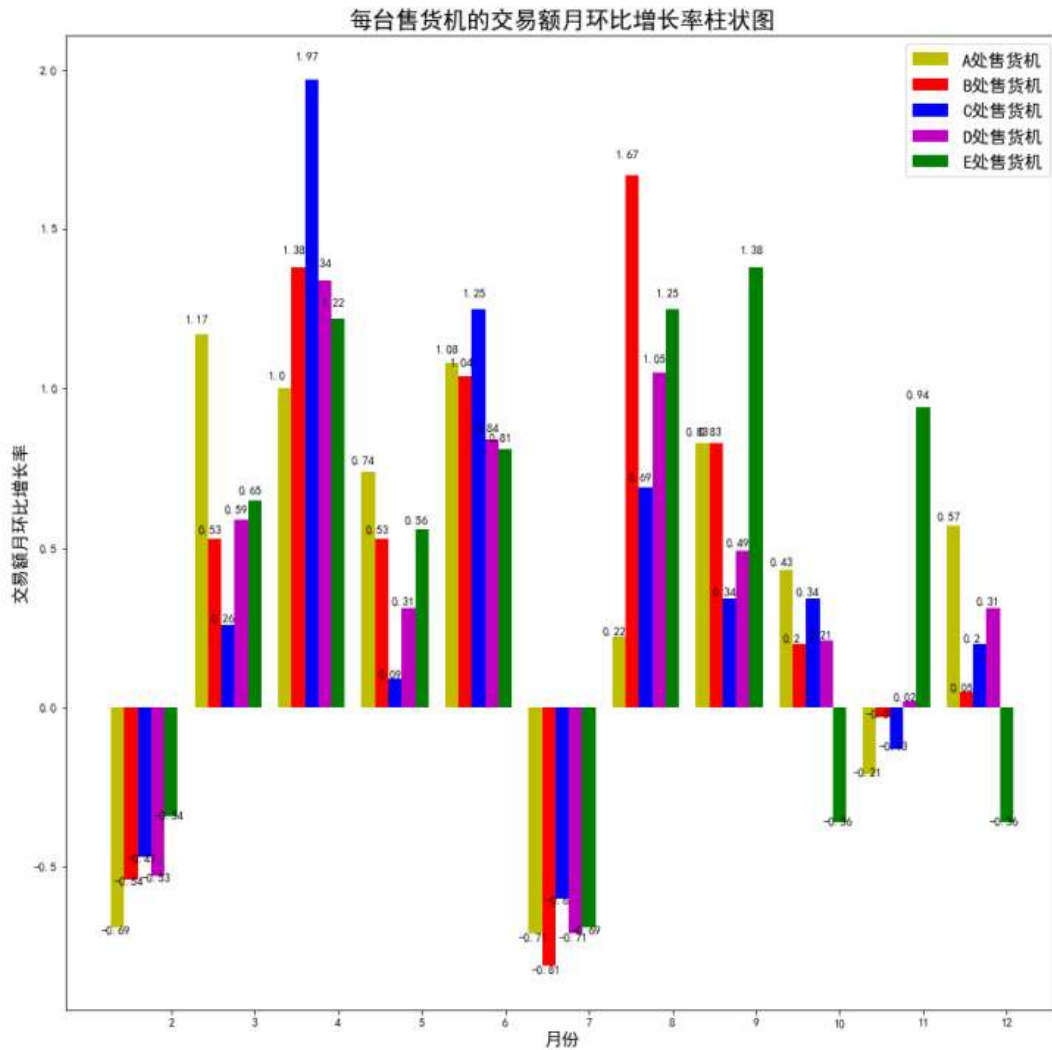


图3 每台售货机的交易额月环比增长率柱状图

3. 每台售货机毛利润占总毛利润的饼图

在图4中,假设饮料类的毛利率为25%,非饮料类为20%,由每台售货机的饮料类和非饮料类的总交易额分别乘以相应的毛利率再相加得到各台售货机的毛利润,由此画出每台售货机毛利润占总毛利润的饼图。由饼图可知,E出售货机所赚的毛利润占比最大,为33.6%,接着依次递减的是C处售货机、B处售货机、A处售货机,D处售货机的毛利润占比最小,为11.6%。

