

目录

一、背景介绍.....	2
1.1、行业背景.....	2
1.2、分析目标.....	2
二、数据预处理与分析.....	2
2.1、数据的预处理.....	2
2.2、数据分析.....	3
三、数据的可视化.....	5
3.1、2017 年 6 月销量前 5 的商品分析.....	5
3.2、每月交易额分析.....	6
3.3、各售货机月环比增长分析.....	7
3.4、各售货机的毛利润分析.....	12
3.5、气泡图分析.....	13
3.6、热力图分析.....	14
四、生成自动售货机画像.....	17
4.1、标签.....	18
4.2、画像.....	18
五、业务分析.....	22

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，自主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

本报告通过对 2017 年 1 月 1 日至 2017 年 12 月 31 日在不同地点安放的 5 台自动售货机的商品销售数据进行数据分析，并完成下列两个目标：1、对自动售货机的销量，库存和盈利三个方面各项指标的计算与图表绘制，并对每台售货机的销售额进行预测。2、为每台售货机所销售的商品贴上标签，使其能够展现销售商品的特征。

二、 数据预处理与分析

2.1、数据的预处理

先下载 anaconda，然后从网站上下载项目数据。打开其中的附件 1，发现其中有一个不合法的数据：最后一个数据的支付时间为 2017/2/29 3:44:00 PM，而且有一些数据的应付金额和实际金额均为 0。故要先对附件 1 的数据进行预处理，把其中的不合法数据删除。将预处理后得到的 csv 文件命名为 data，下面对其按照售货机的类型分类，分别提取每一台售货机的销售数据，得到 5 个 csv 文件，这些文件分别包含了五个不同售货机的销售数据，如下（说明：我用我的电脑运行后得到的文件打开后得到的是乱码，我觉得可能是编译的时候出了问题，我用了 gbk 和 utf-8 两种方式来编译，要么就是可以运行得到有乱码的文件，要么就是运行的时候报错了。但是我尝试用 jupyter notebook 输出得到的数据，并没有问题，而且将所得到的文件以记事本的方式打开也是没有问题的，且对后面的任务的执行并没有产生影响。所以我认为这是可以暂时忽略的小问题）：

 task1-1A	2019/8/29 20:44	Microsoft Excel ...	1,376 KB
 task1-1B	2019/8/29 20:44	Microsoft Excel ...	1,753 KB
 task1-1C	2019/8/29 20:44	Microsoft Excel ...	1,881 KB
 task1-1D	2019/8/29 20:44	Microsoft Excel ...	1,130 KB
 task1-1E	2019/8/29 20:44	Microsoft Excel ...	3,060 KB

2.2、数据分析

要想对 2017 年 5 月的销售数据进行分析，首先要提取 data 中的 5 月份的数据，将提取得到的 5 月份数据保存在 csv 文件中，并命名为“5 月数据”。

提取了 5 月份的数据后，分别算出 5 台售货机的交易额、订单量和所有售货机的交易额和订单总量，处理后可得到以下两个表格：

售货机	A	B	C	D	E	所有售货机
交易额 (元)	3385.1	3681.2	3729.4	2392.1	5699.0	18886.8
订单量 (件)	750	860	782	553	1287	4232

表 1 2017 年 5 月各售货机的交易额和订单量及其总和

由表一可知 A、B、C 三台售货机的交易额和订单量都比较接近，D 售货机交易额和订单量就相对较低，但是 E 售货机交易额和订单量都很高，甚至超过了 D 售货机的 2 倍。由此说明售货机的种类对交易额和订单量还是有一定的影响的。

现在计算每台售货机每月的每单平均交易额与日均订单量，计算结果如以下表所示：

月份	每单平均交易额	日均订单量
1 月	4.52005988023952	10.774193548387096
2 月	3.8640350877192984	4.071428571428571
3 月	3.5854901960784313	8.225806451612904
4 月	4.0641891891891895	14.8
5 月	4.513466666666667	24.193548387096776
6 月	4.066887417218543	55.36666666666667
7 月	4.106315789473684	15.32258064516129
8 月	3.36375939849624	21.451612903225808
9 月	4.311357074109721	34.63333333333333
10 月	4.031005765534913	50.354838709677416
11 月	4.475409836065574	38.63333333333333
12 月	3.799248873309965	64.41935483870968

表 2 A 售货机每月的每单平均交易额与日均订单量

月份	每单平均交易额	日均订单量
1 月	3.753005464480874	11.806451612903226
2 月	3.2556756756756755	6.607142857142857
3 月	3.6147169811320756	8.548387096774194
4 月	4.095666666666667	20.0
5 月	4.2804651162790694	27.741935483870968
6 月	4.083450513791239	61.63333333333333
7 月	4.41453488372093	11.096774193548388

8 月	3.5841997961264016	31.64516129032258
9 月	4.142126436781609	58.0
10 月	4.122513607125185	65.19354838709677
11 月	4.277207696102615	67.56666666666666
12 月	3.6719981875849568	71.19354838709677

表 3 B 售货机每月的每单平均交易额与日均订单量

月份	每单平均交易额	日均订单量
1 月	4.363031914893617	12.129032258064516
2 月	3.8260869565217392	7.392857142857143
3 月	3.7699619771863118	8.483870967741936
4 月	4.415710382513661	24.4
5 月	4.769053708439897	25.225806451612904
6 月	4.518506666666667	62.5
7 月	4.00407358738502	24.548387096774192
8 月	3.913582208101668	40.61290322580645
9 月	4.440526001195457	55.766666666666666
10 月	4.290756683280471	71.19354838709677
11 月	4.356877897990727	64.7
12 月	3.9530130636325325	76.54838709677419

表 4 C 售货机每月的每单平均交易额与日均订单量

月份	每单平均交易额	日均订单量
1 月	3.7359375000000004	8.258064516129032
2 月	3.0886524822695036	5.035714285714286
3 月	4.305729166666667	6.193548387096774
4 月	3.8161363636363634	14.666666666666666
5 月	4.325678119349005	17.838709677419356
6 月	4.061105722599418	34.366666666666667
7 月	4.256507936507937	10.161290322580646
8 月	3.3211484593837532	23.032258064516128
9 月	3.907339449541285	32.7
10 月	3.9006773920406435	38.096774193548384
11 月	3.8623140495867765	40.333333333333336
12 月	3.5833534378769607	53.483870967741936

表 5 D 售货机每月的每单平均交易额与日均订单量

月份	每单平均交易额	日均订单量
1 月	4.680225988700565	11.419354838709678
2 月	3.638372093023256	9.214285714285714

3 月	4.305714285714286	11.290322580645162
4 月	4.159888268156425	29.833333333333332
5 月	4.4281274281274285	41.516129032258064
6 月	3.8400698215671065	85.93333333333334
7 月	3.928976572133169	26.161290322580644
8 月	3.804470854555744	57.0
9 月	4.139393203883495	137.33333333333334
10 月	3.680100937274693	89.48387096774194
11 月	4.290919976052685	167.03333333333333
12 月	4.1689729397293975	104.90322580645162

