

# 新零售无人智能售货机商务数据分析

# 目录

<b>一、背景介绍</b>	<b>3</b>
1、行业背景	3
2、分析目标	3
3、数据说明	4
<b>二、项目任务</b>	<b>4</b>
项目要求	4
<b>三、数据预处理与统计分析</b>	<b>4</b>
任务 1.1	4
任务 1.2	5
任务 1.3	6
<b>四、数据的可视化</b>	<b>6</b>
任务 2.1	6
任务 2.2	8
任务 2.3	13
任务 2.4	13
任务 2.5	13
<b>五、地区画像</b>	<b>14</b>
任务 3.1	14
任务 3.2	19
<b>六、销售额的预测</b>	<b>23</b>
任务 4.1	23
任务 4.2	23

# 一、背景介绍

## 1、行业背景

新零售，即个人、企业以互联网为依托，通过运用大数据、人工智能等先进技术手段并运用心理学知识，对商品的生产、流通与销售过程进行升级改造，进而重塑业态结构与生态圈，并对线上服务、线下体验以及现代物流进行深度融合的零售新模式。

新零售行业在我国发展时间较短，2016 年新零售概念提出，2017 年进入新零售元年，新零售市场规模为 389 亿元。随着用户习惯的养成及新零售模式的创新，未来增长潜力巨大，预计 2022 年整个市场规模有望突破 18000 亿元。

其中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，为用户共精准贴心的服务，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

本次选取了某商场在不同地点摆放的 5 台自动售货机，编号分别为 A、B、C、D、E，以此作为新零售无人智能售货机商务数据分析的样本代表，对该商场自动售货机获取数据从两个方面进行分析：

- 1) 地区画像；
- 2) 销售额的预测。

### 3、数据说明

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。其中附件一包含了订单号、设备 ID、应付金额、实际金额、商品、支付时间、地点、状态、是否提现等数据。附件 2 将商品分成两个级别，首先大类划分为饮料类和非饮料类，二级类将商品划分为更细的类别如乳制品、饼干糕点和坚果炒货等。

## 二、项目任务

### 项目要求

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，分析每台售货机 2018 年 1 月商品销量的预测问题。

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。

## 三．数据预处理与统计分析

### 任务 1.1

首先根据附件一中的数据，提取每台售货机对应的销售数据，分别保存为“task1-1A.csv”、“task1-1B”等文件，方便后续对每一天售货机进行具体的对比分析。

首先进行数据预处理，从下图中我们可以看到附件一中一共有 70680 行数据，数据中不存在缺失值。

```
订单号      70680
设备ID      70680
应付金额    70680
实际金额    70680
商品        70680
支付时间    70680
地点        70680
状态        70680
提现        70680
dtype: int64
```

之后我们提取附件一中每台售货机对应的销售数据，得到如下列售货机 A 的示意图。

1		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
2	1	DD201708E43A6E078		3	3	40g双汇玉米肠	2017/1/1 1:33	A	已出货未退	已提现
3	16	DD201708E43A6E078		5.8	5.8	100g卫龙辣条	2017/1/1 19:29	A	已出货未退	已提现
4	17	DD201708E43A6E078		0.8	0.8	咪咪虾条	2017/1/1 19:33	A	已出货未退	已提现
5	20	DD201708E43A6E078		3	3	250ml维他奶	2017/1/1 19:55	A	已出货未退	已提现
6	22	DD201708E43A6E078		3.5	3.5	东鹏特饮	2017/1/1 20:18	A	已出货未退	已提现
7	24	DD201708E43A6E078		3	3	250ml维他奶	2017/1/1 20:25	A	已出货未退	已提现
8	25	DD201708E43A6E078		3	3	卫龙大面筋	2017/1/1 20:25	A	已出货未退	已提现
9	31	DD201708E43A6E078		3.5	3.5	东鹏特饮	2017/1/2 8:50	A	已出货未退	已提现
10	34	DD201708E43A6E078		1.5	1.5	卫龙亲嘴虾	2017/1/2 12:02	A	已出货未退	已提现
11	37	DD201705E43A6E078		4.5	4.5	阿萨姆奶茶	2017/1/2 13:34	A	已出货未退	已提现
12	43	DD201708E43A6E078		4	4	145ml旺仔牛奶	2017/1/2 13:58	A	已出货未退	已提现
13	49	DD201705E43A6E078		2	2	伊利优酸乳	2017/1/2 15:56	A	已出货未退	已提现
14	74	DD201708E43A6E078		4.5	4.5	2g韩国海牌	2017/1/2 22:42	A	已出货未退	已提现
15	75	DD201708E43A6E078		4	4	脉动	2017/1/2 22:42	A	已出货未退	已提现
16	85	DD201708E43A6E078		3	3	250ml维他奶	2017/1/3 9:50	A	已出货未退	已提现
17	92	DD201708E43A6E078		2	2	75g新麦潮	2017/1/3 14:57	A	已出货未退	已提现
18	94	DD201705E43A6E078		3	3	可口可乐	2017/1/3 15:10	A	已出货未退	已提现
19	103	DD201708E43A6E078		3	3	果粒橙	2017/1/3 15:44	A	已出货未退	已提现
20	105	DD201705E43A6E078		1	1	香豆干	2017/1/3 15:46	A	已出货未退	已提现