每台售货机毛利润占总毛利润的饼图

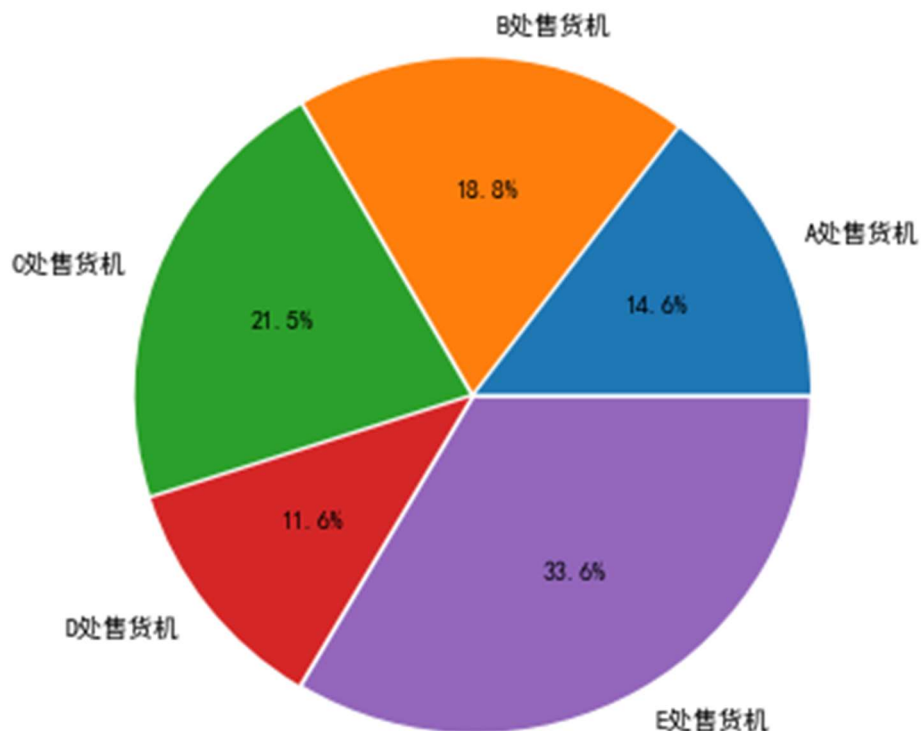


图 4 每台售货机毛利润占总毛利润的饼图

4. 每月交易总额均值的气泡图

图 4 是每月交易总额均值的气泡图，横轴是月份，纵轴是商品的二级类目，气泡的大小对应交易额均值的多少。纵向观察可发现，2 月的交易总额均值整体最小，而 6、9、10、11、12 月份的交易总额均值整体都很大。横向观察可发现，各类商品的月交易总额均值较大的月份主要集中在：