表 6 E 售货机每月的每单平均交易额与日均订单量

由以上表格我们可以看出每单平均交易额都处于一个比较稳定的水平，都是在 3 点多 4 点多之间。每台售货机在 2, 3 月份的日均订单量都会比较少。售货机 A, B, C 的日均订单量都比较相近，售货机 D 的日均订单量相对较低，售货机 E 的日均订单量就相对较高。由此看出人们都比较喜欢在售货机 E 消费。也印证了刚刚说的售货机的种类对订单量有一定的影响的观点。

三、 数据的可视化

数据的可视化主要旨在借助于图形化手段，清晰有效地传达与沟通信息，是数据直观体现的一种方式。得到数据的图像可以直观地看出数据的结构特点。

3.1、2017 年 6 月销量前 5 的商品分析

要画出 2017 年 6 月销量前 5 的商品销量柱状图，就要先得到 6 月份的数据，然后算出各商品的销量并对其进行降序排序，得到前 5 个商品即为 2017 年 6 月销量前 5 的商品。画图如下：

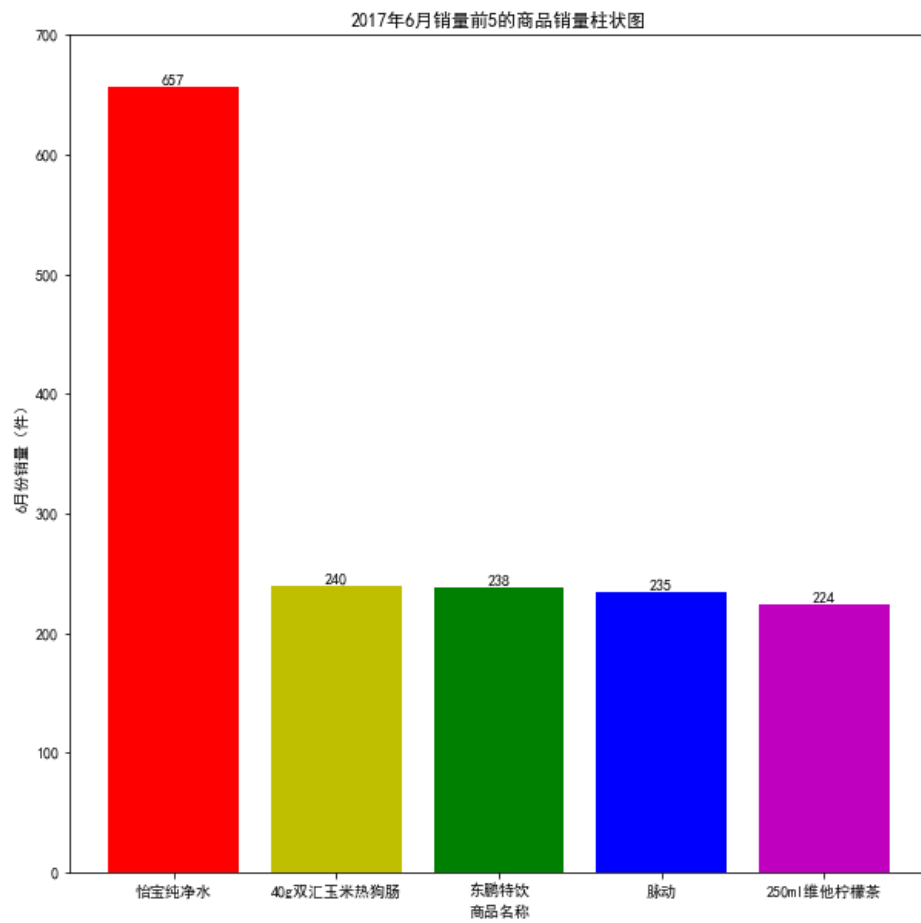


图 1 2017 年 6 月销量前 5 的商品销量柱状图

3.2、每月交易额分析

折线图是一种可以反映变量之间关系的图表，要画出各售货机每月总交易额的折线图，就要先算出每个月交易额分别是多少。画图如下：

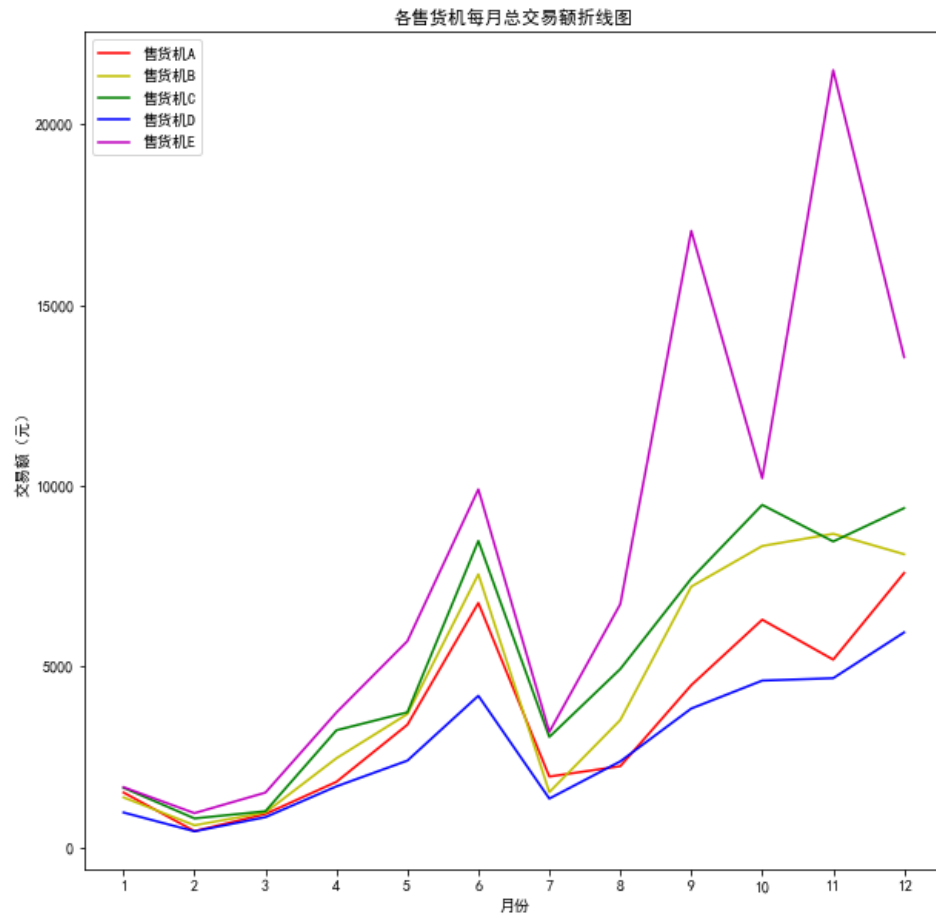


图 2 各售货机每月总交易额折线图

由折线图可看出在 1~3 月每台售货机交易额普遍较低，4 月过后交易额都普遍比较高，但是从 6 月到 7 月各售货机的交易额的下降幅度非常大。

3.3、各售货机月环比增长分析

交易额月环比增长率等于下个月交易额与本月的交易额的差值与本月交易额的比，先算出每台售货机从每个月到下个月的月环比增长率，从而画出每台售货机的月环比增长率的柱状图如以下（要计算的量是增长率，一般以百分数为单位，故在计算过程中都乘以 100）：

