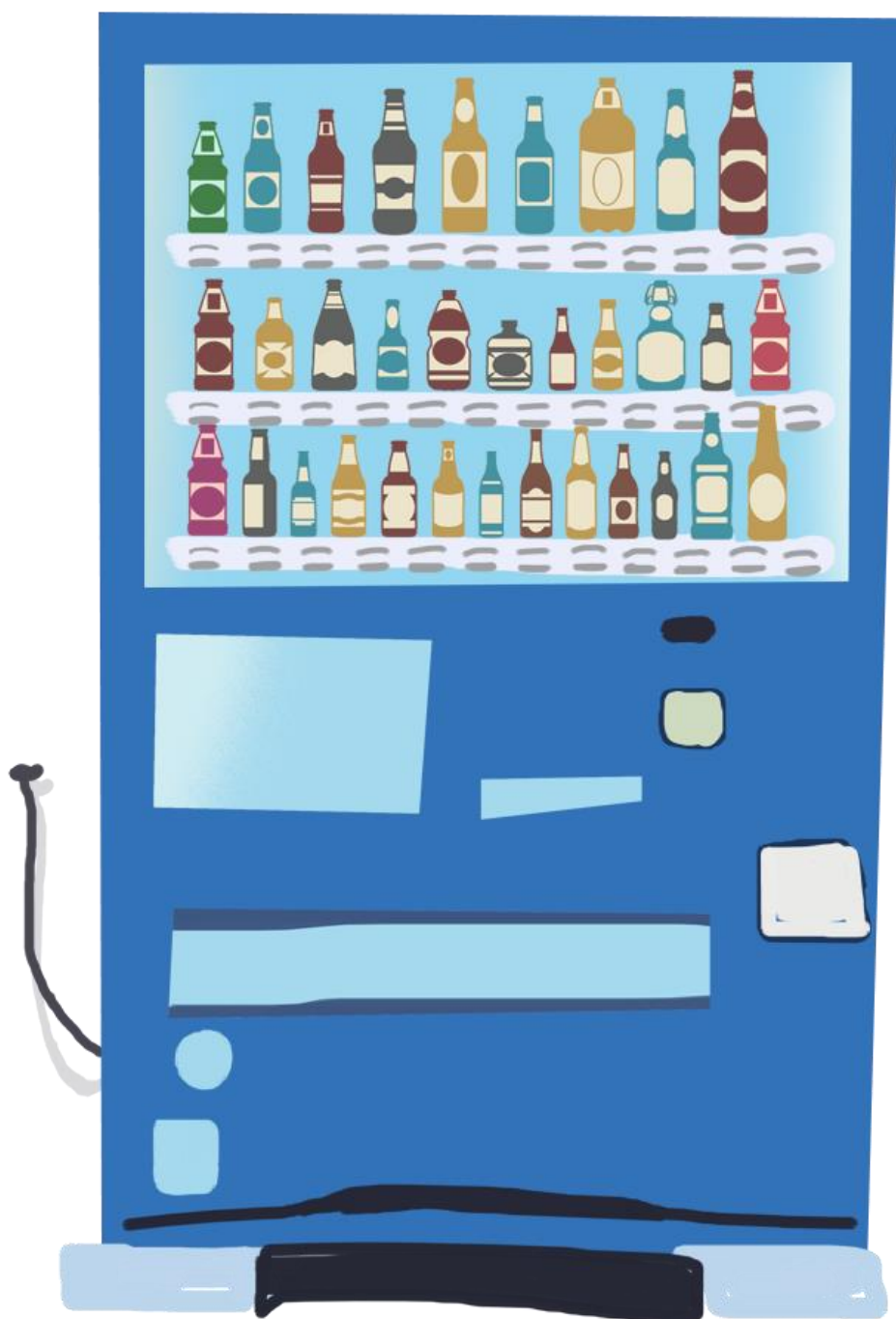


新零售-无人智能售货机商务数据分析报告

郭锦红



目录

- 一、 背景介绍..... 1
 - 1.1. 行业背景 1
 - 1.2. 分析目标 1
- 二、 数据说明..... 1
 - 2.1. 数据来源 1
 - 2.2. 数据内容及示例数据..... 1
 - 2.3. 数据量 2
- 三、 数据预处理与分析..... 2
 - 3.1. 数据预处理..... 2
 - 3.1.1. 合并数据..... 2
 - 3.1.2. 清洗数据..... 2
 - 3.1.3. 转换数据..... 3
 - 3.2. 数据简易分析 3
 - 3.2.1. 售货机五月份交易情况 3
 - 3.2.2. 售货机各月统计情况 4
- 四、 数据分析与可视化..... 6
 - 4.1. 六月售货机销量前 5 商品 6
 - 4.2. 交易额变化分析..... 7
 - 4.3. 毛利润分析..... 8
 - 4.4. 商品在不同月份出售情况分析 9
 - 4.5. 六、七、八 月订单随时间分析 10
- 五、 自动售货机画像 12
 - 5.1. 饮料类商品—销售热度标签 12
 - 5.2. 饮料类商品—消费档次标签 13
 - 5.3. 自动售货机销量画像..... 15
 - 5.4. 营销意见 18
- 六、 业务预测 18
 - 6.1. 预测原理 18
 - 6.2. 预测方法 19

6.2.1. 定性预测.....	19
6.2.2. 定量预测.....	20
6.3. 可预测性判断及解释.....	20
6.4. 预测数据需求	20
七、 小结	21

一、背景介绍

1.1.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。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。

1.2.分析目标

某商场在不同地点安放了 5 台自动售货机，编号分别为 A、B、C、D、E。附件 1 提供了从 2017 年 1 月 1 日至 2017 年 12 月 31 日每台自动售货机的商品销售数据，附件 2 提供了商品的分类。

1. 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，按要求绘制对应图表，并预测每台售货机的销售额。
2. 为每台售货机所销售的商品贴上标签，使其能够很好地展现销售商品的特征。