任务 1.2

计算得到每台售货机 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。

	A	B	C
1		2017年5月份交易额	2017年5月份订单量
2	A	3385.1	756
3	B	3681.2	869
4	C	3729.4	789
5	D	2392.1	564
6	E	5699	1292
7	交易总额	18886.8	
8	订单总量		4270

从结果可以看出，2017 年 5 月份售货机交易额差异不太大，除了个别如售货机 D 该月交易额较低，为 2392.1 元，售货机 E 该月交易额较高，为 5699 元，大部分售货机五月交易额都为 3500 元左右。同样的，2017 年 5 月份售货机订单量除了个别如售货机 D 该月订单量较低，为 564 单，售货机 E 该月订单量较高，为 1292 单，大部分售货机五月订单量都维持为 800 单左右。

### 任务 1.3

计算获得每台售货机每月的每单平均交易额与日均订单量。

	A	B	C	D	E	F	G	H	I	J	K	L
1		售货机A日	售货机A每	售货机B日	售货机B每	售货机C日	售货机C每	售货机D日	售货机D每	售货机E日	售货机E每	单平均交易额
2	一月份	5063	4.064033	6633	3.99204	6917	4.184965	4318	3.746572	11403	4.115119	
3	七月份	476	4.097689	345	4.401739	764	3.988351	317	4.229653	813	3.919311	
4	三月份	255	3.58549	265	3.614717	263	3.769962	192	4.305729	350	4.305714	
5	九月份	1040	4.307212	1745	4.130258	1678	4.427294	983	3.89939	4134	4.125375	
6	二月份	114	3.864035	185	3.255676	208	3.826923	141	3.088652	258	3.638372	
7	五月份	756	4.477646	869	4.236133	789	4.726743	564	4.241312	1292	4.410991	
8	八月份	666	3.358709	981	3.5842	1259	3.913582	715	3.316503	1767	3.804471	
9	六月份	1669	4.047394	1856	4.06805	1882	4.5017	1040	4.025962	2593	3.817856	
10	十一月份	1160	4.471552	2031	4.268784	1943	4.352393	1210	3.862314	5020	4.283227	
11	十二月份	2003	3.787868	2210	3.667014	2379	3.943043	1663	3.57258	3252	4.168973	
12	十月份	1565	4.020703	2026	4.11234	2216	4.27333	1186	3.884233	2777	3.676125	
13	四月份	447	4.036913	603	4.07529	734	4.403678	443	3.790293	895	4.159888	

结果显示，对于五台售货机，1 月每单平均交易金额与其他月份相差不大，但是一月份相较其他十一个月份，它的日均订单量高出很多，在 5000 到 6000 的范围内，约为五至六倍，售货机 E 一月份日均订单量更是高达 11403。其后 2 月和 3 月的每单平均交易金额差异不大，但是 2 月和 3 月的日均订单量迅速降低，这可能是 2 月和 3 月处于春节期间，大部分人都回家过年的原因。