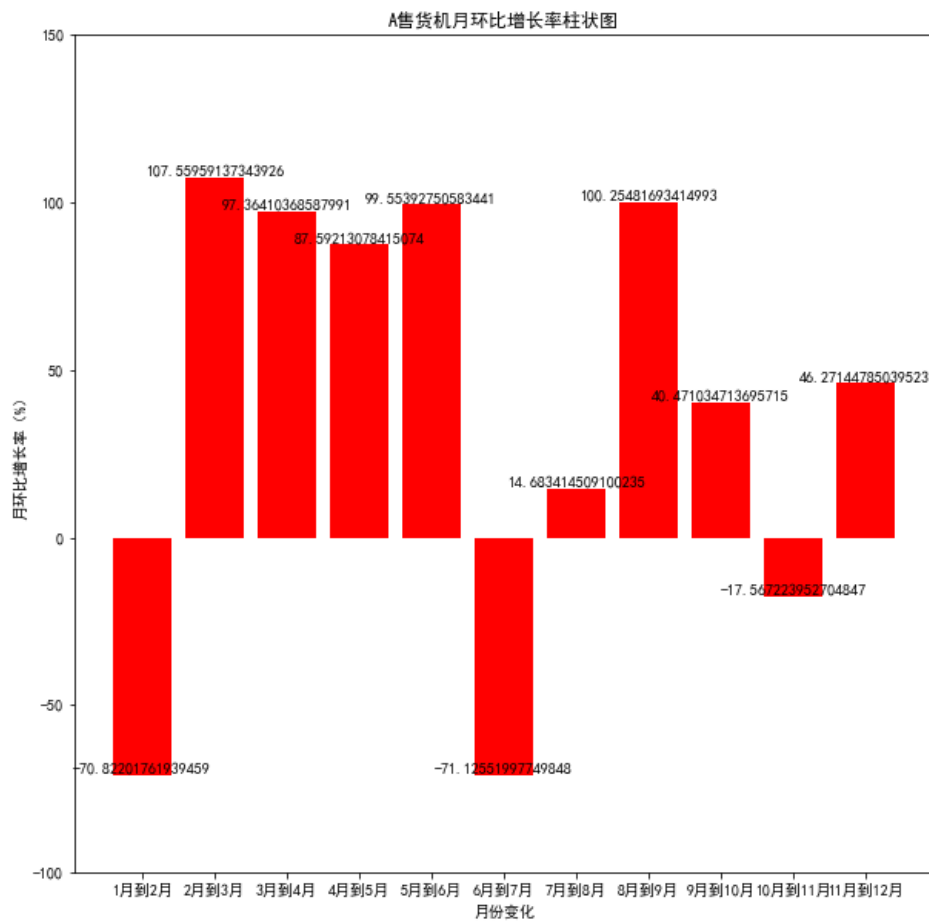


图 3 A 售货机月环比增长率柱状图

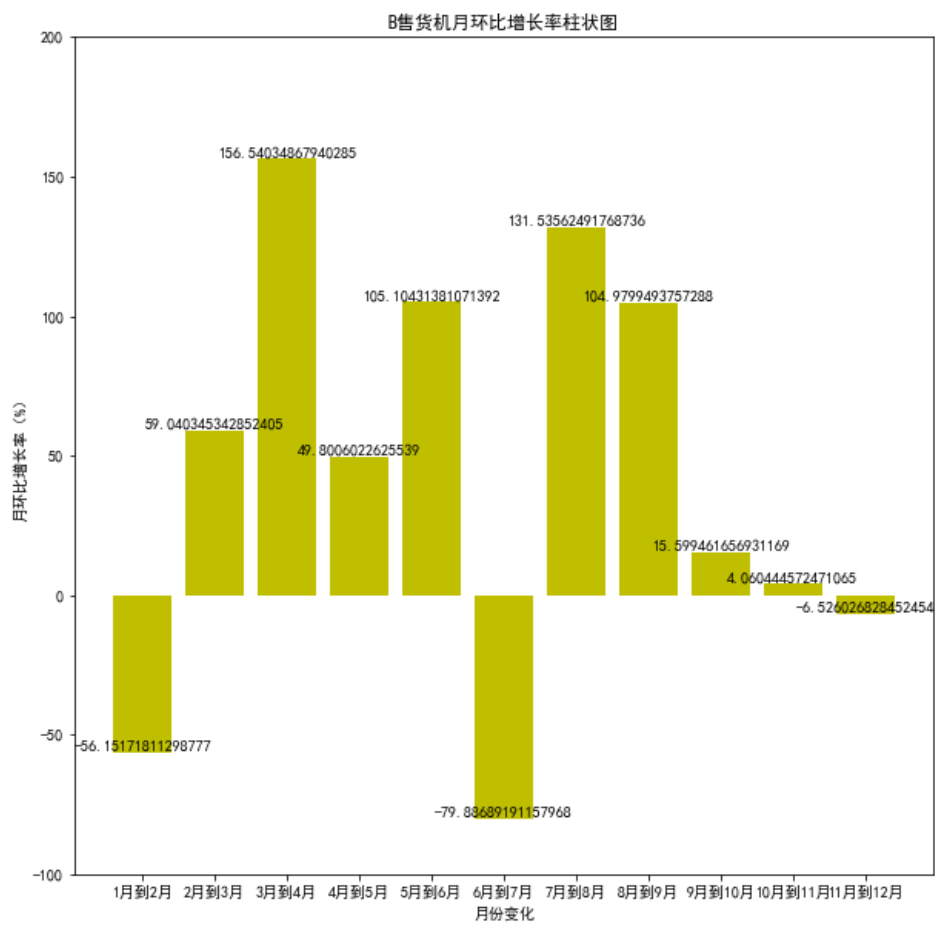


图 4 B 售货机月环比增长率柱状图

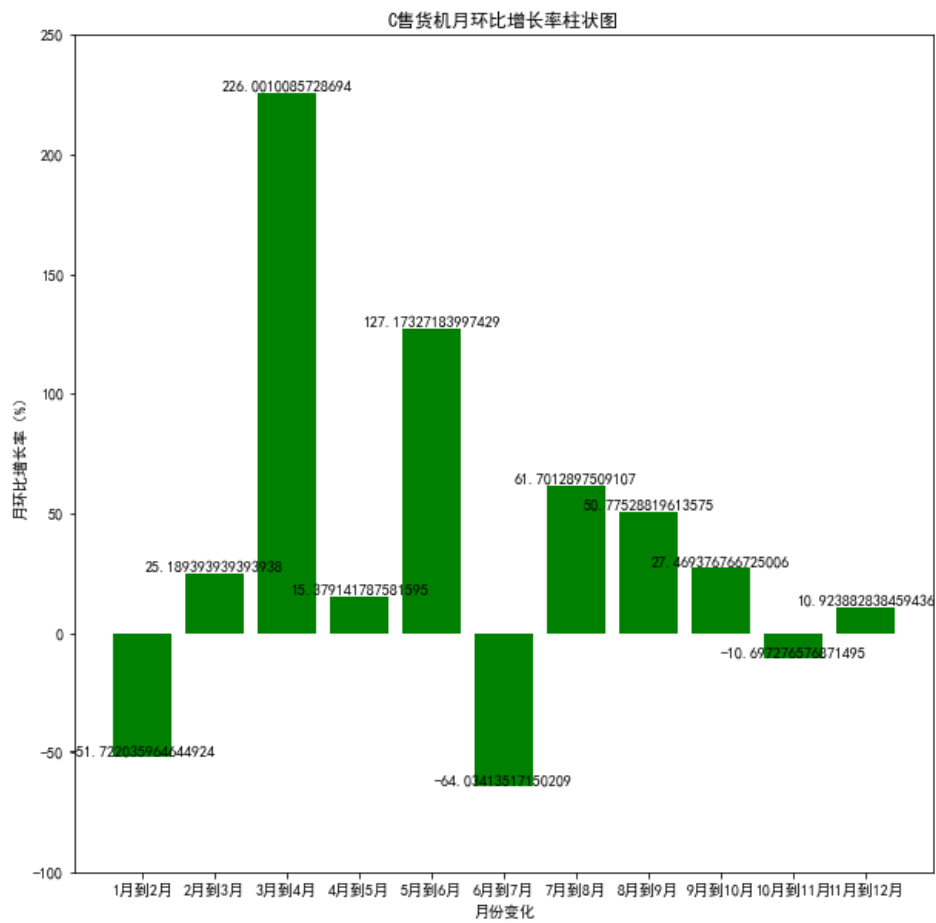


图 5 C 售货机月环比增长率柱状图

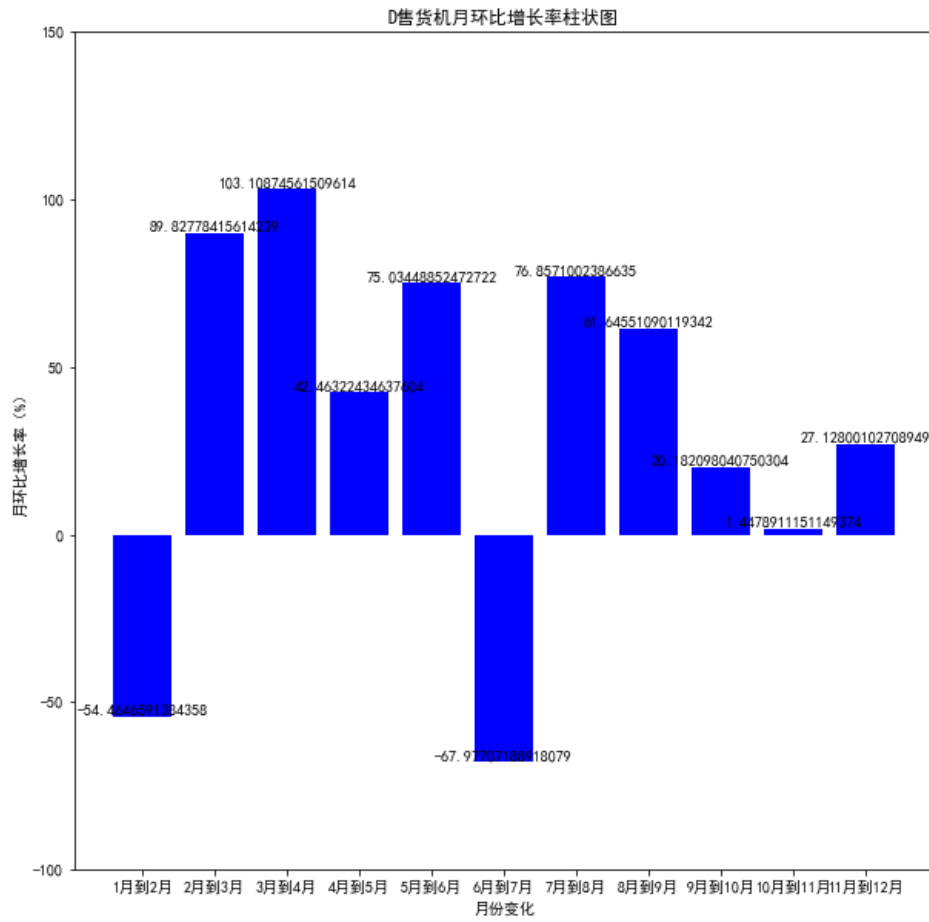


图 6 D 售货机月环比增长率柱状图

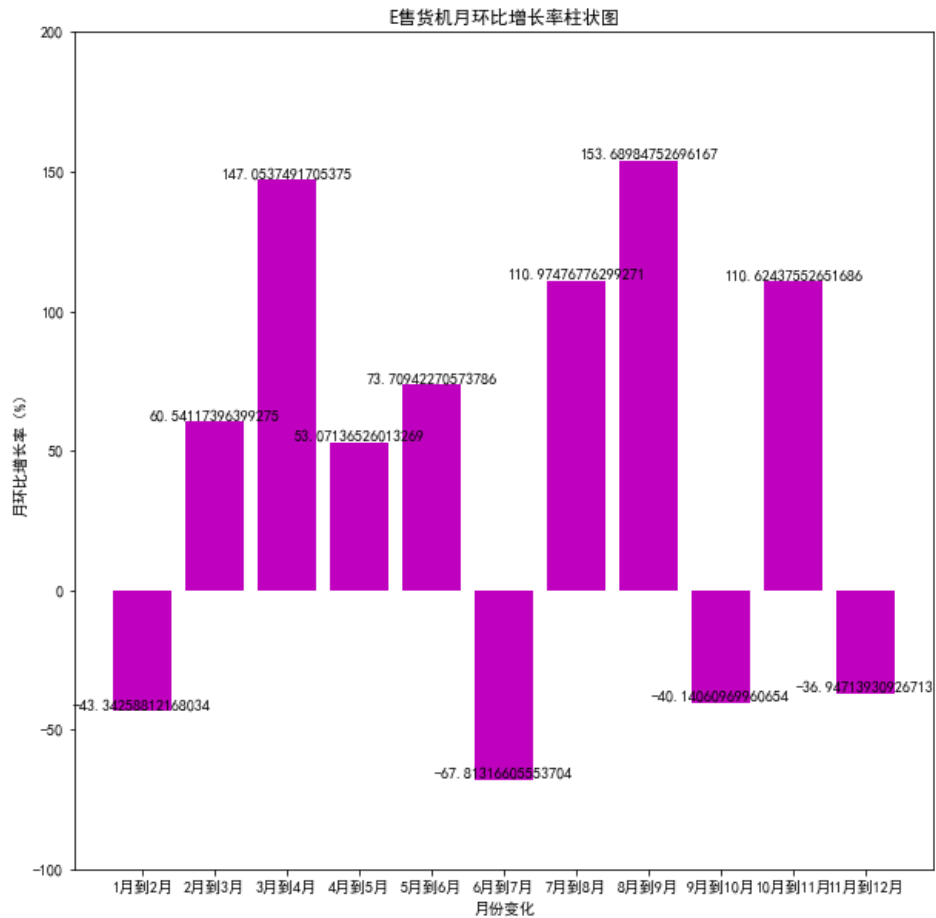


图 7 E 售货机月环比增长率柱状图

由以上 5 个柱状图可发现每个售货机在 1 月到 2 月和 6 月到 7 月的交易额都呈现下降的趋势, 其他月份的交易额大部分都是呈现上升的趋势 (有一小部分以不太大的幅度下降), 不过 E 售货机 9 月到 10 月和 11 月到 12 月的交易额都呈现幅度较大下降的趋势。