二、数据说明

2.1.数据来源

数据来源：某商场的 5 台不同地点的自动售货机

数据时间：2017 年 1 月 1 日至 2017 年 12 月 31 日

2.2.数据内容及示例数据

销售数据（附件 1）包括了全部 5 台售货机的每笔订单的具体情况（包括订单号，设备 ID，应付金额，实际金额，商品，支付时间，地点，状态，提现情

况)。

订单号	设备ID	应付金额	实际金额	商品	支付时间	地点	状态	提现
DD201708	E43A6E078	4.5	4.5	68g好丽友	2017/1/1 0:53	D	已出货未退款	已提现
DD201708	E43A6E078	3	3	40g双汇玉	2017/1/1 1:33	A	已出货未退款	已提现
DD201708	E43A6E078	5.5	5.5	430g泰奇/	2017/1/1 8:45	E	已出货未退款	已提现
DD201708	E43A6E078	5	5	48g好丽友	2017/1/1 9:05	C	已出货未退款	已提现
DD201708	E43A6E078	3	3	600ml可口	2017/1/1 9:41	B	已出货未退款	已提现

图表 2.2-1 销售数据示例

分类数据 (附件 2) 提供的则是售货机出售的每种商品对应的大类 (饮料类、非饮料类) 和二级类 (包括乳制品、饼干糕点、肉干/豆制品/蛋等在内)。

商品	大类	二级类
100g*5瓶益力多	饮料	乳制品
100g越南LIPO奶味面包干	非饮料	饼干糕点
10g卫龙亲嘴烧香辣味	非饮料	肉干/豆制品/蛋
10g越南LIPO奶味面包干	非饮料	饼干糕点
110g顺宝九制话梅	非饮料	蜜饯/果干

图表 2.2-2 分类数据示例

2.3.数据量

销售数据 (附件 1): 70680 条数据。

分类数据 (附件 2): 315 条数据。

三、数据预处理与分析

3.1.数据预处理

3.1.1.合并数据

销售和分类数据合并, 即为每一个商品写入对应的大类以及二级类。

3.1.2.清洗数据

1) 格式内容清洗:

观察数据, 发现最后一行数据时间存在异常, 决定删去。

2) 检测和处理重复值:

对“订单号”进行查重检查, 发现并无重复订。

3) 检测和处理缺失值：

发现并无任何缺失值。

4) 删除多余的字段：

计算“应付金额”与“实际金额”的相似度，两列数据完全相同，因此选择删除“应付金额”列以提高效率；统计“状态”列和“提现”列数据，发现这两列都没有提供有效信息，故决定删去这两列。

5) 检测和处理异常值：

观察“实际金额”列数据，发现有部分数据存在异常，因此本文使用箱线图找出异常值及其索引，并用对应商品“实际金额”的中位数去替换异常值。再检查异常值检查遗漏的实际金额为 0 的数据，并删除。

3.1.3.转换数据

1) 增加特征列并设置时间格式：

把“支付时间”列转换为时间格式。为了方便后续分析，还增加了“Month”、“Day”、“Hour”列。为方便计数，为每一订单数据增加“销量”为 1

2) 提取每台售货机对应的销售数据：

提取每台售货机对应的销售数据，分别保存在“result”文件内的“task1-1A.csv”、“task1-1B.csv”、“task1-1C.csv”、“task1-1D.csv”、“task1-1E.csv”中。

3.2.数据简易分析

3.2.1.售货机五月份交易情况

计算每台售货机 2017 年 5 月份的交易额、订单量及 5 台售货机交易总额和订单总量，计算结果如下表：

地点	交易额	订单量
A	3067.1	753
B	3496.5	866
C	3357.3	785
D	2151.2	561
E	5329.8	1292
合计	17401.9	4257

图表 3.2-1 各台售货机 2017 年 5 月交易额及订单量统计表

从 图表 3.2-1 我们可以看出：2017 年 5 月份 A、B、C 售货机的销售情况差不多，它们交易额均是三千多，订单量也都八百左右。E 售货机的交易额和订单量明显高于其它售货机；而 D 售货机的销售数据则不太理想，相比于其它售货机它的交易额和订单量明显较少。

3.2.2.售货机各月统计情况

计算每台售货机 2017 年每月的每单平均交易额与日均订单量，每台售货机每月的每单平均交易额如下表：

地点 月份	A	B	C	D	E
1 月	3.99	3.65	3.98	3.60	4.05
2 月	3.79	3.29	3.84	3.17	3.75
3 月	3.45	3.60	3.59	3.94	4.22
4 月	3.86	3.73	4.02	3.58	3.93
5 月	4.07	4.04	4.28	3.83	4.13
6 月	3.75	3.91	4.03	3.79	3.70
7 月	3.79	4.09	3.85	3.85	3.84
8 月	3.32	3.58	3.91	3.31	3.75
9 月	4.01	3.96	4.15	3.68	3.93
10 月	3.78	3.88	3.97	3.62	3.69
11 月	4.13	4.00	4.12	3.69	4.07
12 月	3.63	3.64	3.82	3.49	4.02

图表 3.2-2 各台售货机 2017 年每月每单平均交易额

从 图表 3.2-2 我们可以看出：每台售货机每月每单的平均交易额相差不大，每台售货机每月每单平均交易额保持在 [3, 5] 元之间。由此可以初步断定，一年中五个地点的售货机的顾客所购买的商品的平均价格水平较为集中。

每台售货机每月的日均订单量如下表：

地点 月份	A	B	C	D	E
1 月	10.81	11.81	12.16	8.29	11.42
2 月	4.07	6.61	7.39	5.04	9.21
3 月	8.23	8.55	8.48	6.19	11.29
4 月	14.87	20.10	24.47	14.77	29.83
5 月	24.29	27.94	25.32	18.10	41.68
6 月	55.50	61.80	62.67	34.60	86.27
7 月	15.35	11.13	24.61	10.23	26.19
8 月	21.45	31.65	40.61	23.06	57.00
9 月	34.63	58.10	55.83	32.77	137.63
10 月	50.45	65.35	71.39	38.23	89.58
11 月	38.63	67.60	64.70	40.33	167.30
12 月	64.58	71.29	76.61	53.65	104.90

