

# 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。

## 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

## 1、数据预处理

经过对附件 1 数据的观察我们发现存在相对一部分商品的应付金额为 0 元，这意味该商品在未进行任何优惠活动之前它的标价已经是 0 元，这显然不符合常理，所以把价格为 0 元的商品删除只保留应付金额大于 0 元的商品。通过对时间序列的查看可以发现有一行商品的时间格式为 12 小时然而其他商品的时间格式都是 24 小时，为了严谨起见我们把该数据删除，同时筛选出应付金额大于实际金额的数据并保存为最终的数据。为了后面分析的方便我们把数据预处理后的附件 1.csv 保存为 onefinal.csv，附件 1 则保持不变。

## 2、数据分析

### 2.1 交易额以及订单量分析

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

表一 2017 年 5 月各售货机交易额及订单量详情表

由上表可以看出 E 售货机的交易额和订单量为五台售货机中最高的，D 售货机的交易额和订单量为五台售货机中最低的。而且 E 售货机的交易额和订单量均为 D 售货机的两倍以上，由此看出地域对售货机的交易额影响还是很大的。

### 2.2 平均交易额以及日均订单量分析

	A 售货机	B 售货机	C 售货机	D 售货机	E 售货机
1 月	4.52	3.75	4.36	3.73	4.68
2 月	3.86	3.25	3.82	3.08	3.63
3 月	3.58	3.61	3.76	4.30	4.30
4 月	4.06	4.09	4.41	3.81	4.15
5 月	4.51	4.28	4.76	4.32	4.42
6 月	4.06	4.08	4.51	4.06	3.84
7 月	4.10	4.41	4.00	4.25	3.92
8 月	3.36	3.58	3.91	3.32	3.80
9 月	4.31	4.14	4.44	3.90	4.13
10 月	4.03	4.12	4.29	3.90	3.68
11 月	4.47	4.27	4.35	3.86	4.29
12 月	3.79	3.67	3.95	3.58	4.16

表二 2017 年各月每台售货机每单平均交易额（元）详情表

由上表可以看出每台售货机每个月的每单平均交易额基本稳定在 3-5 元之间，而且售货机之间的每单平均交易额差距也不大。总体来说，售货机每单的平均交易额较为稳定。

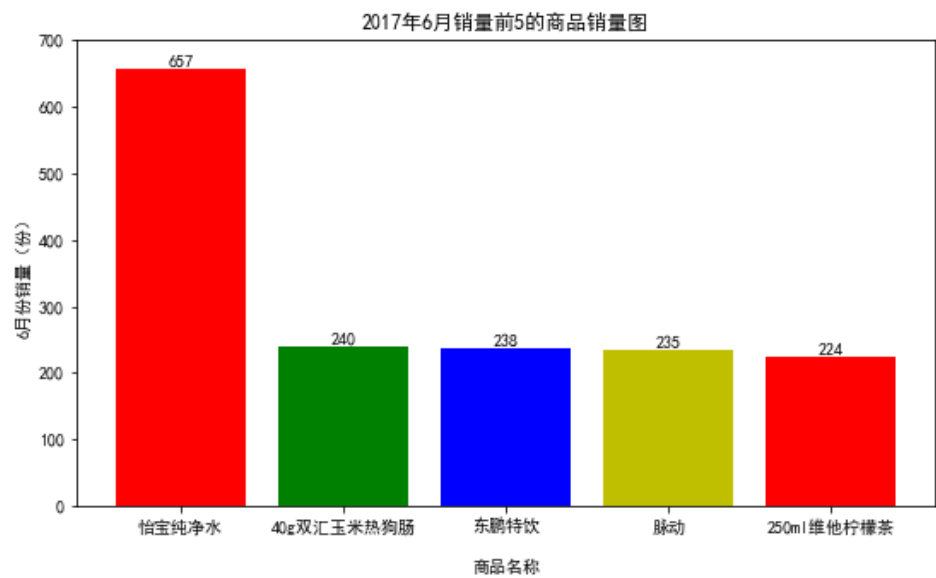
	A 售货机	B 售货机	C 售货机	D 售货机	E 售货机
1 月	10.7	11.8	12.1	8.25	11.4
2 月	4.07	6.60	7.39	5.03	9.21
3 月	8.22	8.54	8.48	6.19	11.2
4 月	14.8	20.0	24.4	14.6	29.8
5 月	24.1	27.7	25.2	17.8	41.5
6 月	55.3	61.6	62.5	34.3	85.9
7 月	15.3	11.0	24.5	10.1	26.1
8 月	21.4	31.6	40.6	23.0	57.0
9 月	34.6	58.0	55.7	32.7	137
10 月	50.3	65.1	71.1	38.0	89.4
11 月	38.6	67.5	64.7	40.3	167
12 月	64.4	71.1	76.5	53.4	104

