

新零售-无人智能售货机商务数据分析报告

目录

1.背景.....	2
2.数据说明	2
3.任务目标	2
4.数据预处理	3
5.统计分析	3
6.数据分析与可视化.....	6
7.自动售货机画像.....	11
8.售货机销售特点分析及营销建议.....	16
9.业务预测	17

1. 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。

本文将根据将对某商场不同地点的 5 台自动售货机 2017 年的商品销售数据进行分析，从而挖掘出有用信息。

2. 数据说明

2.1 销售数据

该部分数据存储在“附件 1”中，包含了编号为 A,B,C,D,E 五台售货机再 2017 年全年的销售数据，共 70680 条。其中每条数据中包含每单交易的订单号、设备 ID、应付金额、实际金额、商品、支付时间、地点、状态、提现几个属性的相关信息。

2.2 商品数据

该部分数据存储在“附件 2”中，包含了售货机所销售的全部商品的全部信息(共 315 条)，并对这些商品进行了大类、二级类的类别划分。其中，大类将商品划分为了饮料类和非饮料类，二级类将商品细分为乳制品、饼干糕点等 20 个小类。

3. 任务目标

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。

4. 数据预处理

4.1 去除无关属性

通过分析，“订单号”、“设备 ID”对于分析无太大帮助，可去除；另外，对于“应付金额”和“实际金额”两列数值相同，因此可以去除其中一列减少数据冗余；对于“状态”属性其所有取值均为“已出货未退款”，“提现”属性的取值均为“已提现”，所以这两个属性也无太大意义，可以去除。

4.2 处理异常时间数据

通过分析再交易数据中存在一条支付时间异常的数据，附件 1 中存在一条交易时间为“2017/2/29”的交易信息，但 2017 年没有 29 日；所以需要将其应该为“2017/3/1”的交易数据。

4.3 销售数据中填入商品标签

由于之后的分析过程中需要根据商品的类别对销售数据进行分析，为了方便之后的分析，所以需要将“附件 2”中的商品类别信息与“附件 1”中的销售数据合并并命名为“销售数据总表(带标签).csv”

4.4 提取每台售货机的销售信息

由于之后需要对每台售货机的销售数据进行分析所以需要每台售货机的交易数据从总表中抽取出来，分别命名为“task1-1A.csv”、“task1-1B.csv”，…，“task1-1E.csv”进行存储。处理部分结果如下图所示：

应付金额	商品	支付时间	地点	大类	二级类
3.0	40g双汇玉米热狗肠	2017-01-01 01:33:00	A	非饮料	肉干/豆制品/蛋
5.8	100g卫龙点心面黑椒牛排味	2017-01-01 19:29:00	A	非饮料	肉干/豆制品/蛋
0.8	咪咪虾条马来西亚风味	2017-01-01 19:33:00	A	非饮料	膨化食品
3.0	250ml维他柠檬茶	2017-01-01 19:55:00	A	饮料	茶饮料
3.5	东鹏特饮	2017-01-01 20:18:00	A	饮料	功能饮料

图 4-1 预处理结果展示图

5. 统计分析

本部分对各个售货机的交易数据的交易额，订单量等信息进行一些基本的汇总分析，目的是先通过简要的分析，分析差别。

5.1 五月份交易数据汇总

本部分对每台售货机 2017 年 5 月份的交易额、订单量及所有售货机交易总额和订单总量进行了汇总，相关结果如下：

表 5-1 各售货机 5 月份交易数据汇总

售货机 指标	A	B	C	D	E	合计
交易额	3385.1	3681.2	3729.4	2392.1	5699.0	18886.8
订单量	756	869	789	564	1292	4270

通过上表数据可以看出 A, B, C 三个地区的售货机在五月份的交易总量相差不多，但 D 地区的交易额及订单量明显低于其它地区，E 地区的售货机交易额与订单量明显高于其它地区。造成这一结果的原因可能是五个地区的位置及人流因素的影响。因此后续的分析要着重注意 D 和 E 两个区域的售货机与其它售货机的差异，从而得到更多有价值的信息。

5.2 各售货机每月每单销售数据汇总

根据每台售货机的销售数据，分别计算了每月的每单平均交易额与日均订单量，具体结果如下：

表 5-2 各售货机每月每单平均交易额

售货机 月份	A	B	C	D	E
1 月	4.51	3.75	4.33	3.69	4.68
2 月	3.86	3.26	3.83	3.09	3.64
3 月	3.59	3.61	3.77	4.31	4.31
4 月	4.04	4.08	4.4	3.79	4.16
5 月	4.48	4.24	4.73	4.24	4.41
6 月	4.05	4.07	4.5	4.03	3.82
7 月	4.1	4.4	3.99	4.23	3.92
8 月	3.36	3.58	3.91	3.32	3.8
9 月	4.31	4.13	4.43	3.9	4.13

10 月	4.02	4.11	4.27	3.88	3.68
11 月	4.47	4.27	4.35	3.86	4.28
12 月	3.79	4.67	3.94	3.57	4.17