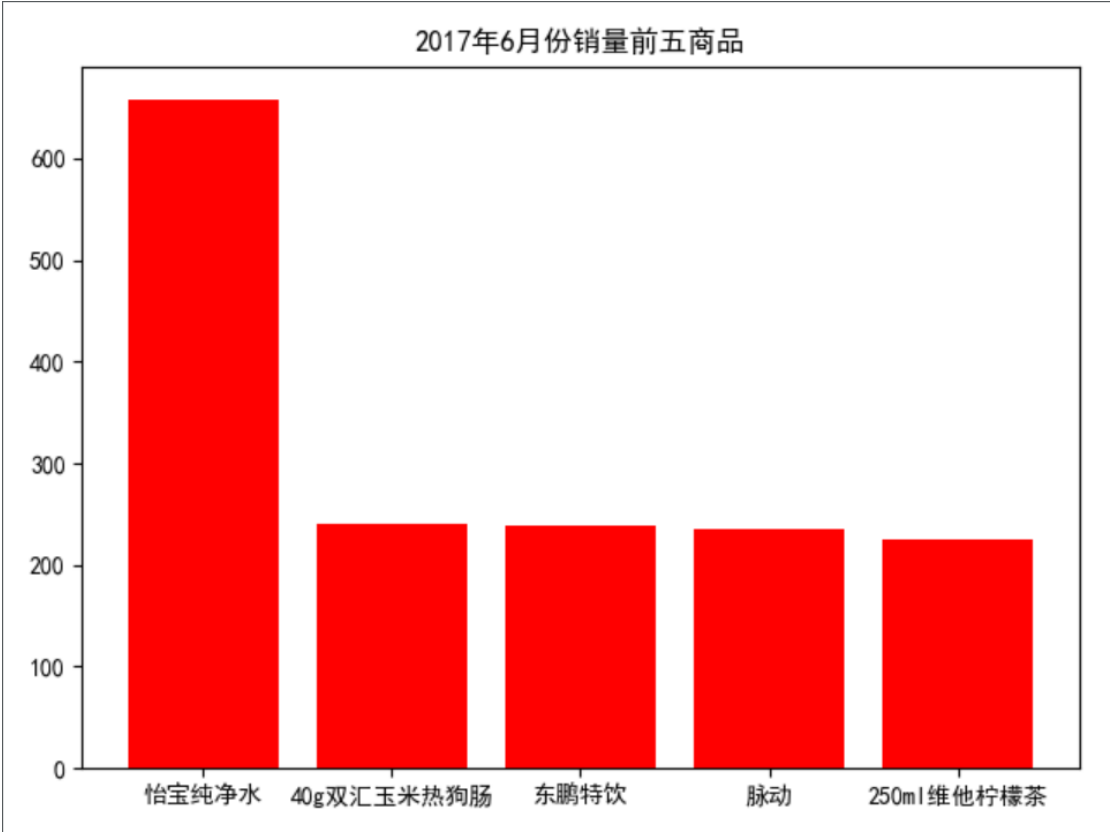
## 四、数据的可视化

### 任务 2.1

首先获取 2017 年 6 月份各商品销量数据，销量从大到小排行，如下图所示。

怡宝纯净水	657
40g双汇玉米热狗肠	240
东鹏特饮	238
脉动	235
250ml维他柠檬茶	225
30g无穷农场盐_鸡蛋	205
250ml维他原味豆奶	199
20g咪咪虾条马来西亚风味	190
营养快线	179
阿萨姆奶茶	151
统一冰红茶	138
王老吉（罐）	127
145ml旺仔牛奶盒装	119
250ml燕塘甜牛奶	113
鸭翅	112
85g新麦潮北海道吐司	112
伊利优酸乳	111
250ml燕塘红枣牛奶	109
86g合味道海鲜风味	103