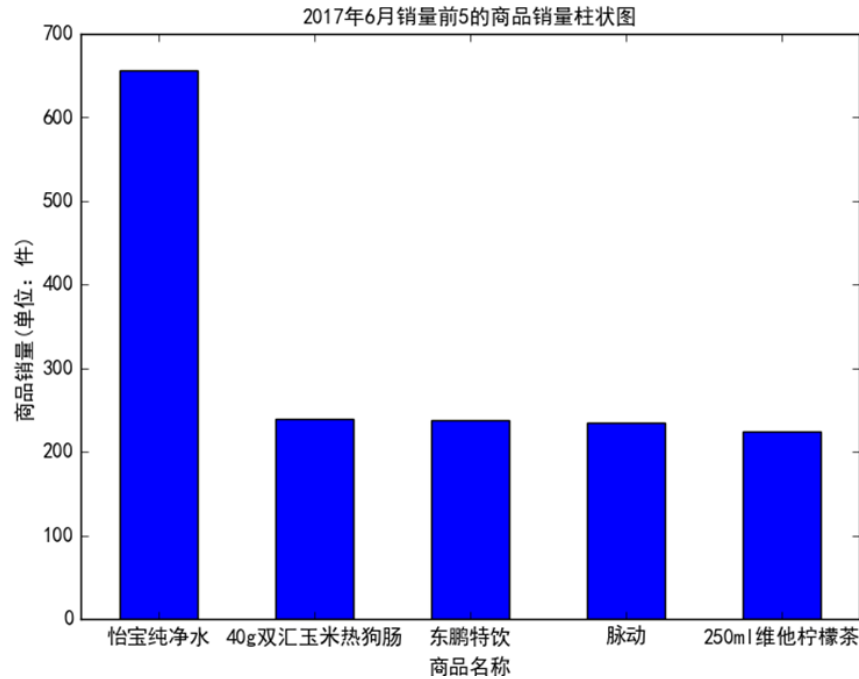
图表 3.2-3 各台售货机 2017 年每月的日均订单量

从 图表 3.2-3 我们可以看出：5 台售货机在 6 月、9~12 月的日均订单量偏高，而 5 台售货机在 2 月份的日均订单量都是最低的。A、B、C、D 售货机在 12 月份的订单量是最高的，而 E 售货机在 11 月份的订单量是最高的。

四、数据分析与可视化

4.1.六月售货机销量前 5 商品

绘制 2017 年 6 月销量前 5 的商品销量柱状图：



图表 4.1-1 2017 年 6 月销售前 5 的商品销量柱状图

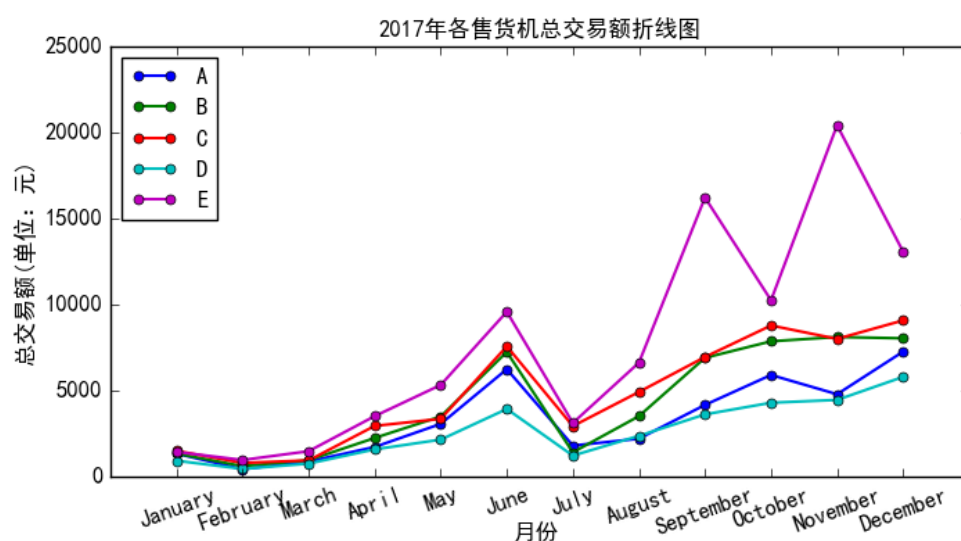
商品	销量	大类	二级类
怡宝纯净水	657	饮料	水
40g 双汇玉米热狗肠	240	非饮料	肉干/豆制品/蛋
东鹏特饮	238	饮料	功能饮料
脉动	235	饮料	功能饮料
250ml 维他柠檬茶	225	饮料	茶饮料

图表 4.1-2 2017 年 6 月销量前 5 的商品销售量

从 图表 4.1-1 和 图表 4.1-2 可以很清晰看到，怡宝纯净水的销量遥遥领先，其销量超过其它产品一倍以上。可能由于六月份天气炎热，整个六月 top5 中，饮料类的商品占了大部分，且水和功能饮料都较为热销。因此，自动售货机运营者可以在六月增加“怡宝纯净水”和功能饮料的供给量。