从表 5-2 中可以看出，再 12 个月份中每个月，每个售货机每单交易额都相差不多，这表明各个售货机所售出的商品基本都处于相同的价格水平，且每单平均交易额都在 3-4 元，这说明售货机所售商品均为较廉价的商品，或者是顾客倾向于在售货机购买一些物美价廉的商品。

表 5-3 各售货机每月日均订单量

售货机 月份	A	B	C	D	E
1 月	10	11	12	8	11
2 月	4	6	7	5	9
3 月	8	8	8	6	11
4 月	14	20	24	14	29
5 月	24	28	25	18	41
6 月	55	61	62	34	86
7 月	15	11	24	10	26
8 月	21	31	40	23	57
9 月	34	58	55	32	137
10 月	50	65	71	38	89
11 月	38	67	64	40	167
12 月	64	71	76	53	104

从表 5-3 中可以看出 A, B, C, D, E 五个地区的售货机每月的日均订单量都随着月份的变化呈相同的趋势波动，在 1, 2, 3, 7 几个月份普遍偏低，在 6, 9, 10, 11, 12 几个月份普遍偏高，这说明时间是影响售货机销量的一个主要因素；另外，E 地区的售货机的销量普遍高于其它地区售货机的销量，这说明该商场的 E 地区的人流量应该更大，所以使得 E 地区的售货机的销量更高。

6. 数据分析与可视化

将数据通过图像的形式展现出来往往能够更能直观地发现数据中所包含的规律，本部分将通过数据可视化的手段进一步研究数据中包含的知识。

6.1 六月份商品销量前五柱状图

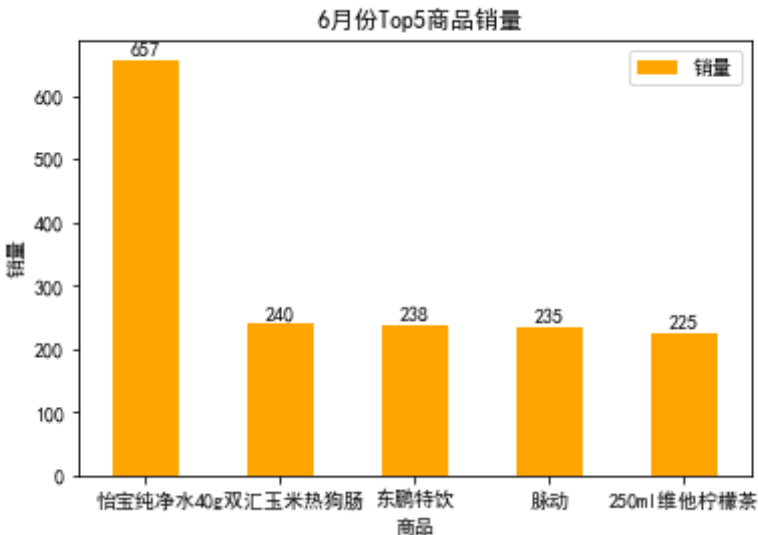


图 6-1 六月份总销量前五商品

从图 6-1 中可以看出，在六月份销量前 5 的五种商品中有四种都是饮品，这说明人们在售货机购买商品分主要是因为口渴，其次才是饥饿；通过“怡宝纯净水”这种只是用来解渴的商品的销量是其他商品销量的近三倍更能证明这一观点。

6.2 每台售货机每月总交易额折线图

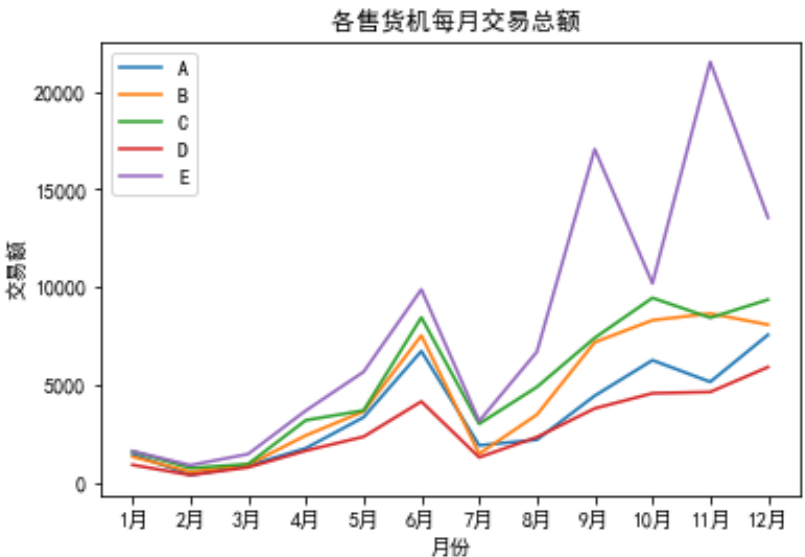


图 6-2 各售货机每月总交易额折线图

由图 6-2 的各售货机每月交易总额折线图可以看出 1, 2, 3, 7 四个月的交易额较低, 且五台售货机随着时间呈相似的趋势变化, 从 3 月份至 6 月份, 交易额逐渐增加, 但 7 月份交易额突然下降, 7 月过后销售额又开始呈上升趋势, 造成这一原因可能是 1, 2, 3 月正处于春节前后, 7 月份正处于暑期。所以, 在这几个月可以考虑适当减少商品投放量, 以及降低补货频率以降低成本。