绘制 2017 年 6 月销量前 5 的商品销量柱状图

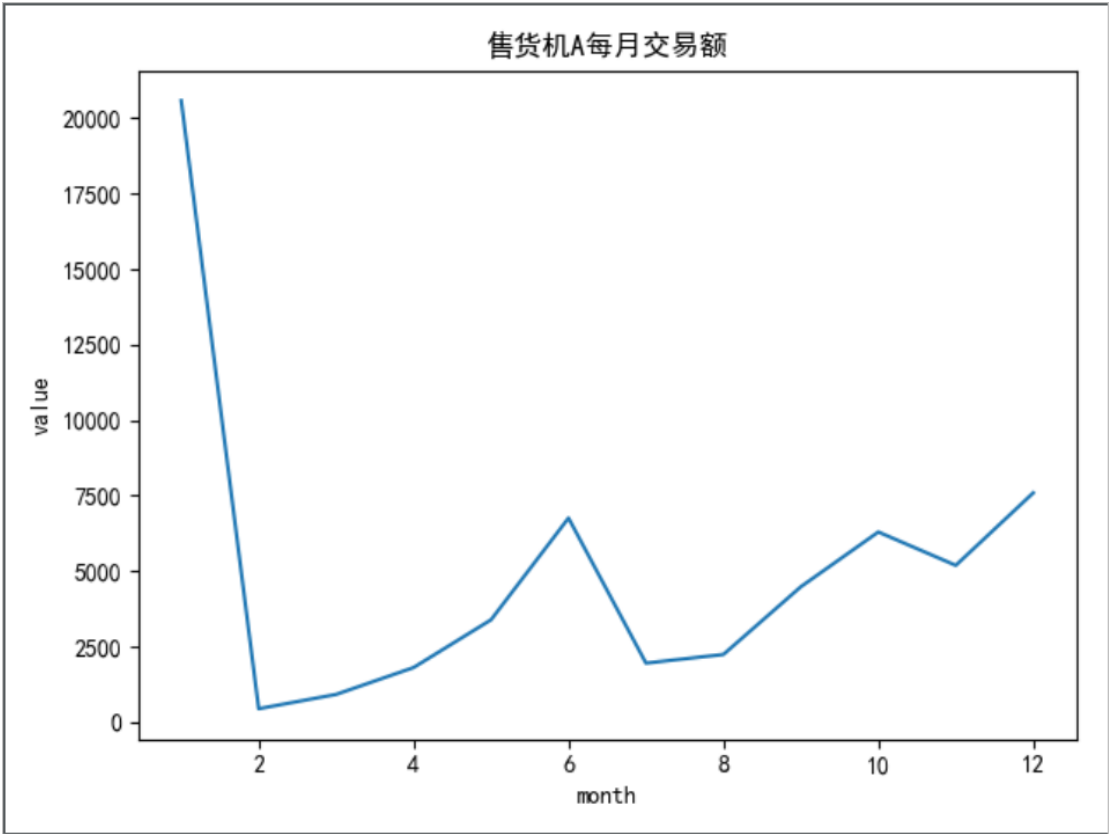


从上图可以得知，怡宝纯净水销量最多，其次是 40g 双汇玉米热狗肠、东鹏特饮、脉动和 250ml 维他柠檬茶。其中怡宝纯净水销量高达其他前四名销量商品

的 3 倍。

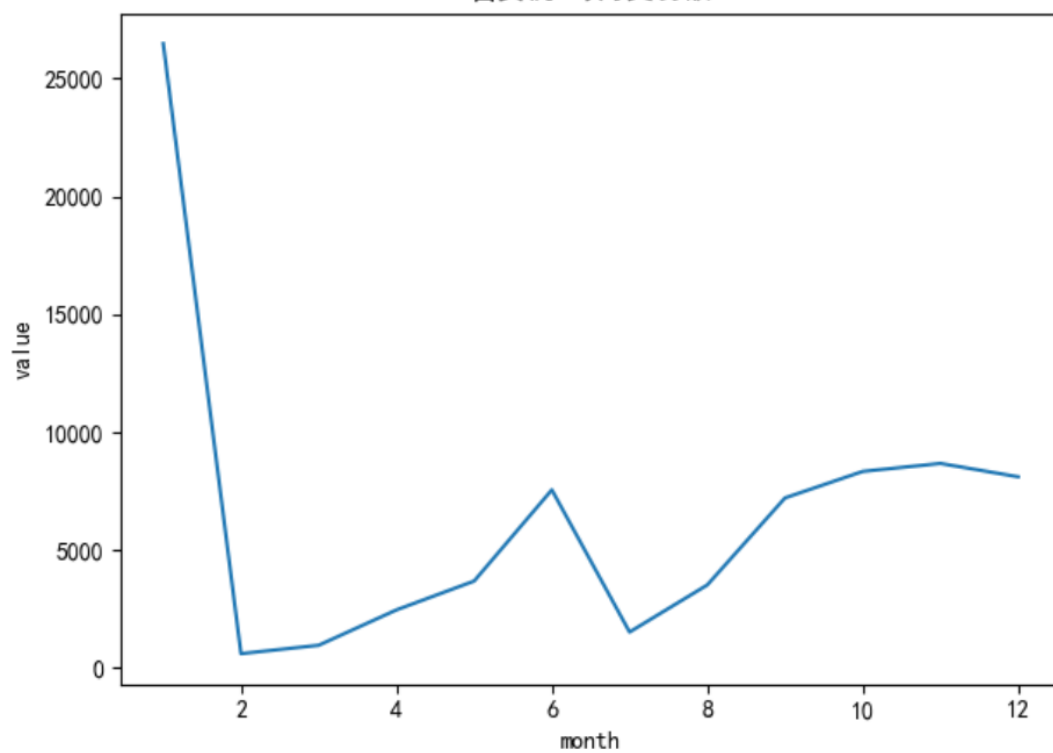
## 任务 2.2

绘制每台售货机每月总交易额折线图及交易额月环比增长率柱状图，如下图所示。可以得知，所有售货机都是一月份交易额为全年最高，2 月份最低，3 月份其次较低。可能是 1 月份人们有准备回家的需求，提前储备好物品和出外游玩的缘故。而 2 月和 3 月处于春节期间，大部分人都回家过年，所以交易额处于全年最低区间。随着 2、3 月份交易额的大幅回落，后面除了 7 月，其余全年每个月份的交易额态势处于逐步增长阶段。具体从下图的交易额月环比增长率也可得知，除了 2、7 月份，其余都较上月呈现良好态势。