4.2.交易额变化分析

每台售货机每月总交易额折线图：



图表 4.2-1 2017 年各售货机总交易额折线图

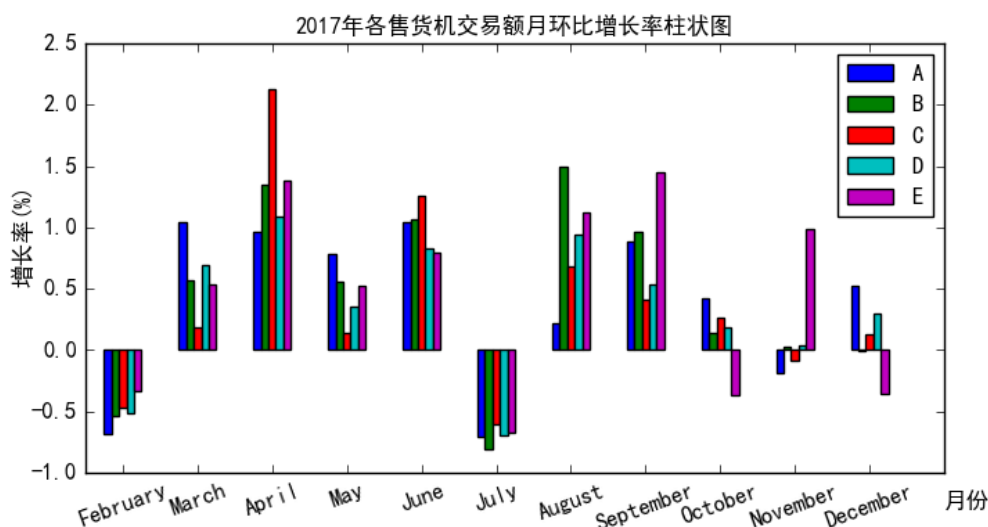
从 图表 4.2-1 可以看出：若从整体交易额来看，E 售货机每月的总交易额较高，D 售货机的每月总交易额较低。

若从折线的趋势来看，5 台的售货机的交易额都是从 2 月上升到 6 月，在 6 月达到一个极大值，之后整体下降至 7 月，然后整体上升。

值得注意的是，E 售货机在 9 月和 11 月都出现了比较突出的极大值，总交易额突然大增。

因此，对于 5 台的售货机，运营者在 6 月、9 到 12 月都要适当地增加商品供给量和供给频率。而对于 E 售货机，在 9 月和 11 月则要更大力地增加商品供给量和供给频率。

每台售货机每月交易额月环比增长率柱状图：



图表 4.2-2 2017 年各售货机交易额月环比增长率柱状图

交易额月环比增长率=(本月交易额-上月交易额)/上月交易额×100%，所反映的是连续两个月的交易额变化情况。

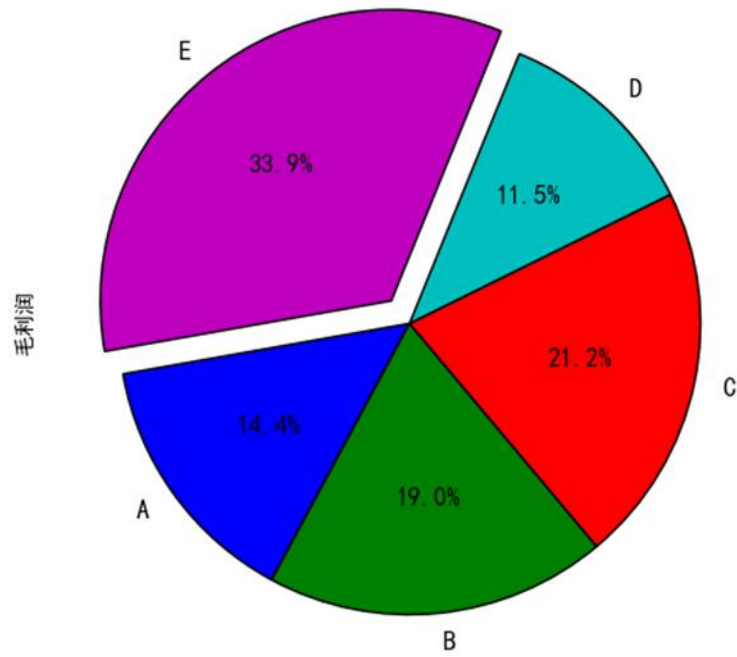
从 图表 4.2-2 可以看出：在 2 月和 7 月，5 台售货机交易额均出现了非常明显的大幅度下降。初步估计，有可能是 2、7 两个月正值假期，收到假期人流量影响。而售货机位置是固定的，虽然我们没有较好的提高人流量的解决办法，但是我们可以减少投放保质期短的商品，防止产品滞销导致出现货物损失。

4.3.毛利润分析

绘制 2017 年各台售货机毛利润占总毛利润比例的饼图，在这里假设饮料类毛利率为 25%，非饮料类为 20%。

从 图表 4.3-1 **错误!未找到引用源。** 我们可以看出：E 售货机的毛利润极高；而 D 售货机的毛利润最少，只有 D 售货机的三分之一。因此，建议运营者增加在 E 售货机的商品投放量，减少在 D 售货机的商品投放量，尤其是在九月到十二月份的时候。