6.3 每台售货机月环比增长率

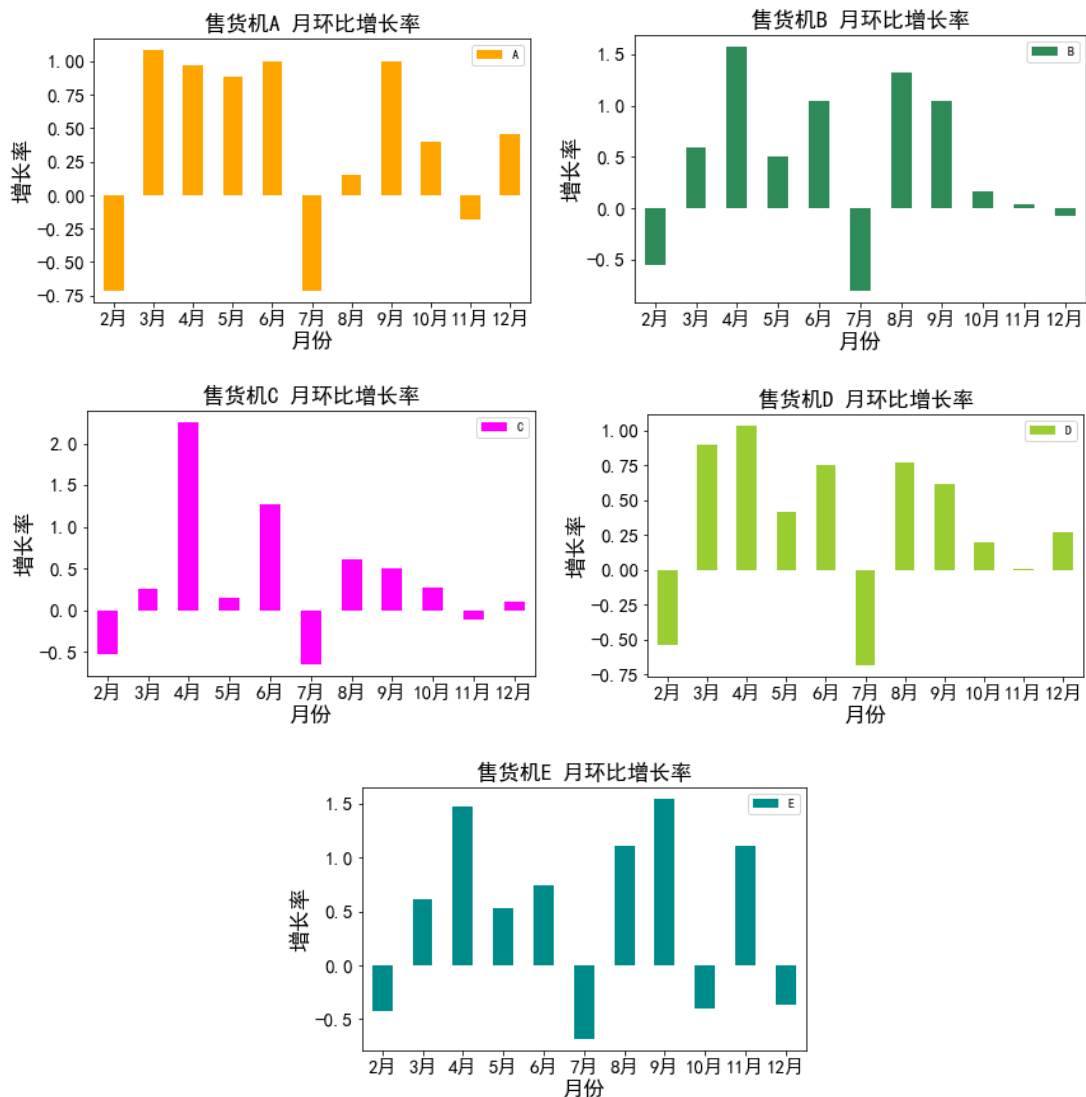


图 6-3 各售货机月环比增长率

从每台售货机的月环比增长率柱状图中可以看出, 2 月和 7 月都属于衰退期, 这就更能证明上面的猜测, 这两个月份的衰退是因为春节和暑假的影响。除去这两个月 A, B, C, D 四个地区在其他时间段不会出现大幅度下降的情况, 比较稳定; 而 E 地区的售货机波动则相对较大。

6.4 各售货机毛利润占比

假设饮料类商品的毛利润为 25%，非饮料类为 20%，为了分析各个售货机的获利情况，分别计算每台售货机总的毛利润，并绘制饼图进行分析。

表 6-1 各售货机各类商品毛利润

售货机	非饮料类毛利润	饮料类毛利润	总毛利润
A	4640.325	4796.26	9436.585
B	4743.775	6999.04	11742.815
C	5504.300	7910.98	13415.280
D	3037.425	4218.72	7256.145
E	7863.100	12840.60	20703.700