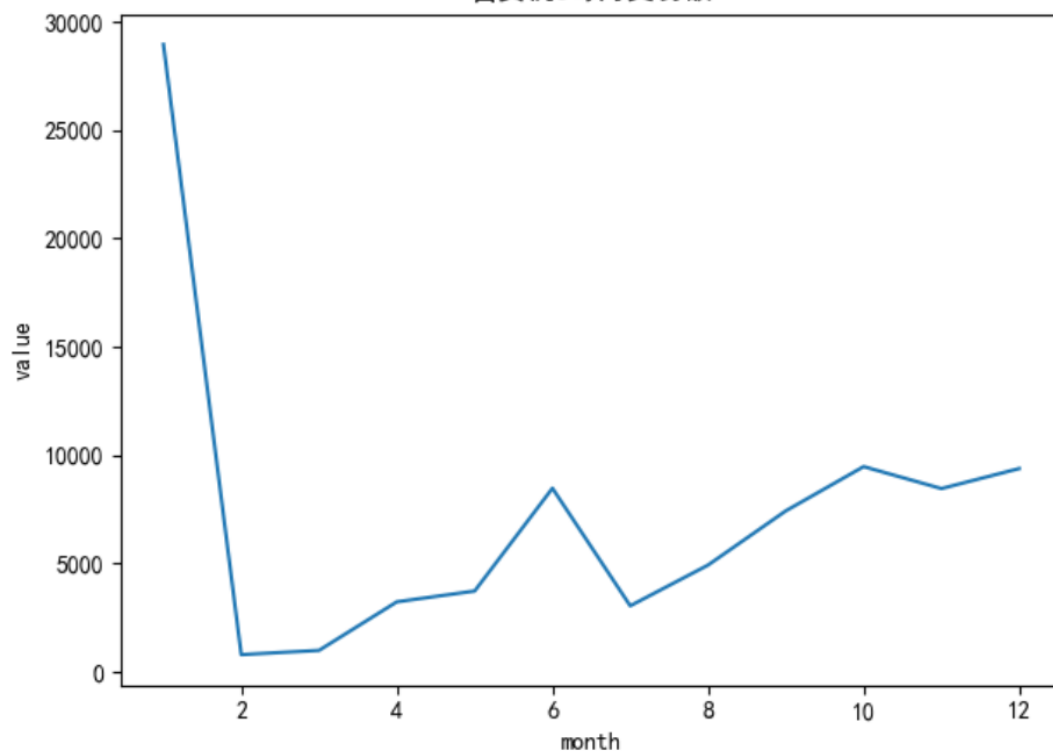




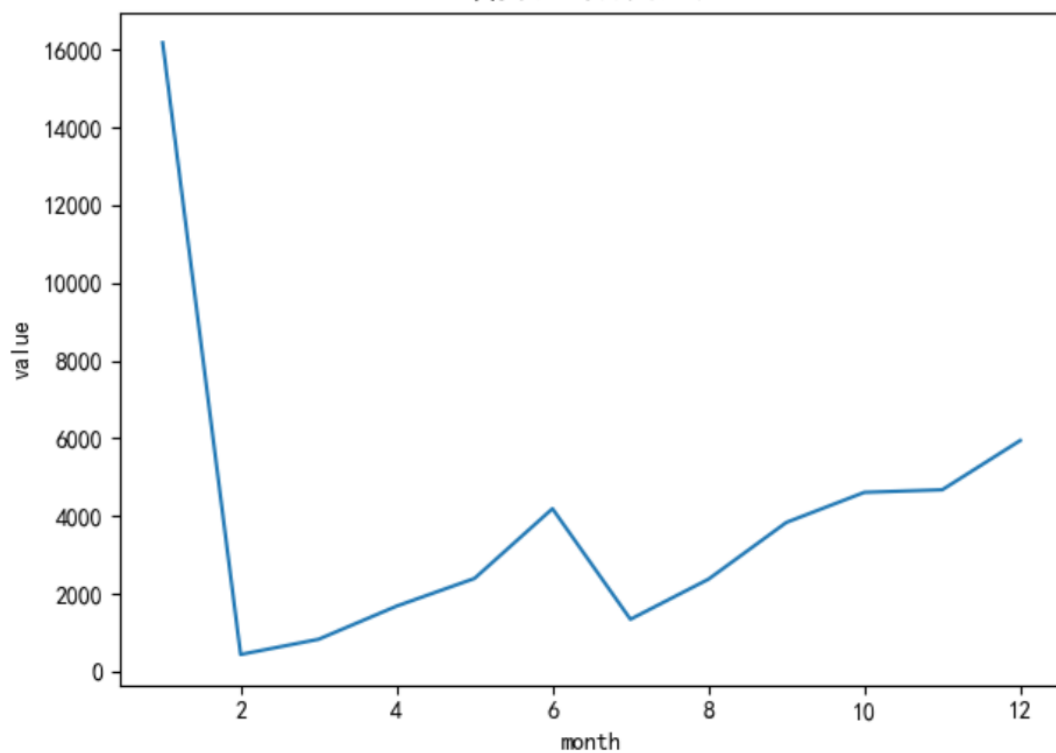
售货机B每月交易额



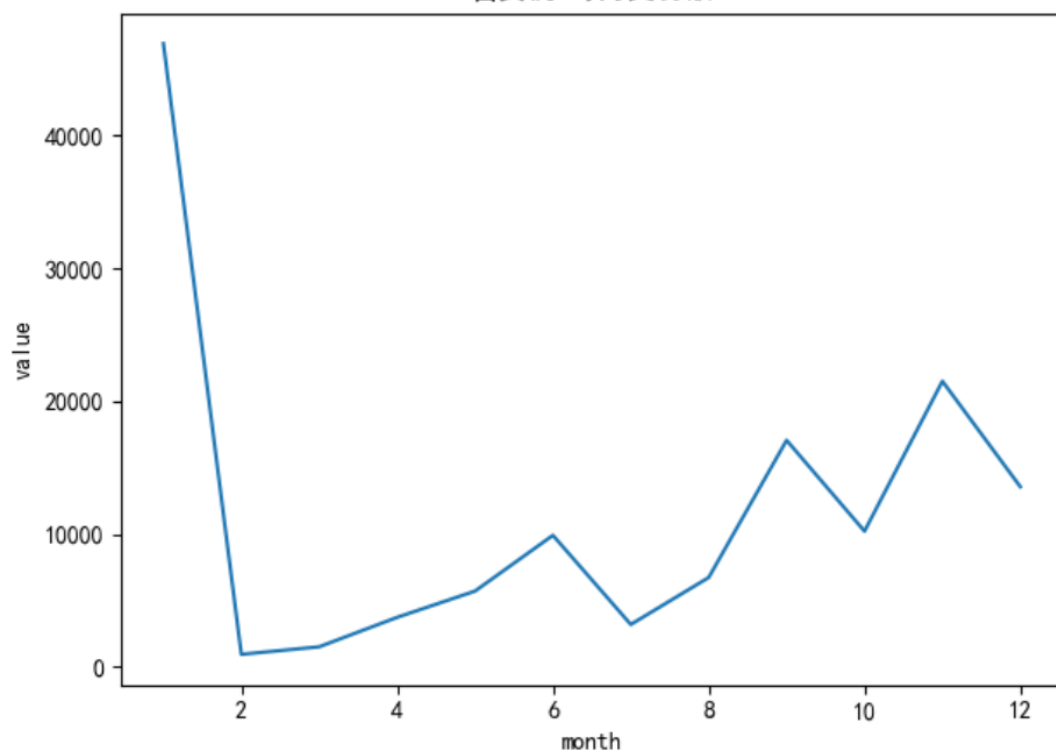
售货机C每月交易额



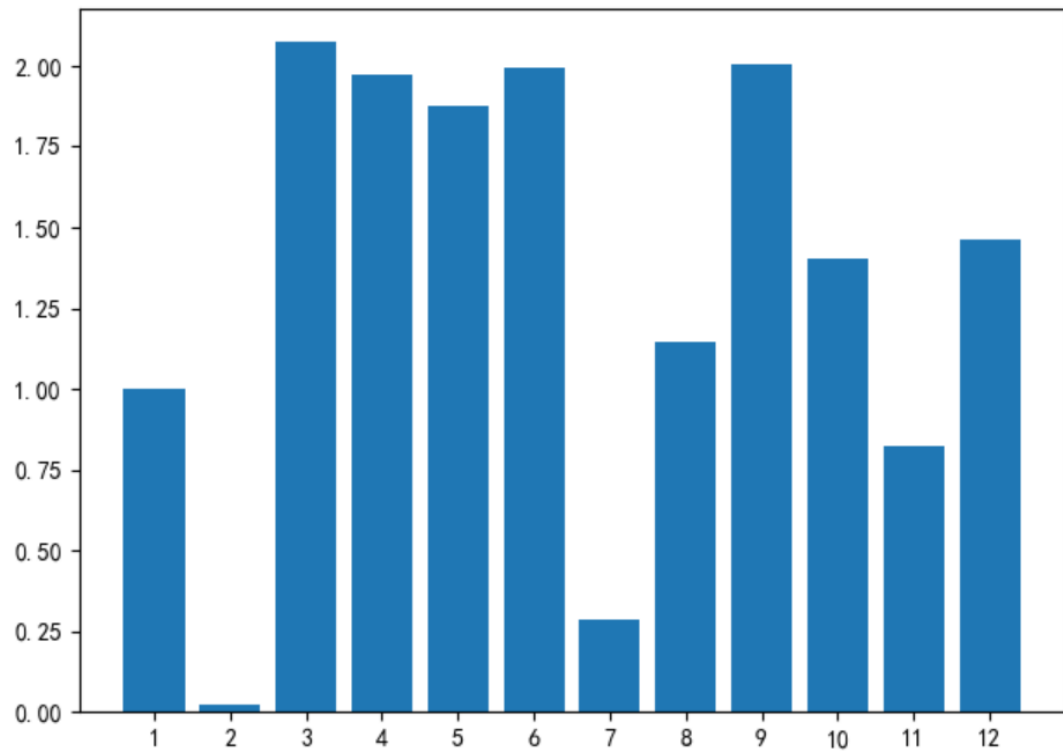