表三 2017 年各月每台售货机日均订单量（件）详情表

由上表可以看出不同的月份每台售货机日均订单量差异较大，总体来说 6、10、11、12 月是销售旺季，日均订单量远高于其他月份。2、3 月是销售淡季，日均订单量远低于其他月份。且不同售货机之间的日均订单量差距较大，从中也可以看出地域对售货机的订单量影响还是较大的。

### 三、数据分析与可视化

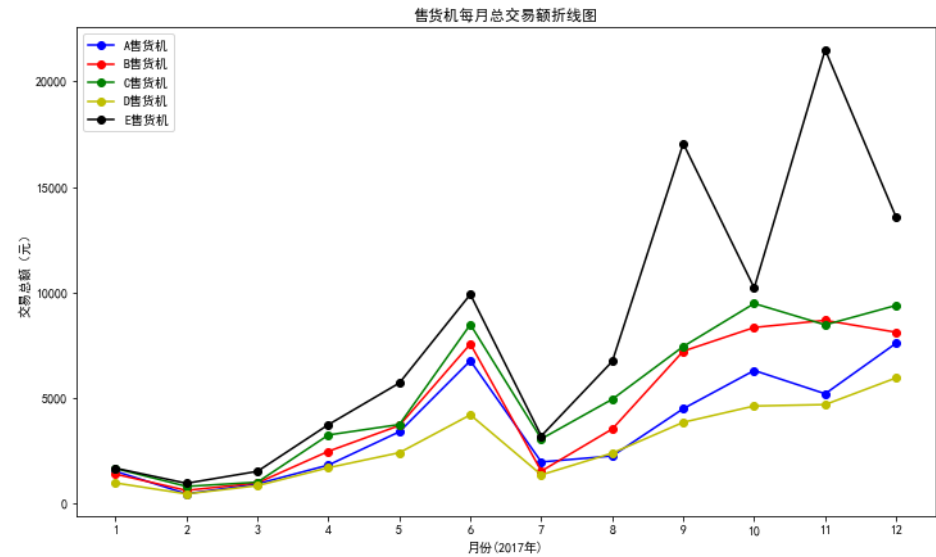
## 1、2017 年 6 月销量前 5 商品分析



图一 2017 年 6 月销量前 5 商品柱状图

由上图可以看出在 2017 年 6 月销量前 5 的 5 种商品中有 4 种是饮料类的商品，只有 1 种是非饮料的商品，因此可以初步认为饮料类商品比较受欢迎。此外，可以看出怡宝纯净水的销量远高于其他商品，这一数据体现了客户对怡宝纯净水商品的信任与热爱。