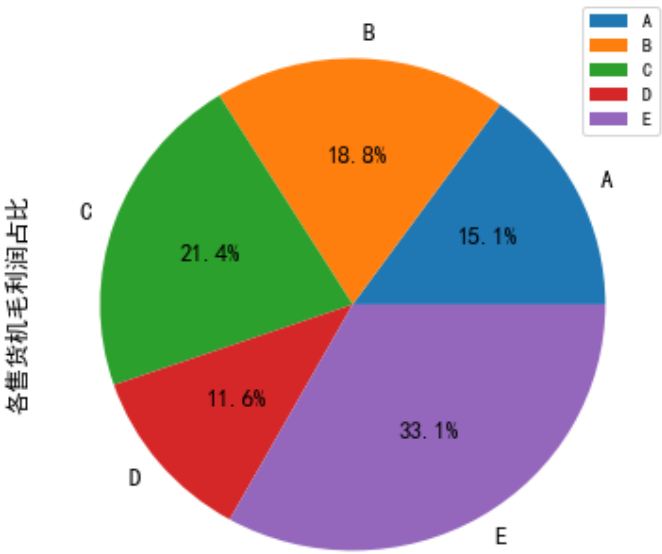


图 6-4 各售货机毛利润占比饼图

通过图 6-4 分析可知，E 地区的售货机获得的毛利润最高，占总体的 1/3，D 地区最少。在通过观察表 6-1 的数据可知饮料类商品获取的毛利润比非饮料

类更高，尤其是 E 地区饮料类毛利润占据 E 地区总毛利润的 2/3。

6.5 消费者商品类别偏好分析

了解消费者对于商品的喜好，更有利于商家根据偏好制定营销策略，从而抓住消费者的心理获取更多利润。

在“附件二”中提供了售货机所售商品的二级类别信息，根据所有售货机一年的销售数据和类别信息，即可统计出消费者的偏好信息。

气泡图能够很直观的将这些信息反映出来，下面将根据商品的二级类目绘制出每月交易额均值气泡图。

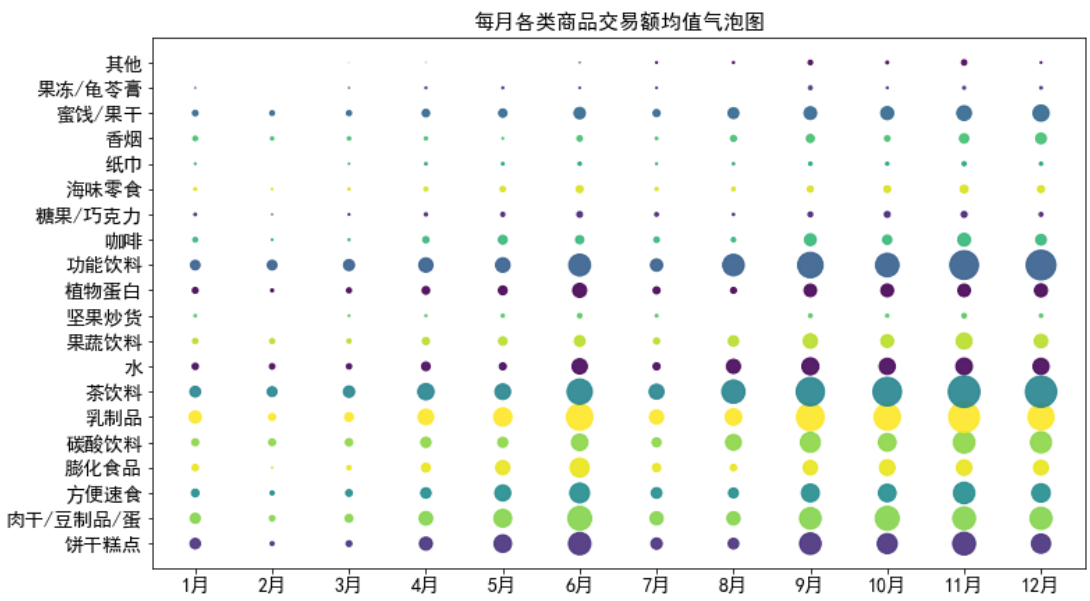


图 6-5 每月各类商品均值交易额气泡图

从图 6-5 可以看出顾客对功能饮料、茶饮料、乳制品、肉干/豆制品/蛋、饼干糕点几类商品比较欢迎，且 6，9，10，11，12 月为需求量较高的时间段，商家可以考虑在这几个月时间内下架一些冷门类别的商品，增加不同类型的热门类别商品，以增加多样性，从而促进顾客的购买欲望。另外，从总体上讲饮料类商品的需求量还是普遍高于非饮料类。