售货机D每月交易额



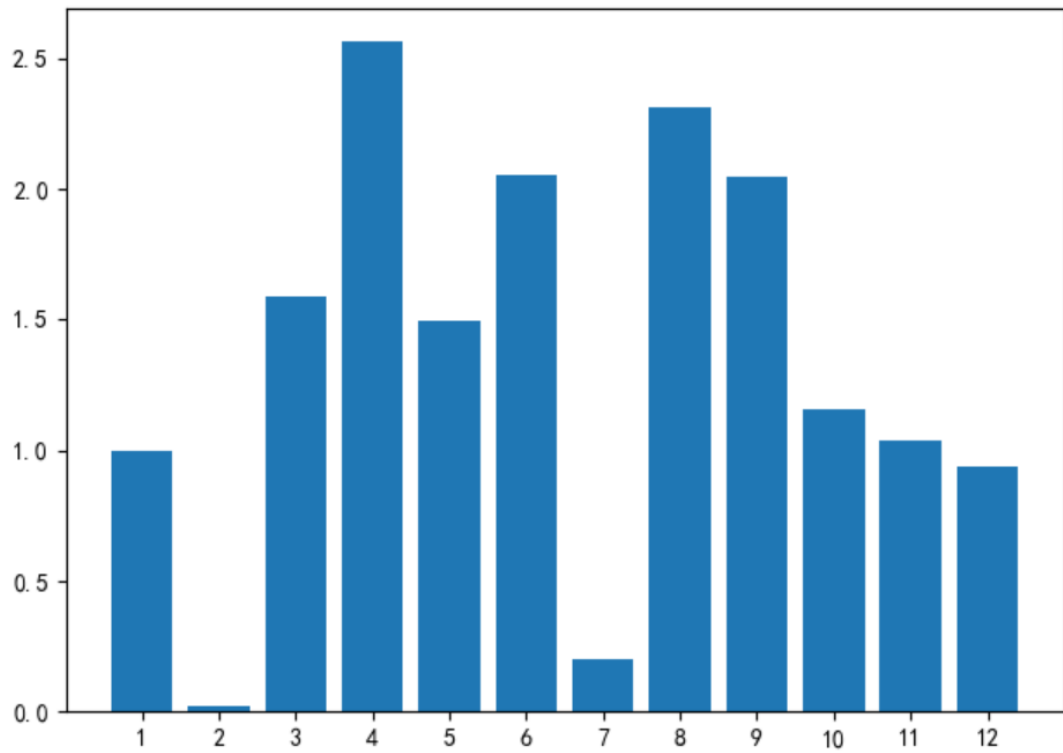
售货机E每月交易额



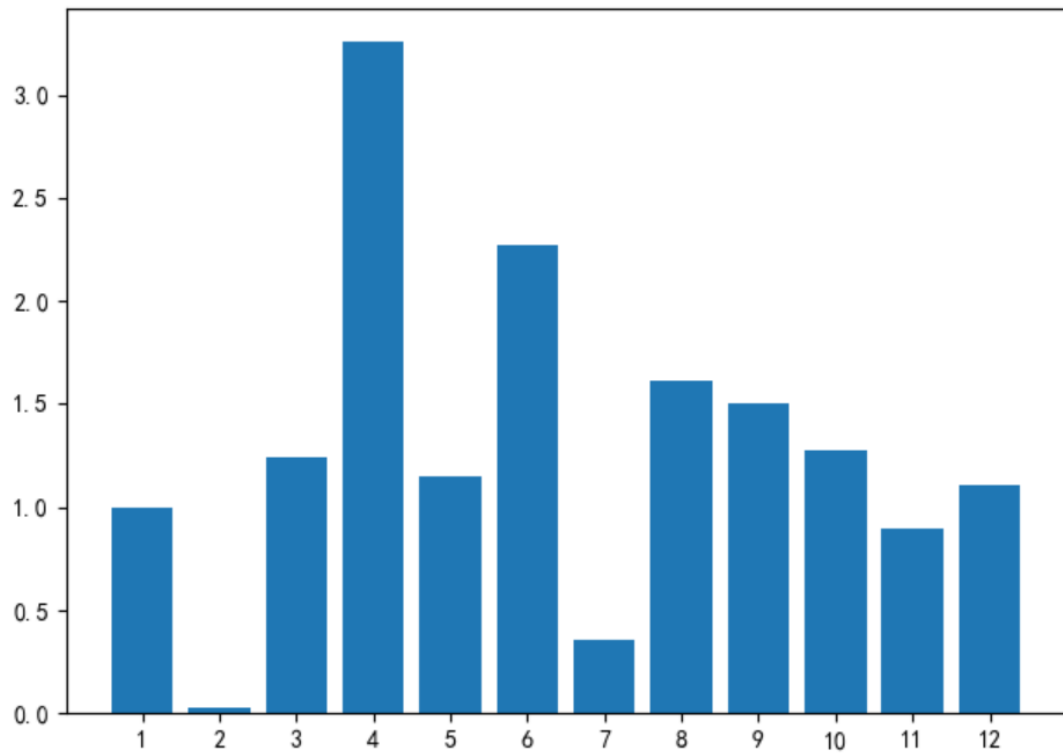
售货机A交易额月环比增长率



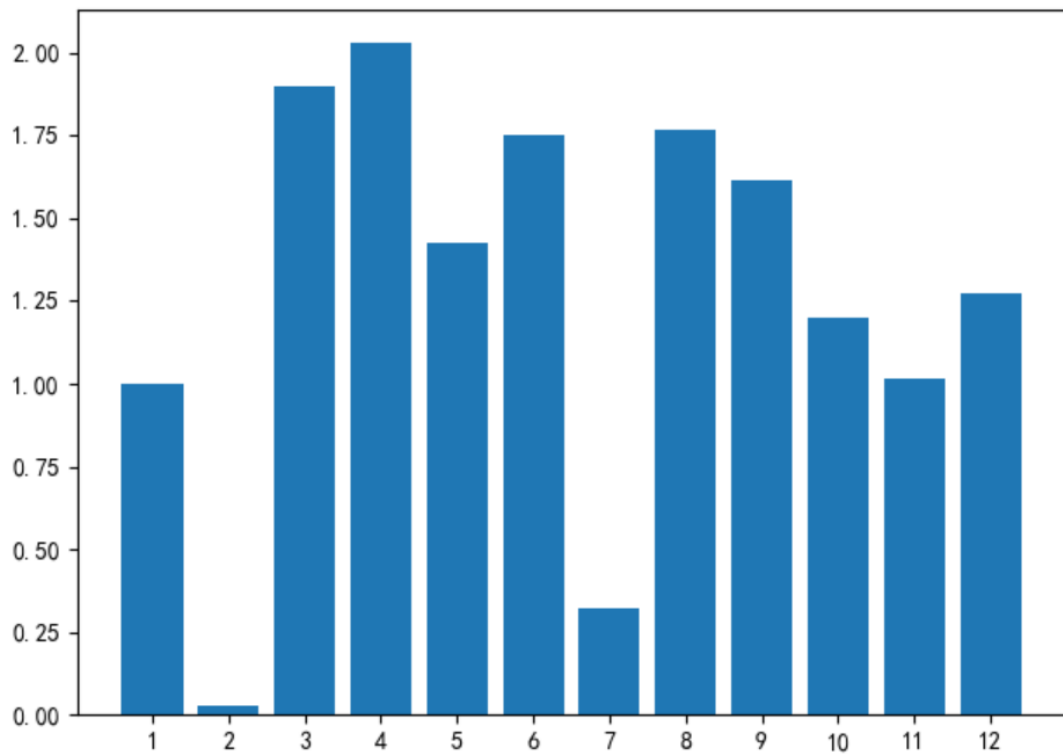