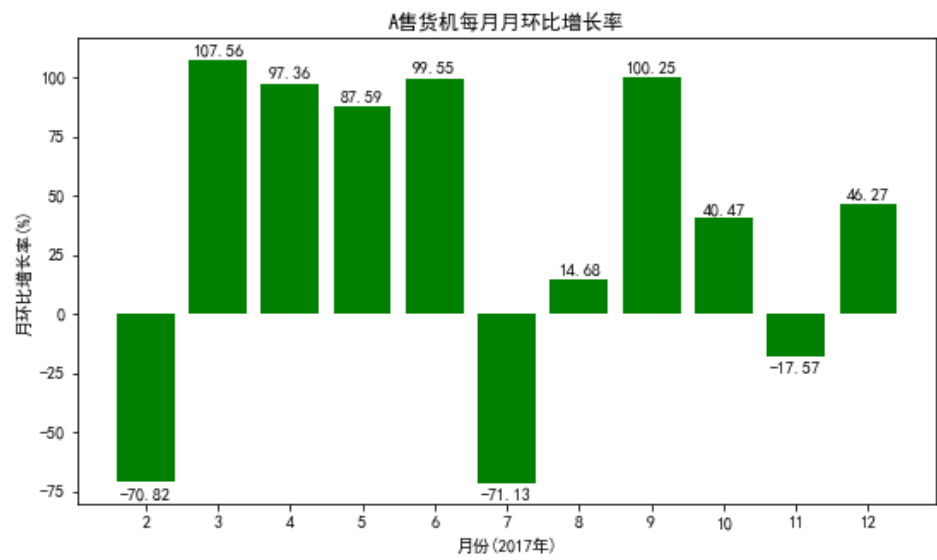
## 2、各售货机每月总交易额分析



图二 各售货机每月总交易额折线图

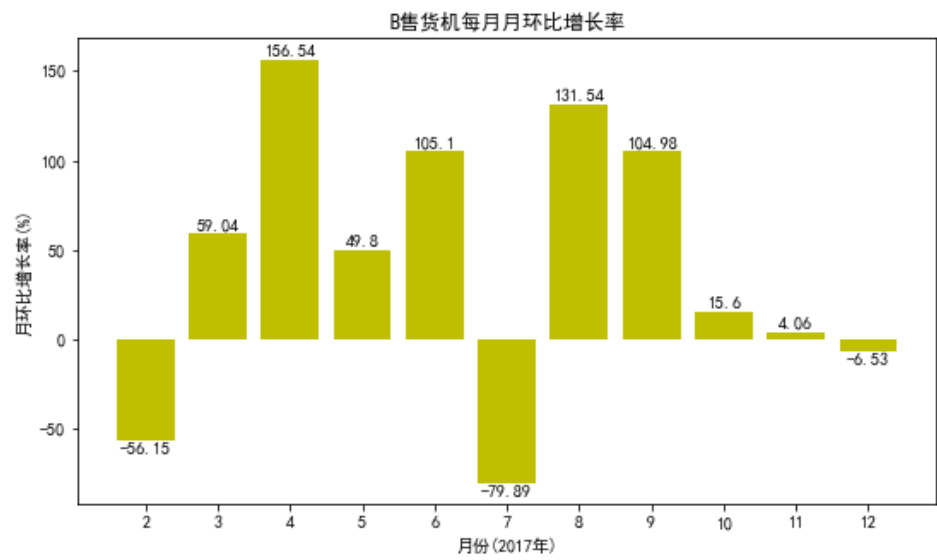
由上图可以看出 E 售货机各月的交易额基本都是五台售货机中最高的，D 售货机则为五台售货机中最低的。而且 E 售货机的交易额在旺季时远高于其他 4 台售货机，由此看出地域对售货机的交易额影响较大且 E 售货机销量较好。

3、售货机交易额月环比增长率分析



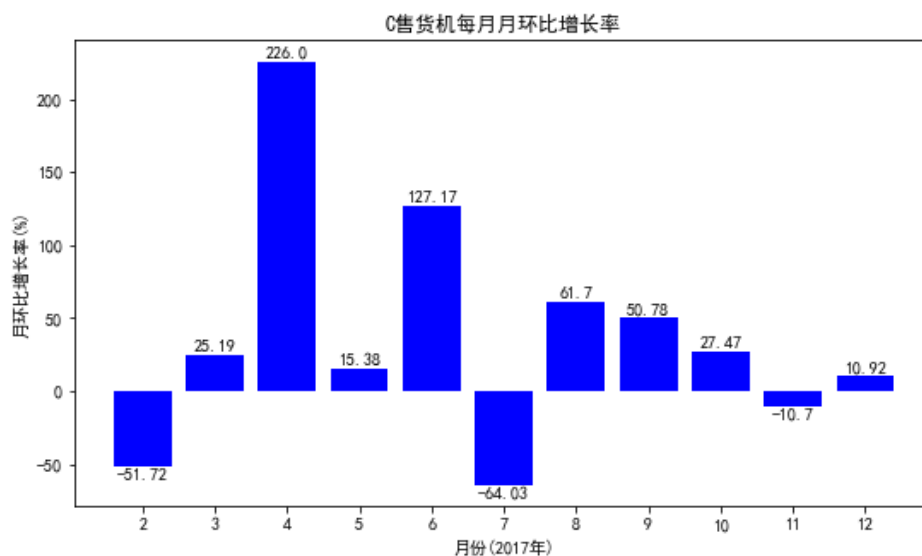
图三 A 售货机交易额月环比增长率柱状图

由上图可以看出 A 售货机 2、7、11 月三个月的交易额月环比增长率为负数，意味着这三个月相对上个月的交易额下降了，而且 2、7 月下降幅度很大。其他月份的交易额月环比增长率均为正数，意味着这些月份相对上月的交易额都有所上升。