6.6 C 售货机 6、7、8 月份各时段订单热力图

通过分析各时段的订单量，可以选择合适的时间点进行补货，下面将通过绘制 C 售货机 6、7、8 月份各时段订单热力图进行分析。

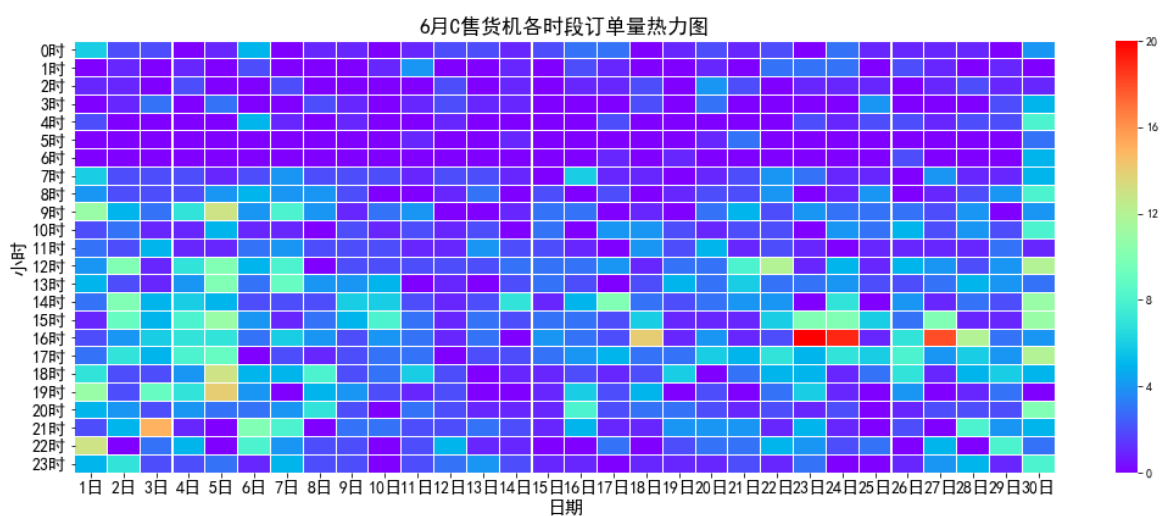


图 6-7 6 月 C 售货机各时段订单热力图

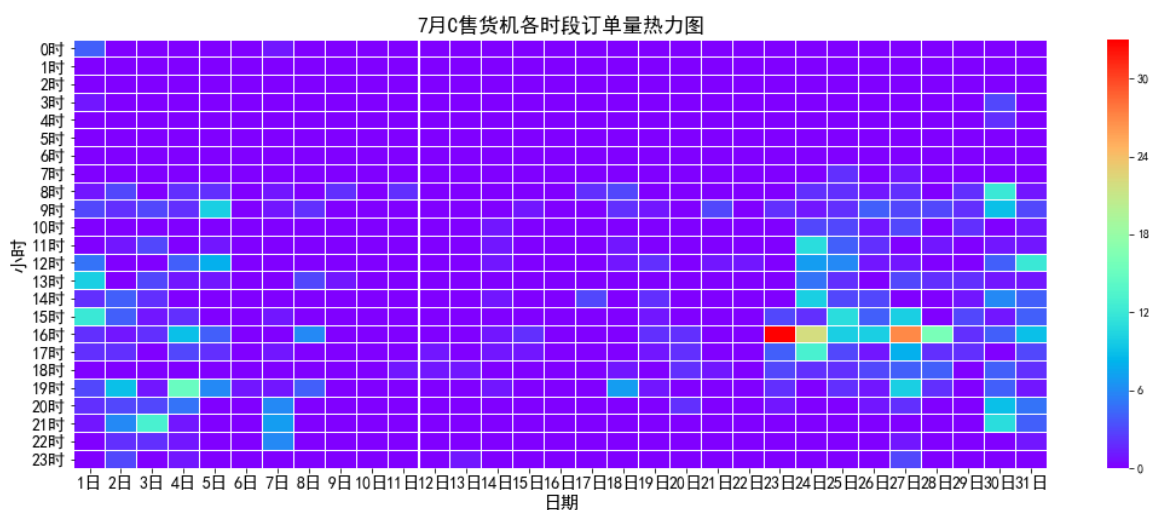


图 6-8 7 月 C 售货机各时段订单热力图

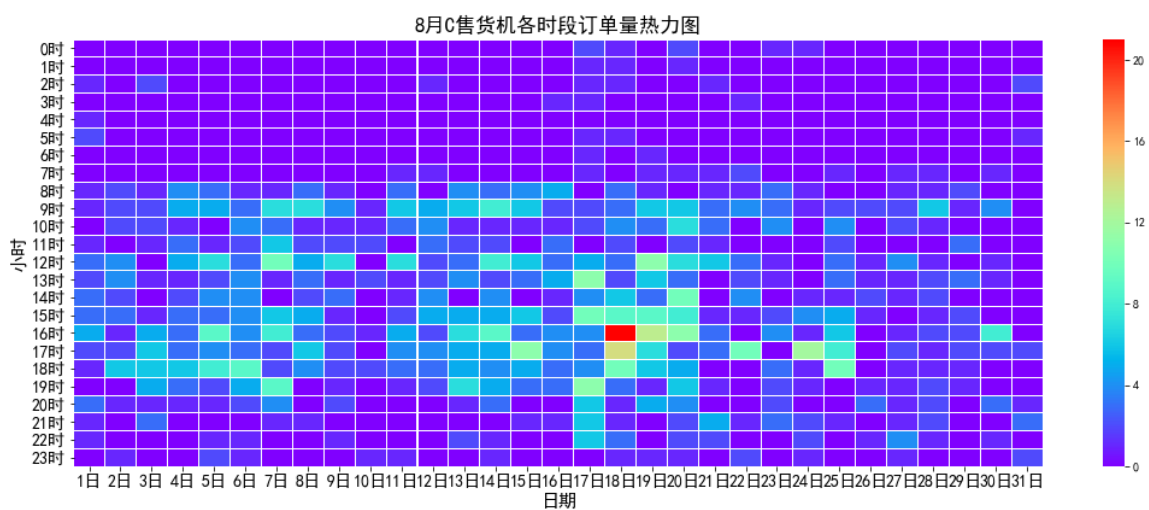


图 6-9 8 月 C 售货机各时段订单热力图

从以上三个月的订单量热力图中可以看出每天 0 时-7 时，订单量较少，其它时间段交易量较多，尤其是下午，所以可以选择每天在 7 点前进行补货。从热力图中可以看出 6 月的销售状况较好，火爆时间主要集中在前期和后期；7 月份因为暑期的原因所以销售比较惨淡，因此可以降低补货的频率，减少人力消耗；八月份交易量开始恢复，主要交易时间段为下午。

7. 自动售货机画像

画像是由一个个特征组成，是对一个对象的整体描述，通过画像我们可以对对象有一个很直观的认识。本部分主要包含两个内容，首先为各个售货机所售卖的商品添加标签，判断该商品为热销、滞销或正常；之后提取商品的一些关键特征，再根据商品的销售情况生成自动售货机画像。