3.4、各售货机的毛利润分析

根据给出的饮料类和非饮料类的率, 我们应该想到要对处理过的附件 1 (也就是 data) 按照大类来分类, 但是在 data 中并没有饮料类和非饮料类的标签, 这时就需要将附件 2 中的数据导入到 data 中, 然后先按照不同售货机分成 5 个文件, 再分别对 5 个文件按照大类来划分, 从而算出每台售货机饮料类和非饮料类的毛利润分别是多少, 再根据它们的毛利率权重 (饮料类 25%, 非饮料类 20%) 算出每台售货机的毛利润, 从而就可以画出各售货机毛利润占总毛利润比例饼图了, 饼图如下:

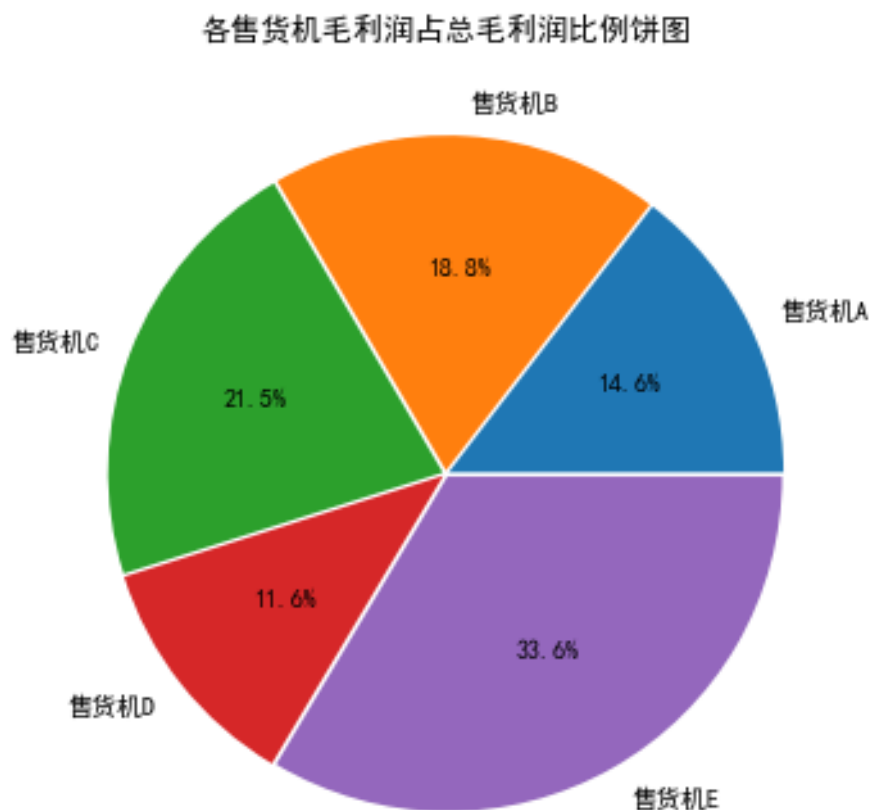


图 8 各售货机毛利润占总毛利润比例饼图

3.5、气泡图分析