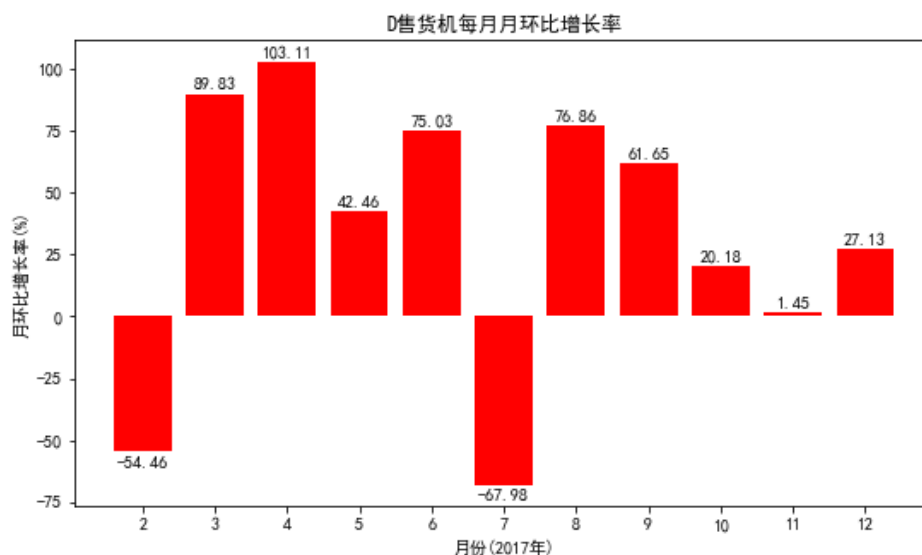
图四 B 售货机交易额月环比增长率柱状图

由上图可以看出 B 售货机 2、7、12 月的交易额月环比增长率为负数，意味着这三个月相对上个月的交易额下降了，而且 2、7 月下降幅度很大。其他月份的交易额月环比增长率均为正数，意味着这些月份相对上月的交易额都有所上升。



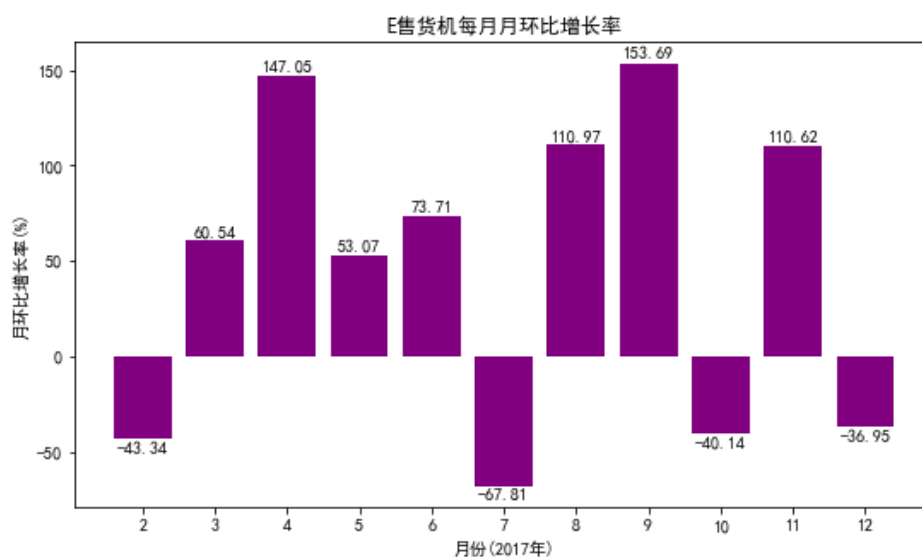
图五 C 售货机交易额月环比增长率柱状图

由上图可以看出 C 售货机 2、7、11 月的交易额月环比增长率为负数，意味着这三个月相对上个月的交易额下降了，而且 2、7 月下降幅度很大。其他月份的交易额月环比增长率均为正数，意味着这些月份相对上月的交易额都有所上升。



图六 D 售货机交易额月环比增长率柱状图

由上图可以看出 C 售货机 2、7 月的交易额月环比增长率为负数，意味着这二个月相对上个月的交易额下降了，而且 2、7 月下降幅度很大。其他月份的交易额月环比增长率均为正数，意味着这些月份相对上月的交易额都有所上升。