7.1 商品标签化

所谓商品标签化，就是找到一个合适的划分标准，将商品划分为热销、滞销或正常。在本项工作中分别计算了总表中五台售货机各个商品平均每月每台的销量，之后又分别计算了各个商品平均每月的销量。

由于通过总表计算出的各个商品平均每月每台的销量反映了整体情况，所以根据该数据来确定划分标准是最合理的，为了更好的确定标准，根据每个商品的销量，按销量值划分出 15 个区间，并画出销量值属于每个区间的商品种类的直方图。如下图所示：

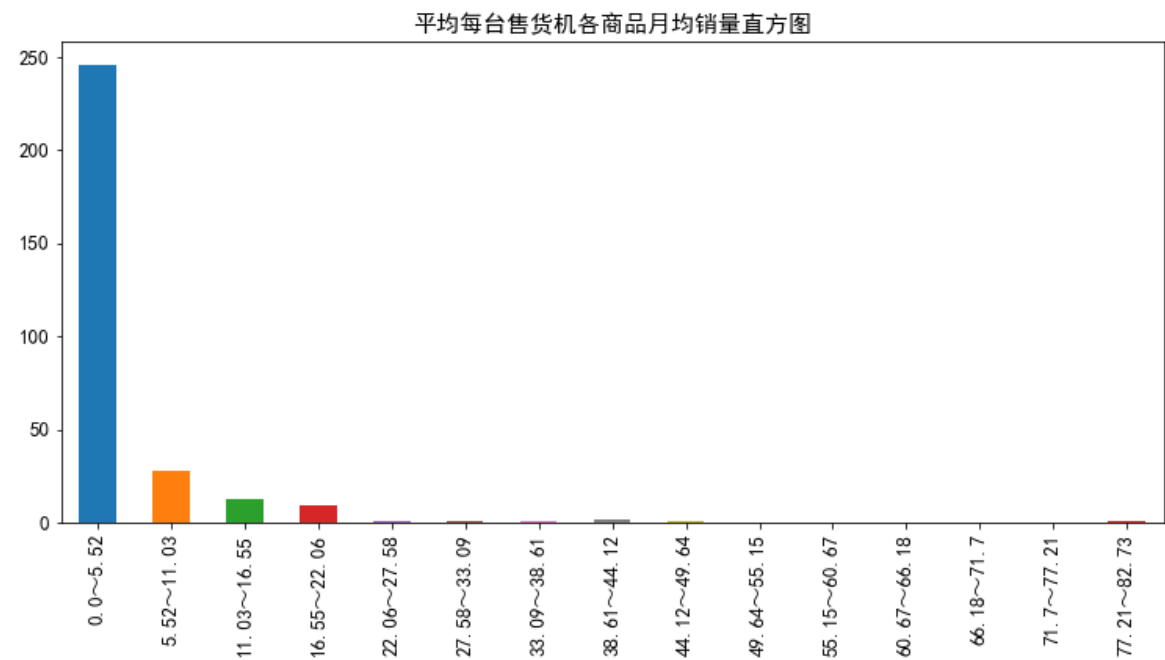


图 7-1 平均每台售货机各商品月均销量直方图

从图 7-1 中可以看出，大约 70% 的商品平均每台售货机的月均销量在 0~5.52 之间，这表明大部分的商品销售情况不太乐观，平均每月售出 5 件都很难保证，因此销量在该区间内的应认为是滞销；如果平均每台售货机每件商品月销量超过 30，即基本每天都有人购买，则可以认定该商品比较受欢迎，应该认为是热销商品，所以根据直方图将平均每台售货机的月均销量大于 27.58 的商品认定为热销；介于 5.52~27.58 之间的商品认定为正常。所以在判断每台售货机所售商品销售状况时按照以下标准进行划分：