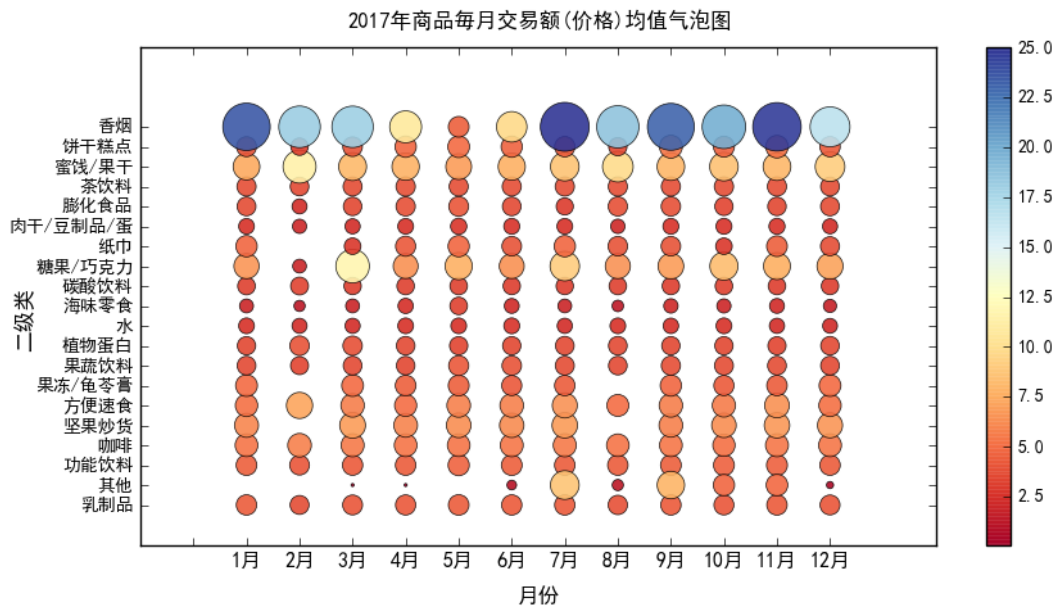
2017年各台售货机毛利润占总毛利润比例的饼图



图表 4.3-1 2017 年各售货机毛利润占总毛利润比例的饼图

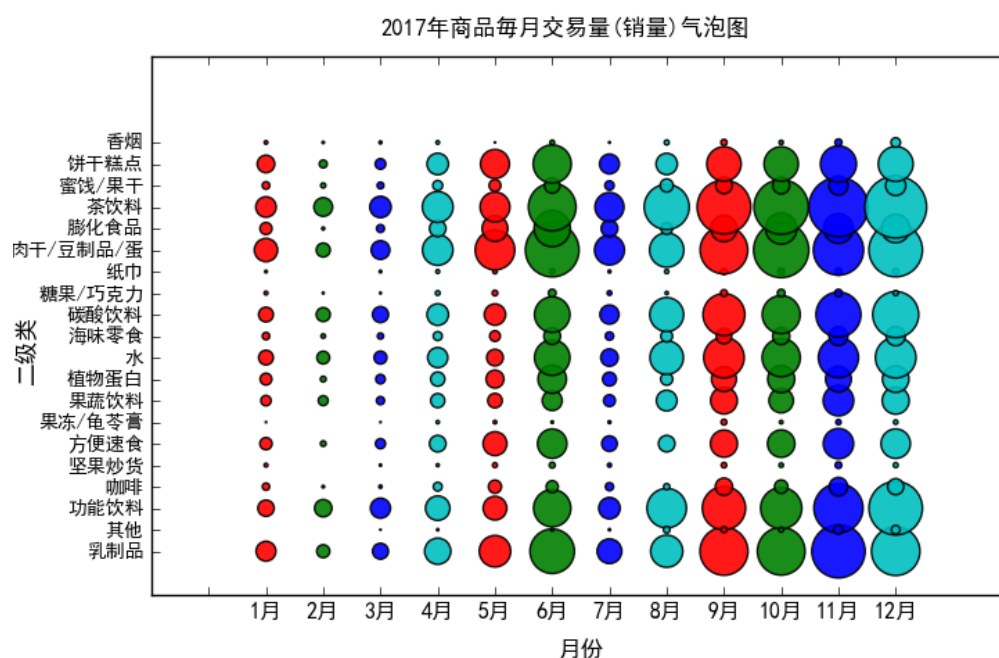
4.4.商品在不同月份出售情况分析

绘制每月交易额(价格)均值气泡图，横轴为时间，纵轴为商品的二级类目：



图表 4.4-1 2017 年商品每月交易额(价格)均值气泡图

绘制每月交易量(销量)气泡图，横轴为时间，纵轴为商品的二级类目：



图表 4.4-2 2017 年商品每月交易量(销量)均值气泡图

从 图表 4.4-1 2017 年商品每月交易额(价格)均值气泡图我们可以看出：每个月“香烟”类的气泡图基本是最大的，这也说明“香烟”类的价格都是较高的，相应也会利润较高。其次是“糖果/巧克力”、“蜜饯/干果”类的气泡图也较大，价格也较高。

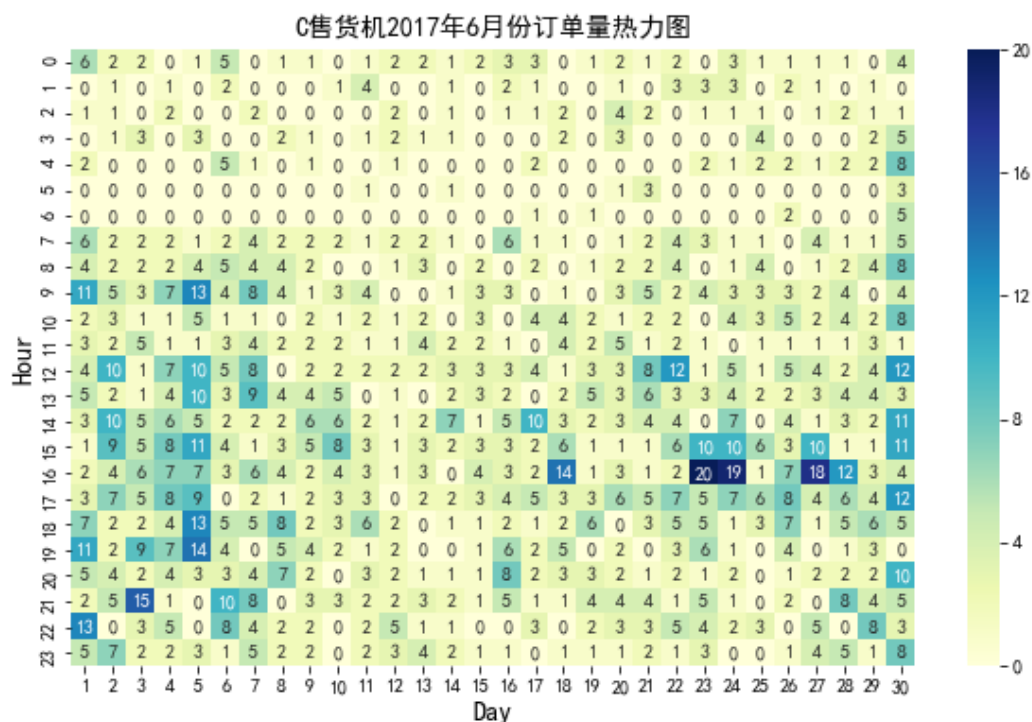
而在 2 月份时“纸巾”、“果冻/龟苓膏”、“坚果炒货”类没有销售记录，8 月份时“坚果炒货”、“果冻/龟苓膏”类也没有销售记录。