香烟：8、9、11、12 月份

方便速食：6、11 月份

纸巾：9、11 月份

饼干糕点、水：6、9、10、11 月份

蜜饯/果干和茶饮料:10、11、12 月份

海味零食、植物蛋白:6、11、12 月份

果冻/龟苓膏、其他:9、11、12 月份

膨化食物、肉干/豆制品/蛋、糖果/巧克力:6、10、11 月份

碳酸饮料、果蔬饮料、咖啡、功能饮料、乳制品:9、10、11、12 月份。

此外，各类商品的月交易总额均值的最大值和最小值相近，没有出现哪种商品的月交易额均值异常小或者异常大。

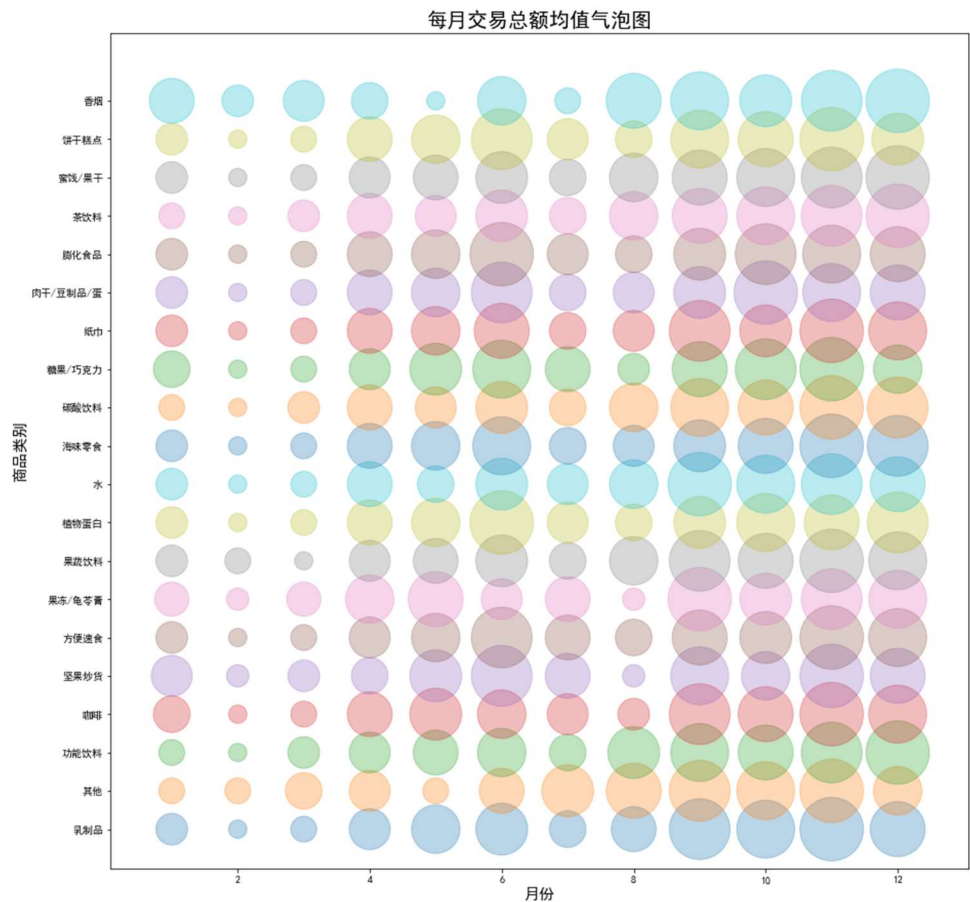


图 5 每月交易总额均值的气泡图

5. 售货机 C 在 2017 年 6、7、8 月份订单量的热力图

图 6 至图 8 分别为售货机 C 在 2017 年 6、7、8 月份订单量的热力图，横轴以天为单位，纵轴以小时为单位，图例表示颜色越黄订单量越多，颜色越绿订单量越少。观察热力图可以发现，售货机 C 在 2017 年 6 月 23 和 24 日下午 4 点这一小时订单量最多，在 6 月 1 号至 8 号出现订单量数量较多的时间段集中在早上 9 点至晚上 10 点。其他时间段都有较少的订单量出现。售货机 C 在 2017 年 7 月份订单量分布较 6 月份明显稀疏了很多。但在 7 月 21 号中午 12 点订单量达到 7 月份的最大值，之后 22 号至 2 号中午 12 点也有较多的订单量，这可能与当地在这几天中午举办了热闹的活动，吸引了人流有关。售货机 C 在 2017 年 8 月份的订单量分布和 6 月份一样，都是比较均匀的，订单量主要集中在早上 8 点至晚上 22 点，在 8 月 18 号下午 4 点这一小时内订单量达到 8 月份的最大值。