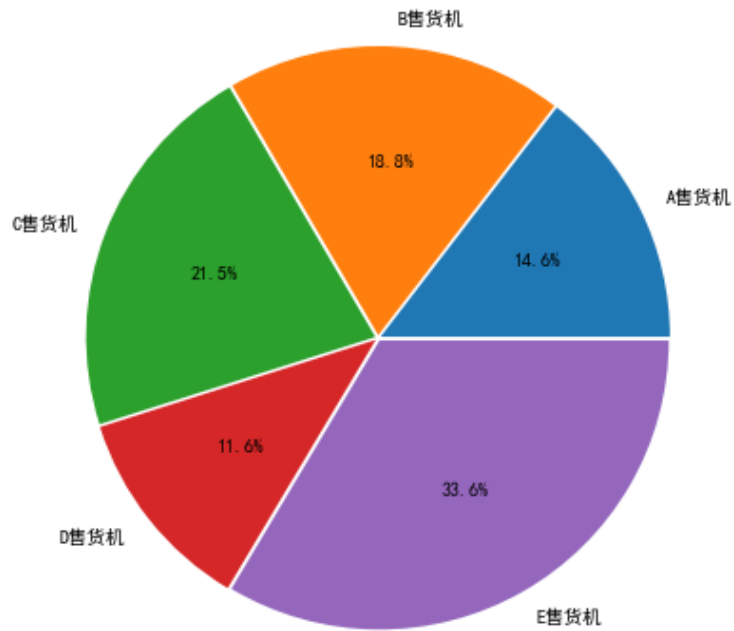
图七 E 售货机交易额月环比增长率柱状图

由上图可以看出 C 售货机 2、7、10、12 月的交易额月环比增长率为负数，意味着这四个月相对上个月的交易额下降了，而且 2、7 月下降幅度很大。其他月份的交易额月环比增长率均为正数，意味着这些月份相对上月的交易额都有所上升。

综合上述五个图来看，五台售货机 2、7 月的交易额月环比增长率均为负数，而且降幅都比较大。11 月份的交易额月环比增长率则基本在 0 附件，意味着这三月的交易额相对上个月的浮动不大（除了 E 售货机 11 月份）。

#### 4、毛利润分析

每台售货机毛利润占总毛利润比例的饼图

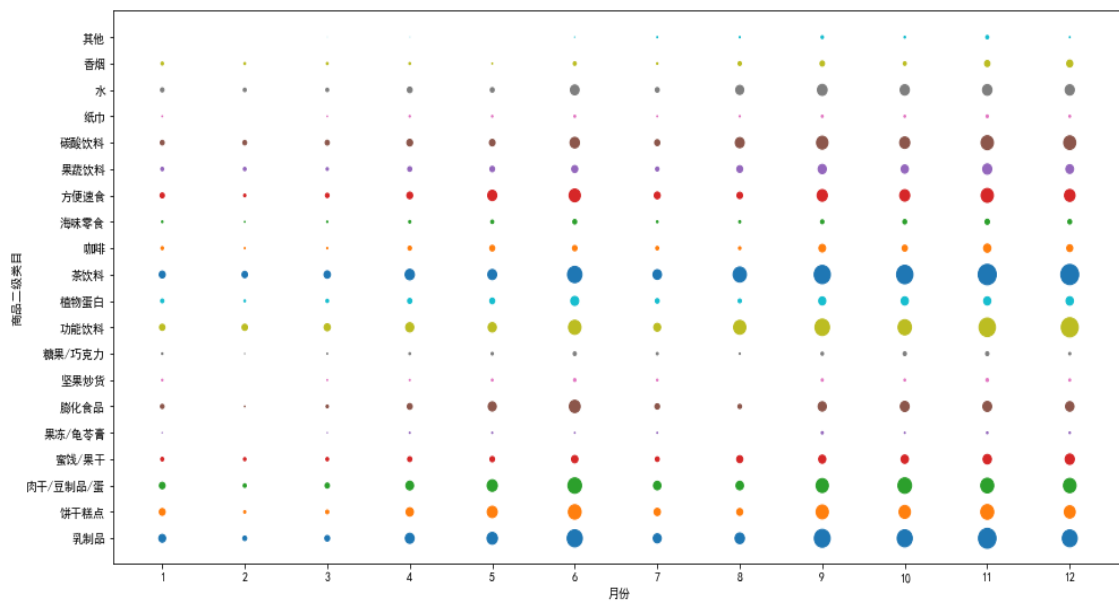


图八 每台售货机毛利润占总毛利润比例饼图

由上图可以看出 E 售货机毛利润在五台售货机中最大, 且为总体毛利润的 33.6%。D 售货机毛利润在五台售货机中最小。E 售货机的毛利润为 D 的 3 倍左右。

具体的数值是 A 售货机的毛利润为：9707.5，B 售货机的毛利润为：12543.8，C 售货机的毛利润为：14291.3，D 售货机的毛利润为：7703.3，E 售货机的毛利润为：22341.2。