再观察 图表 4.4-2 的销售情况，并结合 4.2 交易额变化分析：可知 2 月份的总销量是 2017 年中最少的，最有可能是人流过少，导致 2 月份有三类商品没有销售记录。但是 8 月份的总体销量是不差的，需要调查回顾我们在 17 年的 8 月是否在售货机中投放了“坚果炒货”、“果冻/龟苓膏”类商品，如果没有投放，那我建议保持一定量的对上述产品的投放。

4.5.六、七、八 月订单随时间分析

绘制 C 售货机 6、7、8 三个月订单量的热力图，横轴以天为单位，纵轴以小时为单位。

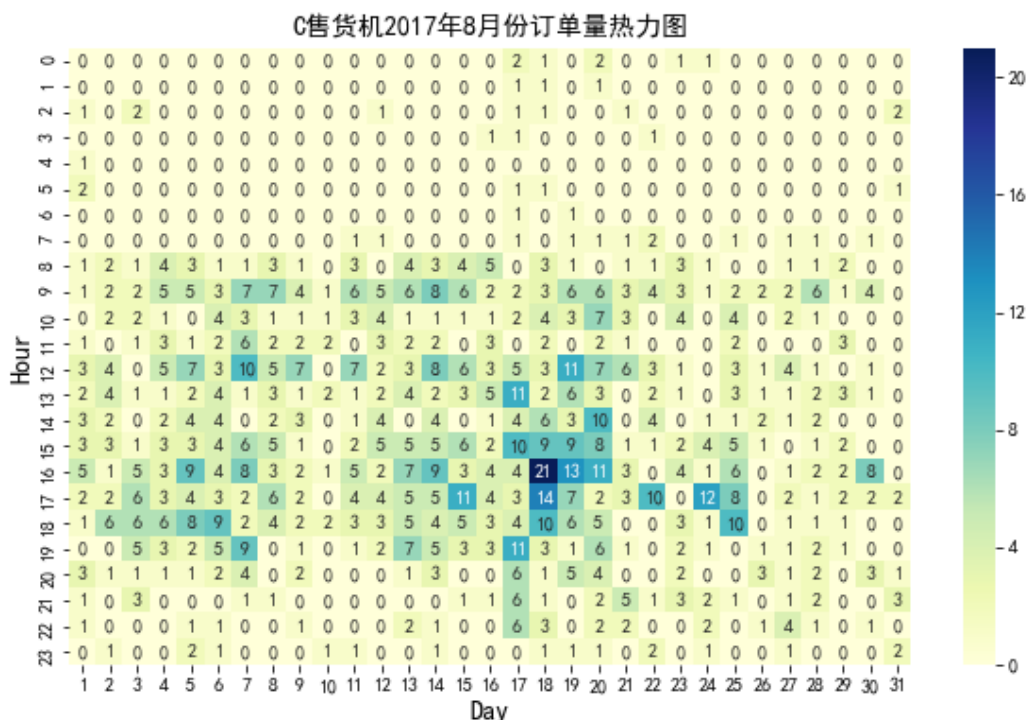
热力图反映在哪天哪个小时的订单密集程度，也就是订单量相对多或少。



图表 4.5-1 C 售货机 2017 年 6 月份订单量热力图



图表 4.5-2 C 售货机 2017 年 7 月份订单量热力图



图表 4.5-3 C 售货机 2017 年 8 月份订单量热力图

从上面三个热力图可以看出：

- 1) 早上八点到中午十二点以及晚上七点到晚上十点这段时间中，订单量稳定但并不是很高；从中午十二点到晚上七点，商品订单量极高，而订单量最高一般出现在下午 4 点的时候。因此，自动售货机运营者的供货时间可以定在早上。
- 2) 七月中旬，C 售货机出现了订单量极低的现象，我们需要调查回顾找原因，减少因出现机器故障等原因而降低了订单销售量。

五、自动售货机画像

5.1. 饮料类商品—销售热度标签

根据帕累托法则得出的“少数中的多数”，即 20% 的商品销售量（毛利润）可实现全部销售量（毛利润）的 80% 左右，而剩下的 80% 商品的销售量（毛利润）则只实现总销售额的 20% 左右。因此，这里对于“热销/正常/滞销”的饮料类商品

有如下标记规则：

热销商品：同时满足单品销量排名前 20%（包含 20%）和单品毛利润排名前 20%（包含 20%）。

滞销商品：同时满足单品销量排名后 20%（包含 20%）和单品毛利润排名后 20%（包含 20%）。

正常商品：除了热销商品和滞销商品外，其余的都是正常商品。

2017 年各售货机饮料类商品销量和毛利润数据描述如 图表 5.1-1：

	A	B	C	D	E
销量 20%	4.2	4	4.6	2	9
销量 80%	77.2	116.8	127	81	204.4
毛利润 20%	4.6	3.7	5.575	2	10.05
毛利润 80%	86.575	123.075	130.65	73.375	213.175

图表 5.1-1 2017 年各售货机饮料类商品销量和毛利润数据

最后其销售热度标签结果见附件“task3_1A.csv”、“task3_1B.csv”、“task3_1C.csv”、“task3_1D.csv”、“task3_1E.csv”。其中一个销售热度标签结果示例如 图表 5.1-2：