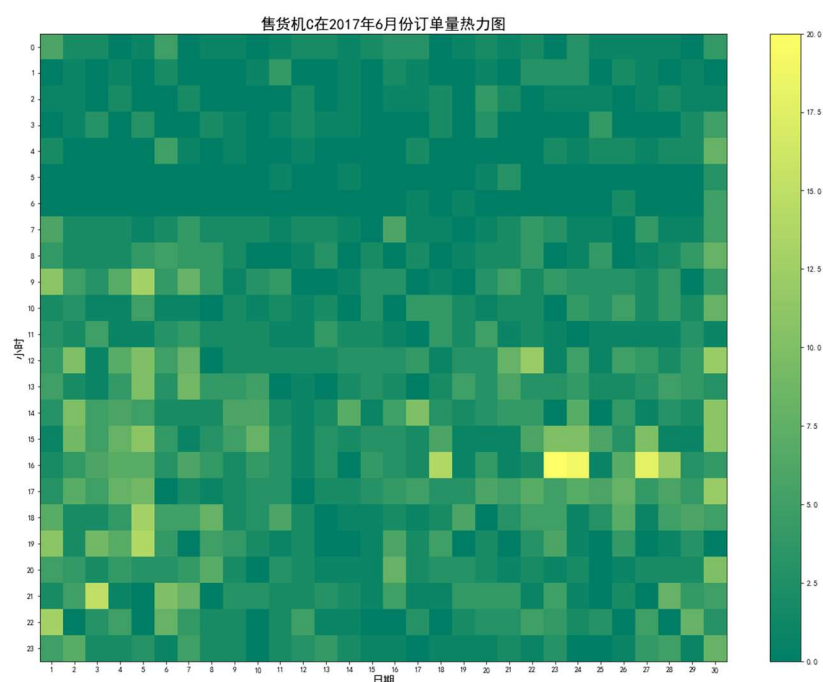


图 6 售货机 C 在 2017 年 6 月份订单量的热力图

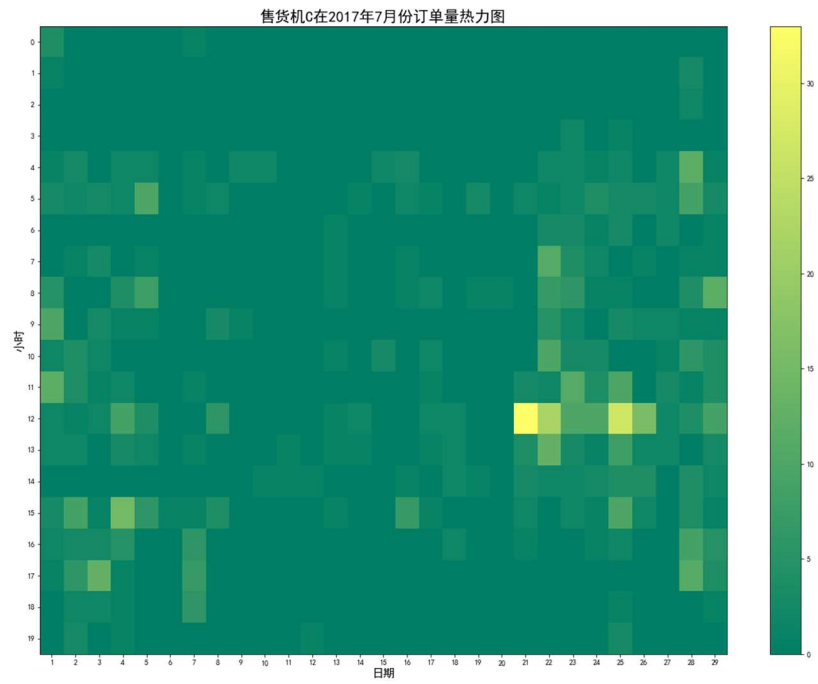


图 7 售货机 C 在 2017 年 7 月份订单量的热力图

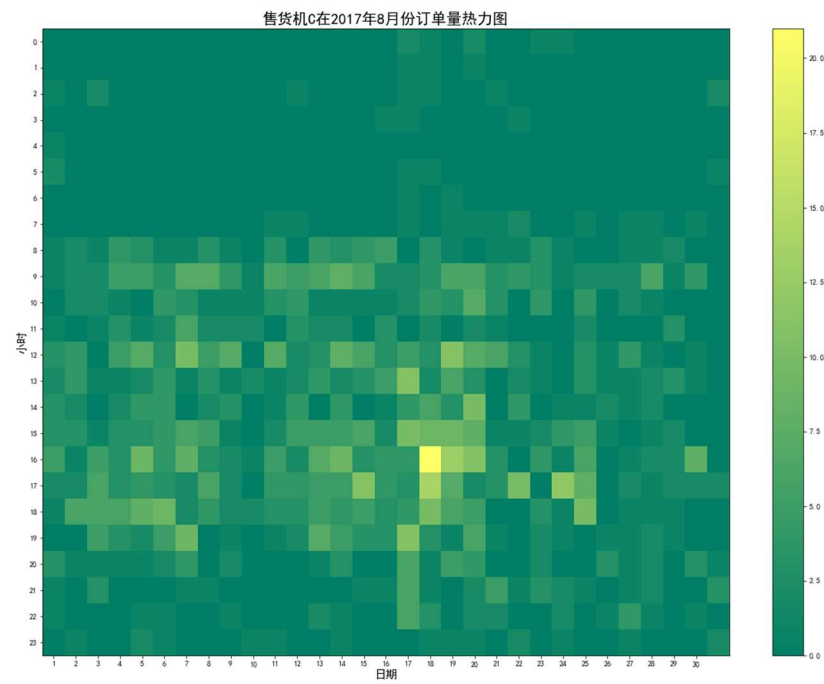


图 8 售货机 C 在 2017 年 8 月份订单量的热力图

三、生成各自动售货机画像

1. 贴标签

为了给出每台售货机饮料类商品的标签：滞销/正常/热销，我定义了“销量占比”的指标：

$$\text{销量占比} = \frac{\text{该饮料商品的销量}}{\text{所有饮料类商品的销量}}$$

算出所有饮料类商品的销量占比后按值其从大到小排列得到一个序列，使用 `quantile()` 函数分别计算出该序列的 20%分位数 `a` 和 80%分位数 `b`。定义销量占比小于 `a` 的商品标签为“滞销”，定义销量占比大于 `b` 的商品标签为“畅销”，销量占比介于 `a` 和 `b` 之间的商品标签是“正常”的。用这种方法，我们对各售货机的所有饮料类商品的热销程度贴上标签保存在 `csv` 文件中，文件名分别为“`task3-1A.csv`”、“`task3-1B.csv`”、“`task3-1C.csv`”、“`task3-1D.csv`”、“`task3-1E.csv`”。