气泡图是可用于展示三个变量之间的关系。它与散点图类似，绘制时将一个变量放在横轴，另一个变量放在纵轴，而第三个变量则用气泡的大小来表示。排列在工作表的列中的数据(第一列中列出 x 值, 在相邻列中列出相应的 y 值和气泡大小的值)可以绘制在气泡图中。气泡图与散点图相似，不同之处在于：气泡图允许在图表中额外加入一个表示大小的变量进行对比。

现在我们根据商品的二级类来分类，先得到每种二级类商品在每个月交易额，然后根据这些数据画出气泡图，如下图所示：

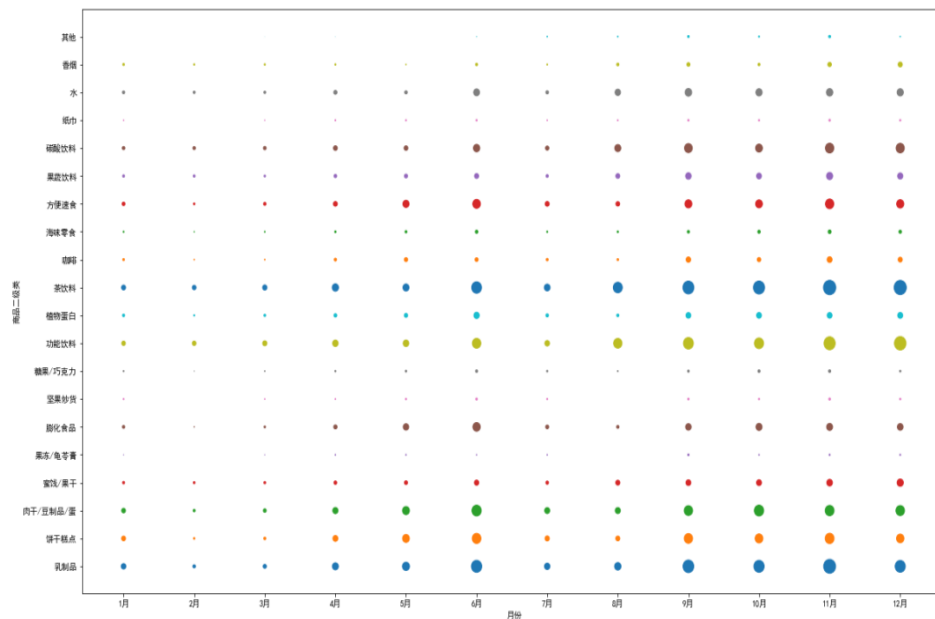


图 9 每月交易额均值气泡图

根据得到的气泡图我们可以看出各售货机在 6 月、9 月、10 月、11 月和 12 月的交易额均值普遍较高，而其他月份的交易额均值就普遍偏低。水、碳酸饮料、方便速食、茶饮料、功能饮料、肉干/豆制品/蛋、饼干糕点、乳制品类的商品的交易额均值比较高，而其他类别的商品的则比较低。