序号	饮料类商品	销售热度
1	100g*5瓶益力多	正常
2	13g雀巢咖啡1+2特浓	滞销
3	145ml旺仔牛奶盒装	热销
4	145ml旺仔牛奶罐装	滞销
5	150g健能酸奶原味	正常

图表 5.1-2 销售热度标签结果示例图

5.2. 饮料类商品—消费档次标签

同 5.1 饮料类商品—销售热度标签，运用“二八原则”，这里对于消费档次“高档/中档/低档”的饮料类商品有如下标记规则：

高档商品：单品平局价格排名前 20%（包含 20%）。

低档商品：单品平局价格排名后 20%（包含 20%）。

中档商品：单品平局价格排名在 20%（不含 20%）到 80%（不含 80%）之间。

2017 年各售货机饮料类商品平均价格数据描述如 图表 5.2-1：

	A	B	C	D	E
平均价格 20%	3.05	3.02	3.18	3.00	3.20
平均价格 80%	4.89	5.00	5.00	4.62	4.97

图表 5.2-1 2017 年各售货机饮料类商品平均价格数据描述

最后其销售热度标签结果见附件“task3_2A.csv”、“task3_2B.csv”、“task3_2C.csv”、“task3_2D.csv”、“task3_2E.csv”。其中一个消费档次标签结果示例如 图表 5.2-2：

序号	饮料类商品	销售热度	消费档次
1	100g*5瓶益力多	正常	高档
2	13g雀巢咖啡1+2特浓	滞销	中档
3	145ml旺仔牛奶盒装	热销	低档
4	145ml旺仔牛奶罐装	滞销	低档
5	150g健能酸奶原味	正常	中档

图表 5.2-2 消费档次标签结果示例图

5.3.自动售货机销量画像

对 5 台售货机的饮料类商品销量分别进行统计，绘制各售货机的商品销量词云图：



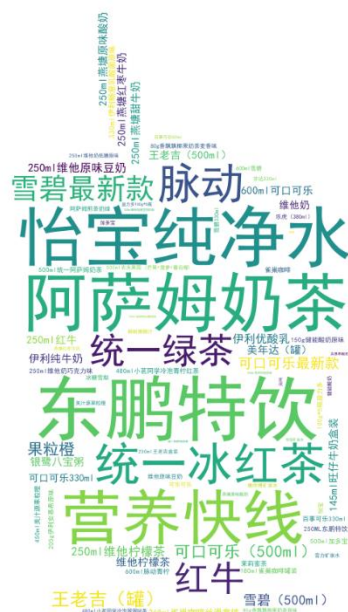
图表 5.3-1 A 售货机商品销量词云图



图表 5.3-2 B 售货机商品销量词云图



图表 5.3-5 D 售货机商品销量词云图



图表 5.3-4 D 售货机商品销量词云图



图表 5.3-3 C 售货机商品销量词云图

从 3.2 数据简易分析 我们可以知道 E 售货机的交易额和订单量明显高于其它售货机；而 D 售货机相比于其它售货机它的交易额和订单量明显较少。所以在这里特别列出是 E、D 售货机饮料类销量 top10 商品数据描述：

商品	销量	平均价格	总毛利润	销售热度	消费档次
东鹏特饮	411	3.52	362.12	热销	中档
怡宝纯净水	387	2.01	194.38	热销	低档
阿萨姆奶茶	365	4.00	365.38	热销	中档
营养快线	319	4.03	321.38	热销	中档
统一冰红茶	309	3.01	232.62	热销	中档
脉动	283	4.07	288.12	热销	中档
统一绿茶	188	3.04	142.75	热销	中档
红牛	180	5.98	269.25	热销	高档
雪碧最新款	178	3.09	137.62	热销	中档
王老吉（罐）	177	3.52	155.88	热销	中档

图表 5.3-6 E 售货机饮料类销量 top10 商品数据描述

商品	销量	平均价格	总毛利润	销售热度	消费档次
怡宝纯净水	1705	2.71	1154.88	热销	低档
脉动	996	4.53	1127.62	热销	中档
营养快线	805	4.46	896.62	热销	中档
阿萨姆奶茶	792	4.40	871.38	热销	中档
东鹏特饮	709	3.87	685.88	热销	中档
统一冰红茶	626	3.47	543.62	热销	中档
统一绿茶	490	3.50	429.00	热销	中档
果粒橙	470	3.71	435.75	热销	中档
王老吉（500ml）	417	4.51	470.00	热销	中档
王老吉（罐）	410	3.85	394.75	热销	中档

图表 5.3-7 D 售货机饮料类销量 top10 商品数据描述

从 图表 5.3-6 和 图表 5.3-7 这个两个表格我们可以看出：

- 1) 不同售货机，其热销的产品都是差不多的。D 售货机和 E 售货机饮料类销量 top10 商品都含有“怡宝纯净水”、“脉动”、“营养快线”、“阿萨姆奶茶”、“东鹏特饮”、“统一冰红茶”、“统一绿茶”、“王老吉（罐）”，饮料类 top10 商品中，有 8 种商品是一样的。并且从五个词云图也可以看出，比较大比较突出的字

体，都有“怡宝纯净水”、“脉动”、“营养快线”、“阿萨姆奶茶”、“东鹏特饮”、“统一冰红茶”、“统一绿茶”、“王老吉（罐）”等。这说明大家在不同的地方，喜欢的饮料很大可能是差不多的。

- 2) **销售量**大时，可适当上提价格而在价格方面。也会发现 top10 商品中，E 售货机的平均价格要比 D 售货机的平均价格要高，这里猜测 E 售货机附近人流比较多，相对应也稍微提高了价格，以获得更大的利润。