2. 生成各自动售货机画像

导入 `wordcloud` 扩展包，并为每台售货机选择特定的背景图，依据商品的标签绘制如下的饮料类销售词云图，根据词云图字体的大小，可以直观地看到：售货机 A 的热销饮料有伊利纯牛奶、雪碧最新款、268ml 雀巢咖啡丝滑拿铁等；售货机 B 的热销饮料有银鹭八宝粥、伊利纯牛奶、雪碧最新款、红牛等；售货机 C 的热销饮料有 250ml 燕塘红枣牛奶、250ml 燕塘甜牛奶等；售货机 D 的热销饮料有美年达（罐）、250ml 维他原味豆奶、伊利优酸乳、王老吉（500ml）等、

[illegible]

银鹭八宝粥
王老吉 (500ml)
250ml 燕塘原味酸奶
250ml 燕塘甜牛奶
燕塘原味酸奶 旺仔牛奶 145ml
雪碧 (500ml)
营养快线
统一绿茶
250ml 维他柠檬茶
王老吉 (罐)
珠江纯生啤酒
250ml 维他原味豆奶
伊利纯牛奶

果粒橙
维他奶 (巧克力味)
怡宝纯净水
芬达 330ml
可口可乐 330ml
阿萨姆奶茶
统一冰红茶
农夫果园
脉动
东鹏特饮

145ml 旺仔牛奶盒装
雪碧最新款
伊利优酸乳 红牛

14





图 13 售货机 E 饮料类销售画像

四、销售额预测

1. 预测未来销售额的原理

销售额预测是销售预测的一个组成部分，销售预测是指对未来特定时间内，全部产品或特定产品的销售数量与销售金额的估计。结合一般规律，在进行本次新零售无人智能售货机商务数据分析中的未来销售额预测时，需要考虑两大类因素：