### 5、每月交易额均值气泡图分析

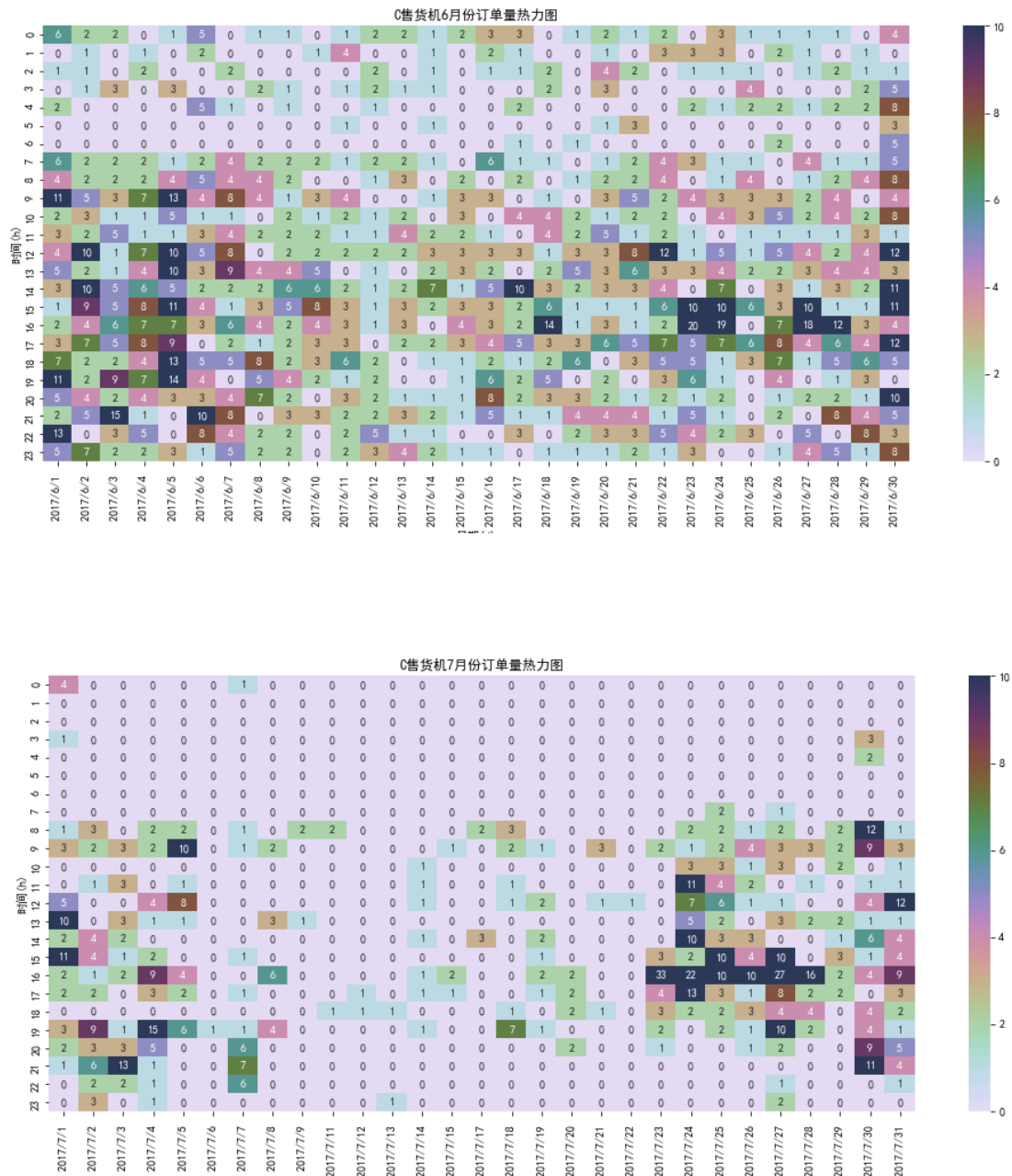


图九 每月交易额均值气泡图

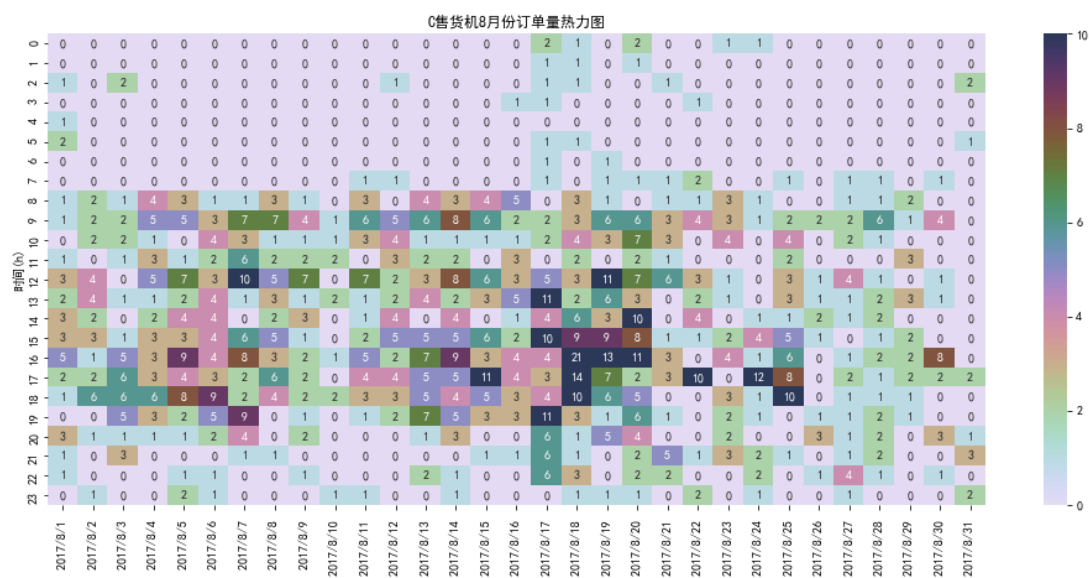
月交易额均值的定义为该二级类的月交易额除以该月的天数。

由上图我们可以看出每个二级类的交易额均值呈现的规律大体上是一样的。4、5、6、8、9、10、11、12 这七个月的气泡比其他月的气泡明显要大，这意味着这几个月的销量比较好。而且乳制品、饼干糕点、肉干/豆制品/蛋、茶饮料、功能饮料这几个二级类的气泡总体较大，意味着这几个品类的商品在这些类别中比较受欢迎。

## 6、C 售货机 6、7、8 月订单量热力图分析







图十 C 售货机 6、7、8 月订单量热力图

由上面三幅图可以看出：

以 24 小时的订单量来分析，每天的 0 点到 7 点订单量都非常少甚至为 0，在每天的 9 点到 20 点的订单量则较高，几乎包含了一天订单量的 90%。

以每天的订单量来分析，6、7 月份的订单量的高峰期主要在月初和月末，月中相对较少。而 8 月份却截然相反，订单量的高峰期主要在月中，月初和月末订单量相对较少。

综上所述，即使是同一台售货机，不同的月份中，订单量的高峰期出现的时间也可能不同。在同一天中，一般情况下凌晨的订单量非常少几乎为 0，中午和傍晚的订单量则较多，占了全天订单量的绝大多数。

## 四、售货机画像

### 1、售货机商品销售分析以及为商品贴标签

通过对商品的销量观察，我们发现订单量很多和很少的商品较少，于是我们把商品的标签值分为热销、正常、滞销，并把订单量大于等于 220 的商品定义为热销，把订单量小于等于 5 的定义为滞销，其他的为正常。并且根据商品的订单量画出了如下画像。

## A 售货机画像

[illegible][illegible]