5.4.营销意见

- 1) 增加畅销商品，并优先上架畅销商品，要保证畅销商品供货的稳定足量，保证畅销商品在所有门店和各个时间都不断档缺货。
- 2) 撤除滞销商品，这可以帮助减少不良库存，从而减少降价商品的比例，品类的减少伴随着人工成本的降低，也有助于整体利润的提升。
- 3) 同时若我们已买进大量滞销商品，我们可以采取捆绑销售、优惠促销等方式去减少库存。
- 4) 积极引入新商品，在有限的卖场空间内推陈出新，追求评效的最大化。
- 5) 适当提高畅销品的价格，尤其在销售情况比较好的 6 月份、9 到 12 月份，这时候的商品销量会比较大，趁着人流多（旺季），考虑价格与销量之间的权重，提高利润。
- 6) 适当提高“怡宝纯净水”的价格，尤其在 6 月份、9 到 12 月份，可以将其价格从 2 元，提升到 2.5 元甚至更高，毕竟在各个售货机中，“怡宝”的销量还是很可观的。

六、业务预测

6.1.预测原理

销售预测是指根据以往的销售情况以及使用系统内部内置或用户自定义的销售预测模型获得的对未来销售情况的预测。销售预测可以直接生成同类型的销售计划。

6.2.预测方法

6.2.1.定性预测

一般来说，在销售预测中常用的定性预测方法有四种：高级经理意见法、销售人员意见法、购买者期望法和德尔菲法。

1) 高级经理意见法

高级经理意见法是依据销售经理（经营者与销售管理者为中心）或其他高级经理的经验与直觉，通过一个人或所有参与者的平均意见求出销售预测值的方法。

2) 销售人员意见法

销售人员意见法是利用销售人员对未来销售进行预测。有时是由每个销售人员单独作出这些预测，有时则与销售经理共同讨论而作出这些预测。预测结果以地区或行政区划汇总，一级一级汇总，最后得出企业的销售预测结果。

3) 购买者期望法

许多企业经常关注新顾客、老顾客和潜在顾客未来的购买意向情况，如果存在少数重要的顾客占据企业大部分销售量这种情况，那么购买者期望法是很实用的。

这种预测方法是通过征询顾客或客户的潜在需求或未来购买商品计划的情况，了解顾客购买商品的活动、变化及特征等，然后在收集消费者意见的基础上分析市场变化，预测未来市场需求。

4) 德尔菲法

德尔菲法又称专家意见法，是指以不记名方式根据专家意见作出销售预测的方法。至于谁是专家，则由企业来确定，如果对专家有一致的认同则是最好不过的。德尔菲法通常包括召开一组专家参加的会议。第一阶段得到的结果总结出来可作为第二阶段预测的基础，通过组中所有专家的判断、观察和期望来进行评价，最后得到共享具有更少偏差的预测结果。

德尔菲法的最大优点是充分民主地收集专家意见，把握市场的特征。但是，德尔菲法一般只能得到企业或行业的预测结果，用此方法所求得的地区、顾客、产品分类等预测结果就没有那么精确了。

6.2.2.定量预测

用来进行销售预测的定量预测方法可以按照不同类型分成两大类：时间序列分析法、回归和相关分析法。

1) 时间序列分析法

时间序列分析法是利用变量与时间存在的相关关系，通过对以前数据的分析来预测将来的数据。在分析销售收入时，大家都懂得将销售收入按照年或月的次序排列下来，以观察其变化趋势。时间序列分析法现已成为销售预测中具有代表性的方法。

2) 回归分析法

各种事物彼此之间都存在直接或间接的因果关系。同样的，销售量亦会随着某种变量的变化而变化。当销售与时间之外的其他事物存在相关性时，就可运用回归和相关分析法进行销售预测。

6.3.可预测性判断及解释

我们需要判断是否可以根据附件提供的数据（仅有 2017 年的数据）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，

因为本次预测是基于已有的 2017 年 5 个售货机的销售数据之上的，这里就不考虑用定性预测的方法，而至考虑定量预测方法：

对于时间序列分析法，时间观察期数对模型影响是非常大的，本次描述性数据分析可以看出，一年中不同月不同商品品类的交易额是存在比较大差异，单从 2017 年来看月与月之间的规律不大，相关度不高。因此如果想要更准确的预测个大类商品在 2018 年 1 月的交易额，是很难的。

对于回归分析法，每个大类商品自变量除了时间外，就只有地点是可作为自变量，不同的售货机销量确实会随着售货机地点的不同而有差异，我们可以选择“地点”也作为其中一个变量。但是由于本次预测也是基于“时间”上面去分析的，所需数据样本还是过少，数据包含的时间段太短。不足以支持预测工作。

6.4.预测数据需求

自动售货机运营者可以提供 5 年以上的销售数据（若只预测 1 月份的数据，可仅提供 5 年间 1 月份的数据），当然，数据越多越好。如果观察的数据时间长

度不够，就很难观察出数据的规律。如果自动售货机运营者能够提供以上数据，这样就可以很好的利用时间序列粉刺法来预测未来的销售额。

根据实际情况，影响自动售货机的交易额的因素还有天气、温度、购买者的年龄、其他商店对商品的定价、当地居民的消费水平等，若自动售货机运营者能提供以上数据，就能以以上因素作为自变量建立回归模型来预测未来的销售额。

七、小结

售货机	总销量	总交易额	销量表现	主要热销产品	热销月份
A	10473	39823.7	销量较好	怡宝纯净水	6 月
				东鹏特饮	10 月
				阿萨姆奶茶	12 月
B	13472	51736.7	销量较好	怡宝纯净水	10 月
				东鹏特饮	11q 月
				阿萨姆奶茶	12 月
C	14472	57797.7	销量较好	怡宝纯净水	10 月
				脉动	11 月
				东鹏特饮	12 月
D	8705	31553.2	销量较低	东鹏特饮	10 月
				怡宝纯净水	11 月
				阿萨姆奶茶	12 月
E	23493	92013.4	销量最好	怡宝纯净水	9 月
				脉动	11 月
				营养快线	12 月

图表 2017 年各售货机对比小结