（1）、外界因素

➤ 需求动向

需求是外界因素之中最重要的一项。如人口流动、流行趋势、爱好变化等，均可成为自动售货机商品需求的质与量方面的影响因素，因此，必须加以分析与预测。

➤ 经济变动

销售额深受经济变动的影响,经济因素是影响商品销售的重要因素,例如,销售额的增加不一定与销售数量直接相关,要考虑市场价格的影响。为了提高未来销售额预测的准确性,应特别关注商品市场中的供应和需求情况、商品原材料的未来发展以及基础工业、加工业生产、经济增长率等指标变动情况。尤其要关注突发事件对经济的影响。

➤ 同业竞争动向

销售额的高低深受同业竞争者的影响,预测自动售货机销售额需要了解竞争对手比如零售店、便利店、其他地点的自动售货机等在市场上的活动。例如,竞争对手的目标市场在哪里,产品价格高低,促销与服务措施等等。

(2)、内部因素

➤ 营销策略

市场定位、产品政策、价格政策、渠道政策、广告及促销政策等变更都会对自动售货机销售额产生影响。

➤ 库存状况

自动售货机里的商品是否充足,能否保证销售需要等。

2.关于本案例中销售额的预测

我认为附件提供的数据不足以对每台售货机的每个大类商品在2018年1月的交易额进行预测。分析如下:

- 1) 本案例中因变量是时间，自变量销售额是时间的函数，符合时间序列分析法的基本特征：根据过去的变化趋势预测未来的发展，它的前提是假定事物的过去延续到未来，即不会发生突然的跳跃变化，是以相对小的步伐前进以及过去和当前的现象可能表明现在和将来活动的发展变化趋向。本案例要求预测第二年 1 月的交易额，属于短期预测。
- 2) 本案例中，时间序列中的每个观察值即销售额的大小，是各种不同因素在同一时刻发生作用的综合结果。由前面数据可视化可以发现，这些因素造成的时间序列销售额数据的变动有如下几个特点：
- 趋势性：2017 年自动售货机销售额随着时间进展，整体上呈现一种上升的趋势。
 - 季节性：2017 年自动售货机销售额由于外部影响随着自然季节的交替出现高峰与低谷的规律。
 - 随机性：个别偶然因素对时间序列产生影响，没有规律。
- 此外，还包括根据附件不能得到的，但是客观存在的自动售货机销售额的另一特点：
- 周期性：循环波动（cyclical fluctuation），是时间序列中呈现出来的围绕长期趋势的一种波浪形或振荡式波动。不同于季节变动，季节变动有比较固定的规律，且变动周期大多为一年，循环波动则无固定规律，变动周期多在一年以上，且周期长短不一。周期性通常是由经济环境的变化引起。

因此，根据传统时间序列分析法原理上可以预测自动售货机未来销售额。但是计算时间序列的成分：趋势（T）、季节性或季节变动（S）、周期性或循环波动（C）、随机性或不规则波动（I）需要连续 2 年或者 2 年以上的有关销售数据，而附件只提供了 1 年的数据，即使预测出来，准确性也是非常令人担忧的。

我认为如果要预测自动售货机未来销售额，经营者需要提供：

- 连续 2 年或者 2 年以上的有关自动售货机销售数据，这是最关键的；
- 有关自动售货机的市场资料、消费者购买动机调查等统计资料，以掌握市场的需求动向等等。

五、小结

本次新零售无人智能售货机商务数据分析先对数据进行了预处理，包括清洗数据，合并数据，转化数据及标准化数据；然后对经营指标数据、商品营销数据及市场需求进行分析，完成对销量、库存、盈利三个方面各项指标的计算，绘制对应图表，可视化数据。接着为每台售货机所销售的商品贴上标签，并绘制各台自动售货机的饮料类的销售画像。最后对附件提供的数据不足以对每台售货机的每个大类商品在 2018 年 1 月的交易额进行预测做了分析说明。