### C 售货机画像



## D 售货机画像



## E 售货机画像



图十一 售货机画像图

由上述画像我们可以看出，在热销商品中饮料类的商品占了大多数，总体来看饮料类的商品销量好于非饮料类的。特别是怡宝纯净水、阿萨姆奶茶、东鹏特饮、营养快线、统一冰红茶等商品的销量远高于其他商品，在非饮料类中卫龙亲嘴，双汇玉米、农场鸡蛋、无穷农场等商品销量较好，经营者在下次的进货中可以考虑多进这些商品，同时减少滞销商品的进货。

## 五、销量预测

### 1、2018 年 1 月销量预测

#### 1.1 分别计算出五台售货机饮料类和非饮料类 1-12 月份的交易额

经过计算得到如下数据，其中 A、B、C、D、E 里面的数值分别代表五台售货机饮料类 1-12 月的交易额，AA、BB、CC、DD、EE 里面的数值分别代表五台售货机非饮料类 1-12 月的交易额。

A=[717.3 313.5 557.5 940.3 1353.1 3748.7 1106.7 1586.5 2782.7 3345.0 3104.8 4424.8]  
B=[870.3 506.5 688.0 1601.7 1950.6 4491.1 850.4 2716.1 4535.8 5567.4 5371.8 5845.5]  
C=[919.6 624.0 780.7 2137.2 1708.9 4783.6 1849.7 3901.7 4653.9 6121.5 5465.9 6608.0]  
D=[479.8 361.0 503.5 1048.0 1066.8 2100.3 770.8 1854.6 2748.9 2859.2 3333.3 3967.4]  
E=[455.7 196.4 399.5 1373.6 2501.8 4026.4 1181.8 1446.5 5251.3 3526.9 6851.3 4241.2]