3.6、热力图分析

下面我们来画出 6、7、8 三个月的订单量热力图，如下：

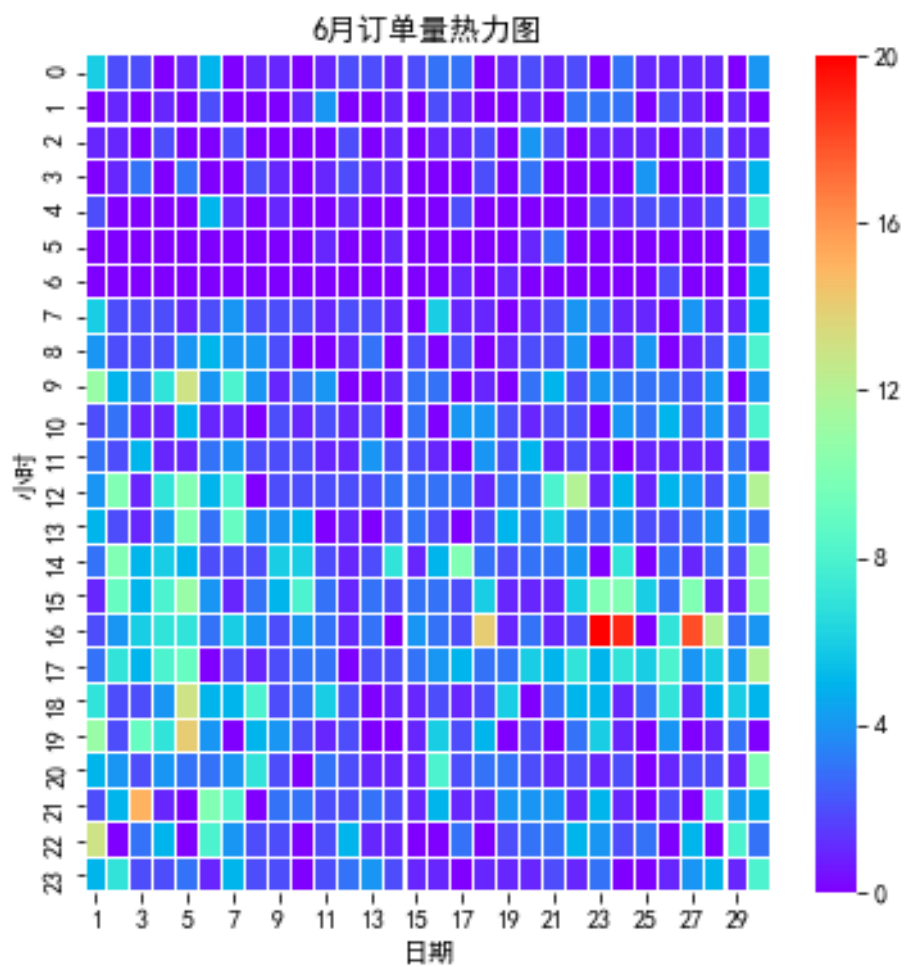


图 10 6月订单量热力图

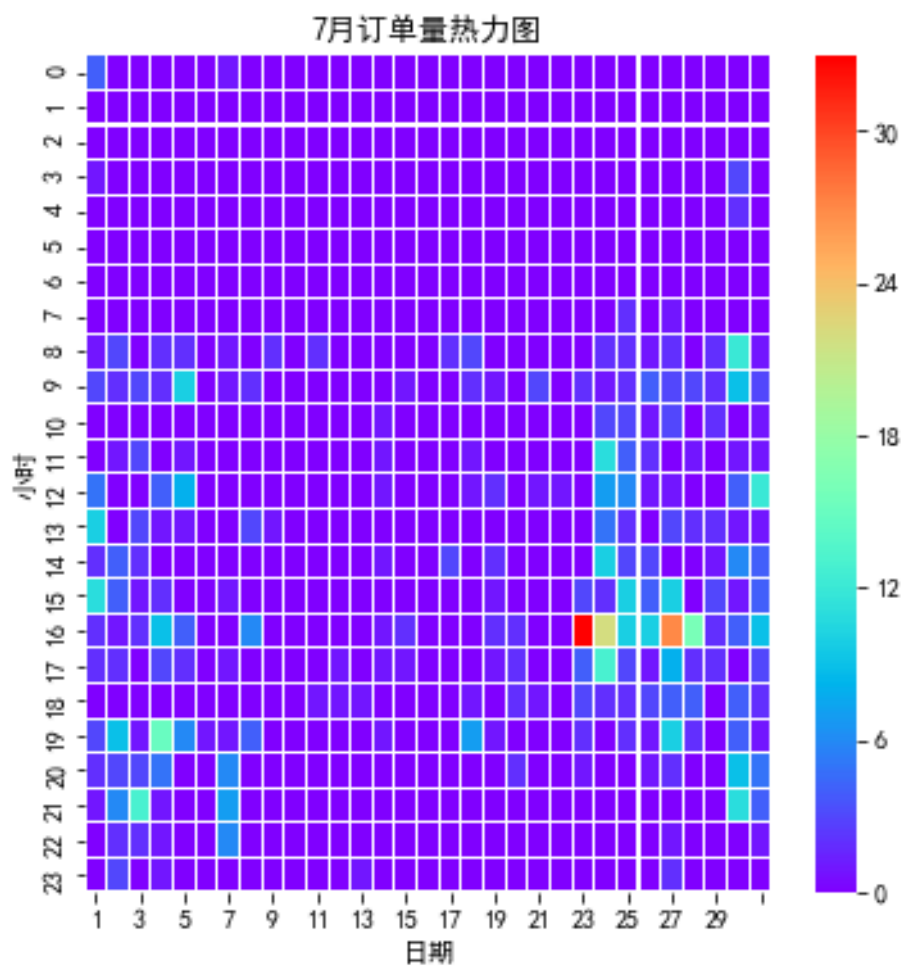


图 11 7 月订单量热力图

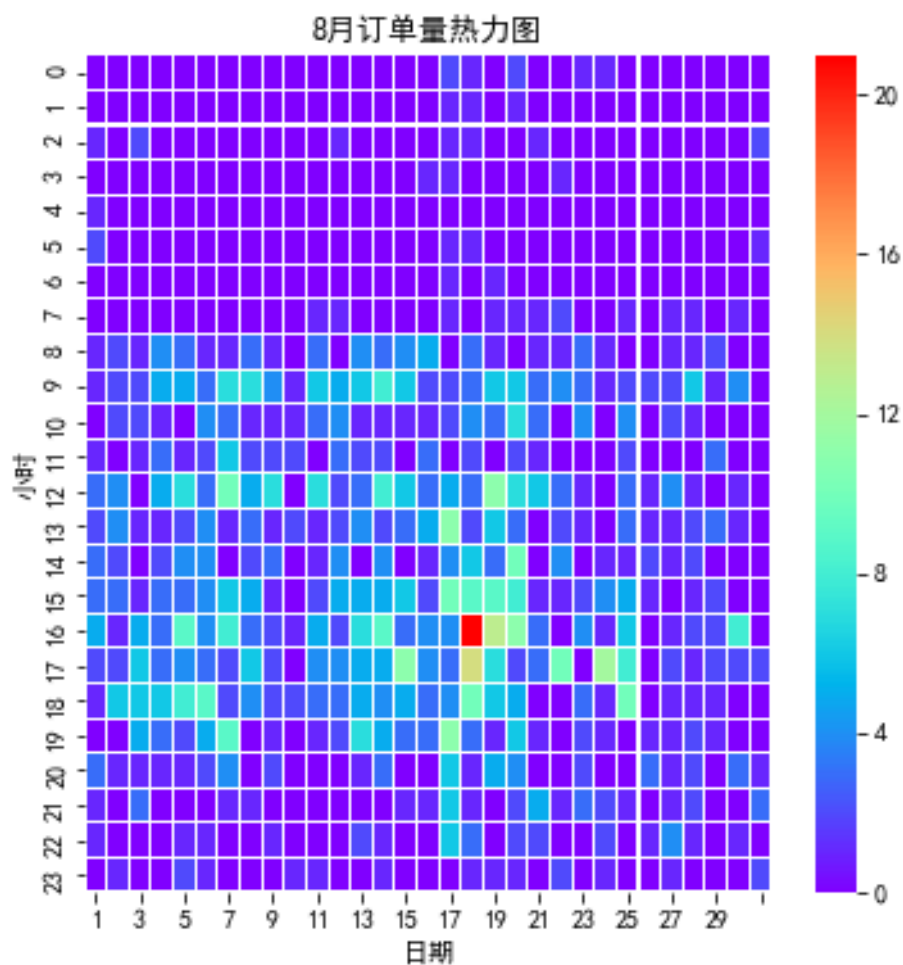


图 12 8月订单量热力图






从这 3 张热力图我们不难发现，这 3 个月的订单量普遍在 9~21 时较多，而 6 月热力图颜色深的地方是最密集的，8 月热力图比较密集，而 7 月热力图就显得非常稀疏了，故可以看出 6 月订单量最多，8 月次之，7 月最少。

四、 生成自动售货机画像

通过表中的销售数据我们可以看出，有些商品销量非常多，而有一些则较少。在不同的售货机中，同一种商品的销售状态可能也会不同。这样就会被贴上不同的标签。我们对饮料类的商品销售数据进行分析，分析它们的销售状态，从而得到它们的标签。标签是对某类特定群体或对象的某项特征进行的抽象分类和概况。

4.1、标签

为了得到各种饮料的标签，我们用四分位数的方法。先得到所有饮料类的商品的销量，然后找出其中的两个四分位数：较大四分位数和较小四分位数。商品销量大于较大四分位数的商品就贴上畅销的标签，小于较小四分位数的商品贴上滞销的标签，而介于较小四分位数和较大四分位数之间的商品则贴上正常的标签。根据这些原理写出代码运行得到 5 个 csv 文件，如下：

 task3-1A	2019/8/29 19:43	Microsoft Excel ...	3 KB
 task3-1B	2019/8/29 19:44	Microsoft Excel ...	3 KB
 task3-1C	2019/8/29 19:45	Microsoft Excel ...	3 KB
 task3-1D	2019/8/29 19:45	Microsoft Excel ...	3 KB
 task3-1E	2019/8/29 19:47	Microsoft Excel ...	3 KB

4.2、画像

画像是由代表性的标签组合而成的。现在我们来分别对每台售货机的各种商品销售状态来做一个画像。要想画出各售货机画像的词云图，就要先下载几个安装包：wordcloud 和 jieba。其次运行画画像的代码，其中先把每台售货机的商品都分别保存到 5 个 txt 文件中，再从网上下载一个 jpg 格式的图片。可得到以下词云图画像：