售货机B交易额月环比增长率

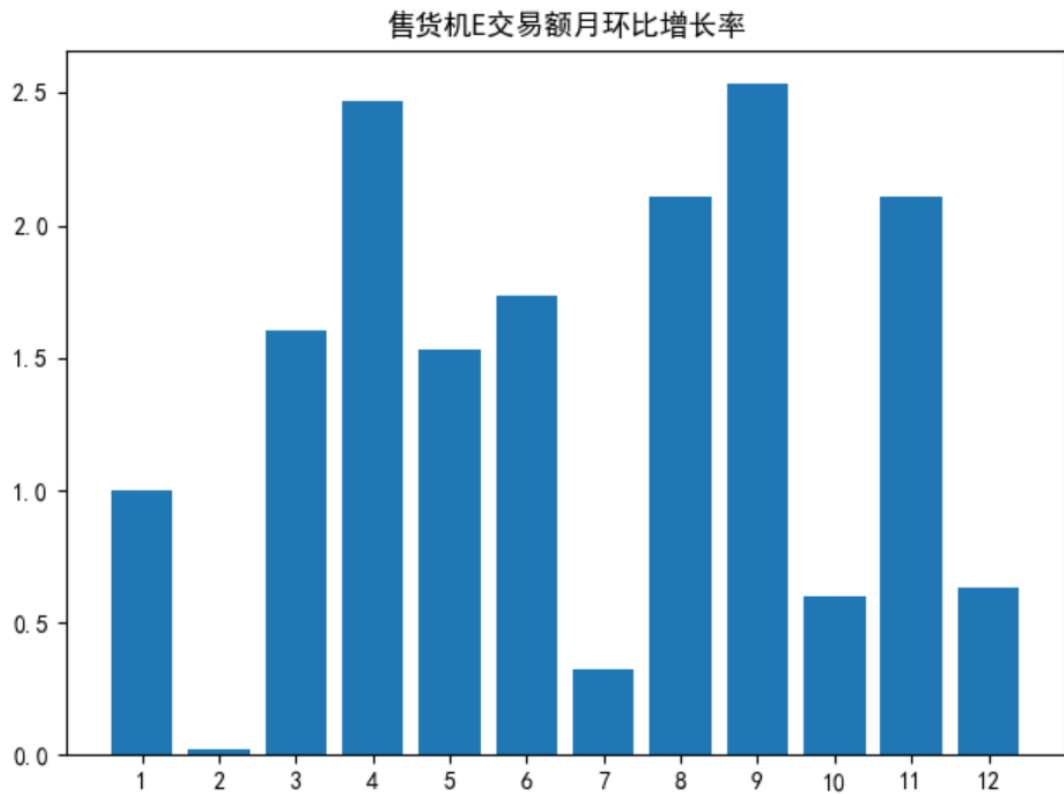


售货机C交易额月环比增长率



售货机D交易额月环比增长率





### 任务 2.3

绘制每台售货机毛利润占总毛利润比例的饼图（假设饮料类毛利率为 25%，非饮料类为 20%）

### 任务 2.4

绘制每月交易额均值气泡图，横轴为时间，纵轴为商品的二级类目