滞销：该台售货机各商品月均销量 ≤ 5.52

正常： $5.52 <$ 该台售货机各商品月均销量 ≤ 27.58

热销：该台售货机各商品月均销量 > 27.58

最终各个售货机标签划分部分结果如下(以 E 售货机为例)：

商品	标签
怡宝纯净水	热销
脉动	热销
营养快线	热销
阿萨姆奶茶	热销
东鹏特饮	热销
统一冰红茶	热销
统一绿茶	热销

图 7-2 E 地区售货机部分商品标签

各售货机销售商品标签数据分别保存在 task3-1A.csv, task3-1B.csv, ..., task3-1E.csv 中。

7.2 绘制售货机画像

本部分是针对所有类型商品的标签和画像，因为饮料类商品销量虽然较高，但非饮料类的销量也是不容忽视的，而且针对所有类型商品的画像更能反应总体特征。

7.2.1 提取商品特征

售货机画像是根据售货机中各商品的销售情况，以及各商品的属性来构建的，通过售货机画像可以了解到顾客的品牌偏好、口味偏好等。由于提供的数据中没有商品特征相关信息所以需要进一步的将商品的品牌，口味等关键属性特征提取出来。

由于常用的分词工具，如 jieba 的词库中并不存在这些商品，所以利用分

词工具很容易丢失关键属性，所以采用分词工具提取有用特征是不合理的。不过，由商品种类有限，所以可以人为的提取各个商品的特征，这样结果较为准确。在提取商品特征时主要提取了商品名、品牌、口味、详细类别等信息，提取结果保存在“商品特征及分类信息.csv”中，部分结果展示如下：

商品	特征
100g*5瓶益力多	益力多
100g越南LIPO奶味面包干	越南面包干,奶味
10g卫龙亲嘴烧香辣味	卫龙亲嘴烧,香辣味,卫龙
10g越南LIPO奶味面包干	越南面包干,奶味
110g顺宝九制话梅	顺宝九制话梅,顺宝九制,话梅
120g达利园蔓越莓提子面包	达利园蔓越莓提子面包,面包,达利园
12g劲仔小鱼(香辣味)	劲仔小鱼,香辣味

图 7-3 商品特征

7.2.2 绘制各售货机画像

根据各售货机的商品销售情况以及商品特征，本部分为每个售货机绘制了两个画像，一个是用于描述售货机销售特征的特征画像，一个用于描述售货机所售商品的类别画像。相关画像如下图所示：



图 7-4 售货机 A 特征画像



图 7-5 售货机 A 类别画像

从售货机 A 的特征画像画像中可以看出，从品牌来讲消费者更喜欢统一、卫龙、怡宝、维他、阿萨姆、伊利、无穷农场、东鹏特饮等几个品牌的产品，从口味上来讲人们普遍喜欢原味经典口味的商品；从商品偏好来看，人们对于纯净水、鸡蛋、奶茶、酸奶、牛奶、维他类功能饮料这些对身体有益处的商品更加喜爱。

从售货机 A 的类别画像可以看到，人们对于“肉干/豆制品/蛋”类型的商

品有较高的需求，其次对于茶饮料、乳制品、功能饮料等对身体有益的饮品有较高需求。



图 7-6 售货机 B 特征画像



图 7-7 售货机 B 类别画像

从售货机 B 的特征画像可以看出，对于品牌的偏好，B 区域的顾客依然喜欢购买卫龙、维他、统一、怡宝、伊利、无穷农场等几个品牌的商品；口味依然是原味类型；不过 B 地区售货机对于维他类型的饮品明显更加喜爱。

从类别画像上可以看出，该地区人们同样喜欢“肉干/豆制品/蛋”类型的商品，不过饮品类该区域的顾客更喜欢茶饮料和乳制品。



图 7-8 售货机 C 特征画像



图 7-9 售货机 C 类别画像

从 C 售货机的特征画像可以看出除了卫龙、怡宝、维他等几个热门品牌外，该区域对燕塘，东鹏特饮、阿萨姆几个品牌的喜好明显高于 A,B 区域。此外奶茶类商品在该地区更受欢迎。



图 7-10 售货机 D 特征画像



图 7-11 售货机 D 类别画像

从 D 区域的特征画像可以看出，统一，卫龙，原味，怡宝，维他几个字样基本占据了整个画像的主要空间，这说明该地区的消费者对于统一，卫龙，怡宝，维他，原味商品这些热门品牌或商品的偏好更加突出。



图 7-10 售货机 E 特征画像



图 7-11 售货机 E 类别画像

从特征画像来看，E地区的售货机画像于A地区的售货机画像比较相似，说明这两个区域的人群偏好也比较相似。不过从类别画像中又可以看出些许差别，E地区的人们对于饼干糕点、功能饮料、碳酸饮料几个类型的商品的喜爱程度明显高于其他地区。

从总体上来看，统一、卫龙、怡宝、维他等几个品牌以及原味商品、肉干/豆制品/蛋、茶饮料、乳制品、功能饮料类型的商品代表了五个区域的人群的共同喜好。即便存在着区域差异，以及人流量的不同，这些关键词依然代表着该区域绝大多数人的偏好。