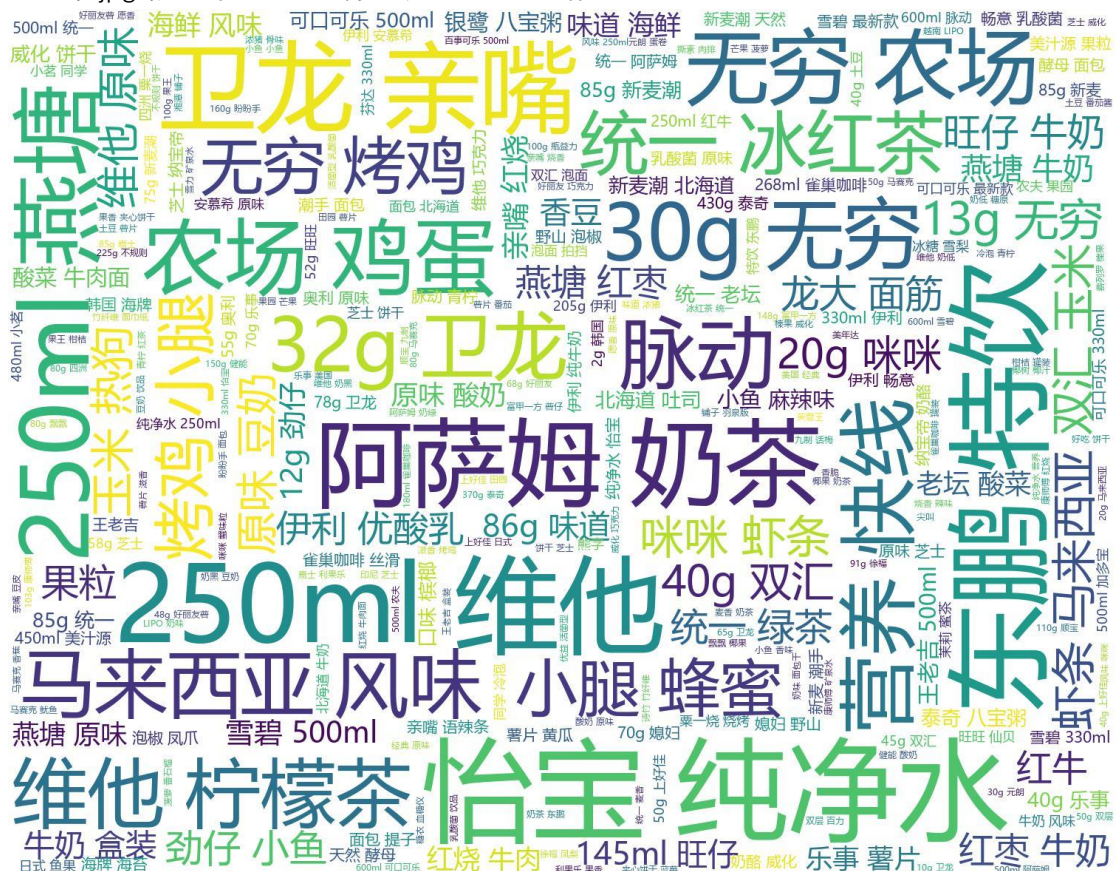


图 13 A 售货机的画像



图 15 C 售货机的画像

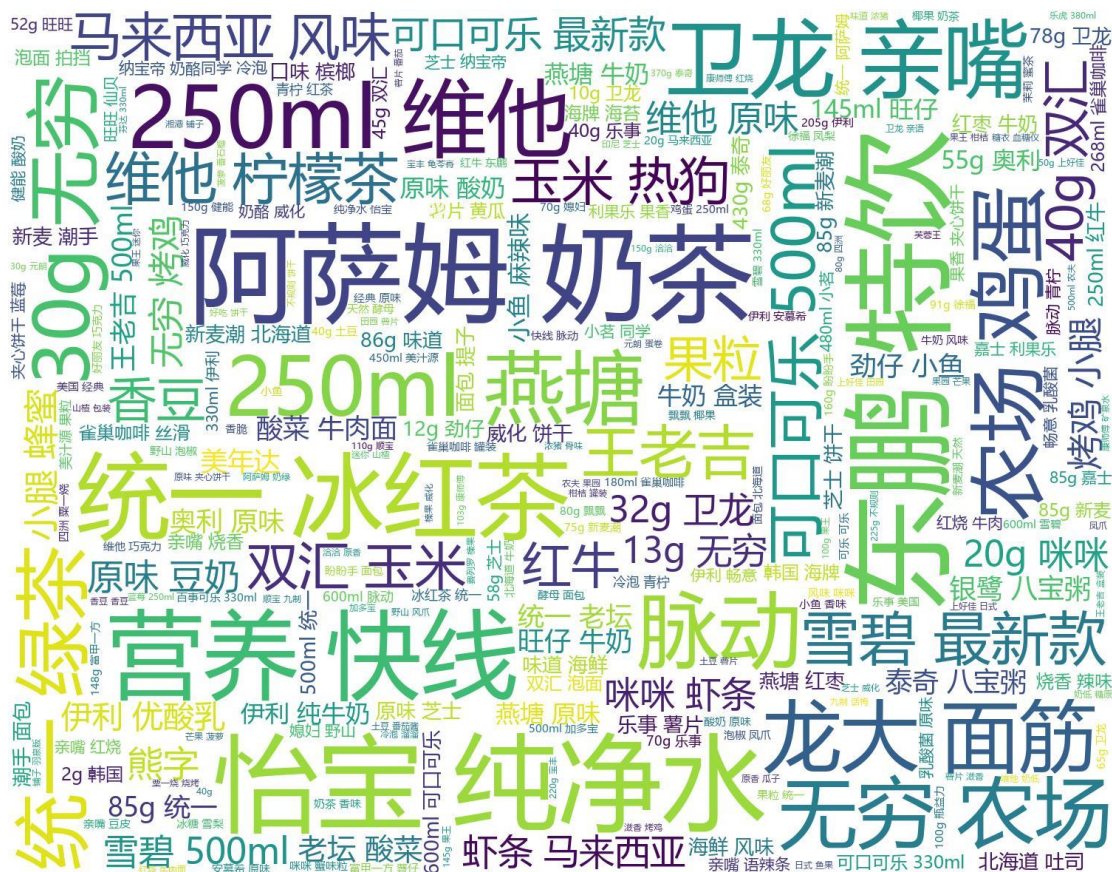


图 16 D 售货机的画像

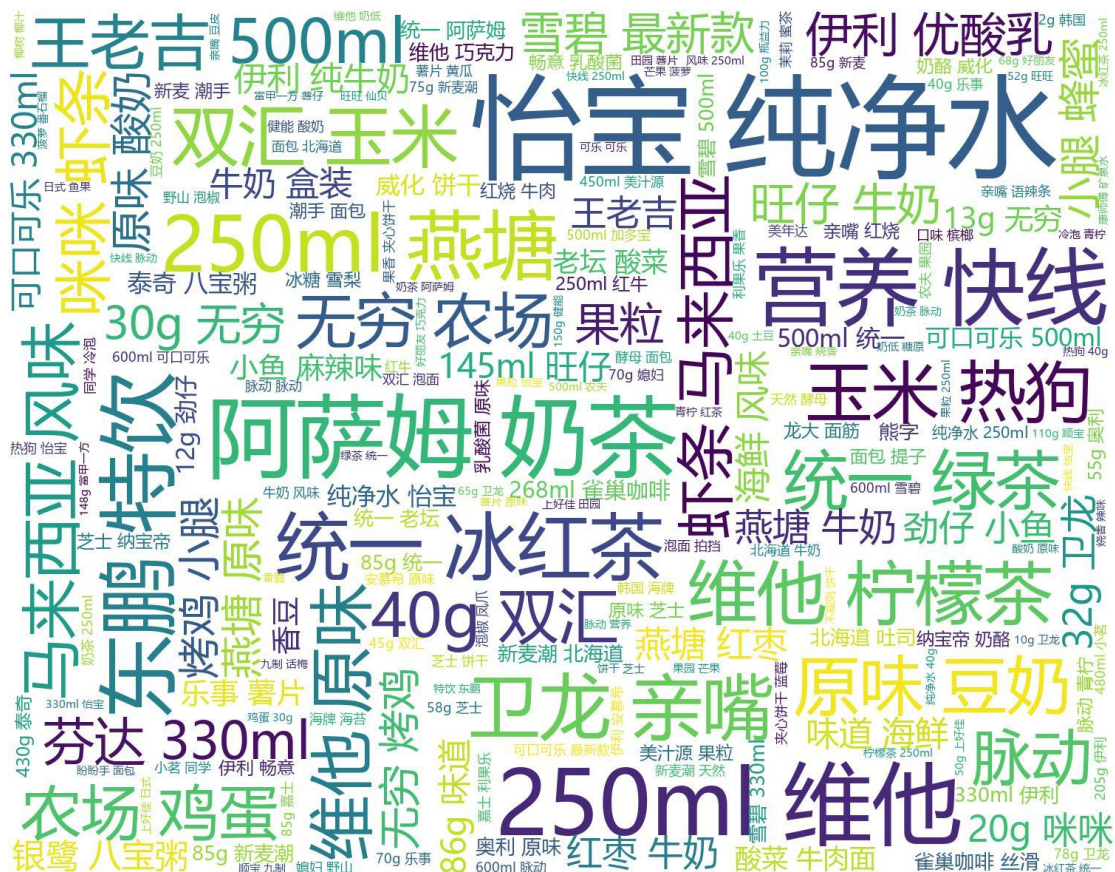


图 17 E 售货机的画像

通过观察这些画像我们可以看出怡宝纯净水在所有画像中的字体都是挺大的,说明怡宝纯净水在每台售货机中都十分畅销。而且饮料类的商品字体普遍比非饮料类的字体基本上都要比非饮料类的大,说明饮料类商品比非饮料类商品更畅销。

售货