AA=[792.4 127.0 356.7 864.1 2031.8 3006.3 843.6 650.4 1696.7 2947.3 2082.1 3162.2]  
BB=[503.3 95.8 269.9 855.6 1730.5 3059.2 668.1 800.0 2671.4 2764.1 3298.1 2258.6]  
CC=[720.9 168.0 210.8 1095.1 2020.3 3688.6 1197.3 1025.5 2775.1 3348.2 2990.8 2772.4]  
DD=[476.6 74.5 323.1 631.1 1325.2 2086.6 570.0 516.7 1084.2 1747.4 1340.1 1973.7]  
EE= [455.6 196.4 399.5 1373.5 2501.7 4026.4 1181.7 1446.4 5251.3 3526.9 6851.3 4241.1]

1.2 构建 bp 神经网络

说明：此神经网络把 2017 年 12 个月的数据用作训练样本，来预测 2018 年 1 月的数据。  
程序：

程序包括两部分，main.m 为主程序，NNetwork.m 为训练及预测程序，NNetwork 输入为真实数据，输出为预测数据。具体程序见附录一

1.3 代入数据得出预测值

把上面 1.1 中的数据代入 main.m 程序中 t=[ ]的位置处预测出五台售货机 2018 年 1 月份的交易额（元）。具体数据如下表格。

	饮料类	非饮料类
A 售货机 1 月预测交易额	4424.5	3161.8
B 售货机 1 月预测交易额	5839.0	198.4
C 售货机 1 月预测交易额	6606.3	2824.9
D 售货机 1 月预测交易额	3962.7	2084.6
E 售货机 1 月预测交易额	551.7	1748.8

表四 2018 年 1 月每台售货机饮料类和非饮料类交易额预测表

由表四我们可以看出 A、B、C、D 售货机 1 月饮料类的预测交易额都远高于 E 售货机的，但是从上面的分析和观察《各售货机每月总交易额折线图》我们发现，E 售货机在 2017 年的交易额基本上都是五台售货机最高的，在 2018 年 1 月份 E 售货机的交易额远低于其他四台售货机不大可能。而且在 2017 年五台售货机 1 月份的交易额都挺低的，理论上在 2018 年 1 月份的交易额也不可能太高。于是我们认为该模型仍然有改进空间。

1.4 模型优缺点分析及问题改进

通过观察《各售货机每月总交易额折线图》后我们发现售货机的交易额受季节和月份的影响非常大，然而附件 1 所给的数据只有 2017 年一年的数据，也就是只有 1 次 1 月份的数据，我们无法观察次年 1 月的交易额到底该如何变化，没有一个完整的周期很难较准确的预测次年 1 月的数据。

而且通过所学知识我们也知道神经网络要想预测的准确率较高需要有较多的数据，一般来说当训练数据达到 20 个或者以上时，预测的误差能维持在 10%之内。

经营者要想得到更好的预测数据，最好能够提供过去两年或两年以上的交易额数据。

## 附录一

main 程序如下

```
clear all      %清除变量空间
clc           %清除命令窗口所有数据
close all     %关闭所有 figure 窗口
% 数据
t=[]; %%原始数据（数据的观测值）
n=1;  %每次预测个数
    for i=12:12
s=t(1:i);
[a, anew]=NNetwork(s,n);    %调用神经网络预测程序
disp(['预测的第',num2str(i+1),'个值为']);
anew                        %预测值
pred=a(1,:);
pred(end+1)=anew;    %加上预测值新数据
    end
```

NNetwork 程序如下

```
function [a, anew]=NNetwork(x,n)
% x 为神经网络训练数组
% n 为预测数据的个数
p=1:length(x); %%数据的个数
warning off
%数据归一化
[pn,minp,maxp,tn,mint,maxt]=premnmx(p,x);
%BP 网络训练
net=newff([-1,1],[5,1],{'tansig','tansig','purelin'},'traingdx');
net.trainParam.show=1000; %每 1000 轮回显示一次结果
net.trainParam.Lr=0.05; %学习速率为 0.05
net.trainParam.epochs=5000; %循环 5000 次
net.trainParam.goal=1e-6; %均方误差
net=train(net,pn,tn);    %训练
a=sim(net,pn);           %神经网络仿真
a=postmnmx(an,mint,maxt); %还原仿真得到的数据，反归一化
%对新数据进行预测
xx=1:length(x);
pnew=[length(xx):length(xx)+n];%预测数据
pnewn=tramnmx(pnew,minp,maxp);%新数据归一化
anewn=sim(net,pnewn);
anew=postmnmx(anewn,mint,maxt);%还原得到预测值
anew=anew(2);
```