8. 售货机销售特点分析及营销建议

8.1 售货机销售特点分析

8.1.1 相同点

通过多方面的分析该商场五个区域的销售情况有如下几个特点：

- ① 人们对于饮料类的购买量明显高于非饮料类，且饮料类商品获得的利润更高。
- ② 人们普遍喜欢统一、卫龙、怡宝、维他几个品牌。
- ③ 人们普遍喜欢原味、肉干/豆制品/蛋、茶饮料、乳制品、功能饮料等类型的商品。
- ④ 平均每单交易额在3-5元，人们更倾向于物美价廉的消费理念。
- ⑤ 1月、2月、3月和7月交易量明显偏低。
- ⑥ 6月、9月、10月、11月和12月为销量高峰期。
- ⑦ 每天0时~7时交易量较少，售货机主要销售时段为下午。

8.1.2 不同点

下面总结了，各个售货机出共同点之外的其他特点：

- ① 根据销量来看，D地区人流量较少，E地区人流量最多。
- ② B地区的人群，相比于其他地区，对于茶饮料，和乳制品的偏好更高。
- ③ C地区对燕塘，东鹏特饮、阿萨姆几个品牌的喜好明显高于A、B区域。
- ④ 奶茶类商品在C、E两个地区比其他地区更受欢迎。
- ⑤ D地区对于统一、卫龙、怡宝、维他几个热门品牌，及原味类型的商品的喜好更加突出。
- ⑥ E地区的人们对于饼干糕点、功能饮料、碳酸饮料几个类型的商品的喜爱程度明显高于其他地区。

8.2 营销建议

根据以上对售货机销售状况的分析，可以给出如下建议：

- ① 在上架商品时优先考虑统一、卫龙、怡宝、维他等几个热门品牌的同类商品。
- ② 如果一种商品有多种口味，那么原味的商品比重应该适当提高。
- ③ 从人们喝的饮品上看，主要是茶饮料、乳制品、功能饮料，以及纯净水，这些饮料都对身体有益，所以商家应该考虑将那些有益的饮品的上架比例提高一些，适当减低一些其他类型饮品的供应量。
- ④ 对于非饮料类商品，人们更喜欢肉干、豆制品、蛋，饼干这些有营养的食物，所以商家可以考虑以“营养健康”为销售理念，优化供给结构。
- ⑤ 由于人们更倾向于物美价廉的商品，所以可以适当减少一些高价商品的供应量。
- ⑥ 由于 D 地区人流量较少，且该地区人们主要购买哪些热门品牌的商品，所以在该地区的商品供应总量可以适当减少，且主要放置一些热门品牌商品及原味商品。
- ⑦ 由于 E 地区销售量较大，所以应该时刻保证该区域售货机的库存充足。
- ⑧ 9，10，11，12 月，为需求量较高的时间段，商家可以考虑在这几个月时间内下架一些冷门类别的商品，增加不同类型的热门类别商品，以增加多样性，从而促进顾客的购买欲望。
- ⑨ 1 月、2 月、3 月和 7 月交易量明显偏低，可以减少商品储备，降低补货频率，从而节约成本。
- ⑩ 如果需要补货，7 点之前为最佳时间。

9. 业务预测

9.1 预测

本部分主要任务为对每台售货机各大类商品 2018 年 1 月的交易额进行预测，由于只有一年的销售数据，所以考虑以 2017 年每一天的交易数据为基础数据，并对 2018 年 1 月份 31 天的交易额进行预测。另外，如果当天没有交易数据，则该天交易额为 0。

该任务属于时序数据预测任务，所以使用了较常用 ARMA 模型，并以均方误差为目标通过网格搜索寻找最优参数： $p(\text{AR})$ ， $q(\text{MA})$ 。最终模型拟合效果，及预测结果如下(以售货机 A 为例)：



图 9-1 售货机 A 饮料类商品销售额预测示意图



图 9-2 售货机 A 非饮料类商品销售额预测示意图

通过图 9-1 和图 9-2 的拟合结果可以看出，拟合结果基本体现了真实数据随时间的变化趋势，所以通过该模型所预测的结果具有一定参考价值。

通过上图的预测结果，最终计算出每台售货机各大类商品 2018 年 1 月的交易额预测结果如下图所示：

表 9-1 2018 年 1 月份交易额预测结果

地区 类别					
	A	B	C	D	E
饮料类	1916.9	2759.3	2989.3	1766.6	4737.7
非饮料类	1387.2	1431.6	1528.6	930.7	2342.6

通过上表的预测结果可以看出，依然是 D 地区的售货机交易额最低，E 地区的交易额最高，且饮料类商品的交易额高于非饮料类。这与实际数据所反应的信息相吻合，这就进一步说明了预测结果对于 2018 年 1 月份的销售额具有一定的参考意义。

9.2 小结

通过上节中对预测结果的分析，虽然拟合结果与真实数据有着相同的变化趋势，具有一定的参考意义，但从准确性上看，还是有着一定的差距。

另外，在之前对 2017 年各售货机的交易数据分析过程中，发现售货机商品的销售状况受春节等节假日的影响较大，且有一定的季节性变化，所以将一年的数据作为预测的基础数据，还不足以准确的把握售货机的销售规律。所以数据量越大才能使模型发挥出更好的效果。