### 任务 2.5

绘制售货机（6、7、8 三个月订单量的热力图，横轴以天为单位，纵轴以小时为单位从热力图，可以分析得出哪些结论？

## 五、地区画像

### 任务 3.1

首先获取商品大类里面的饮料类商品列表，如下图所示。

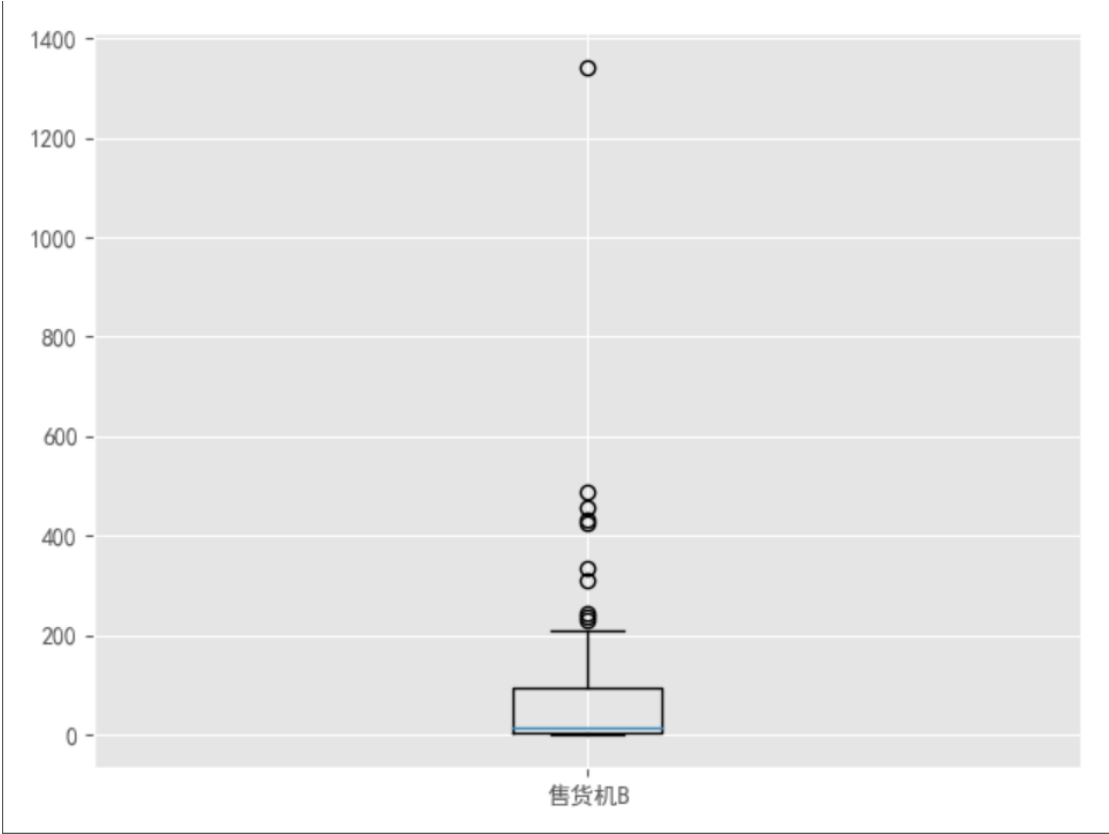
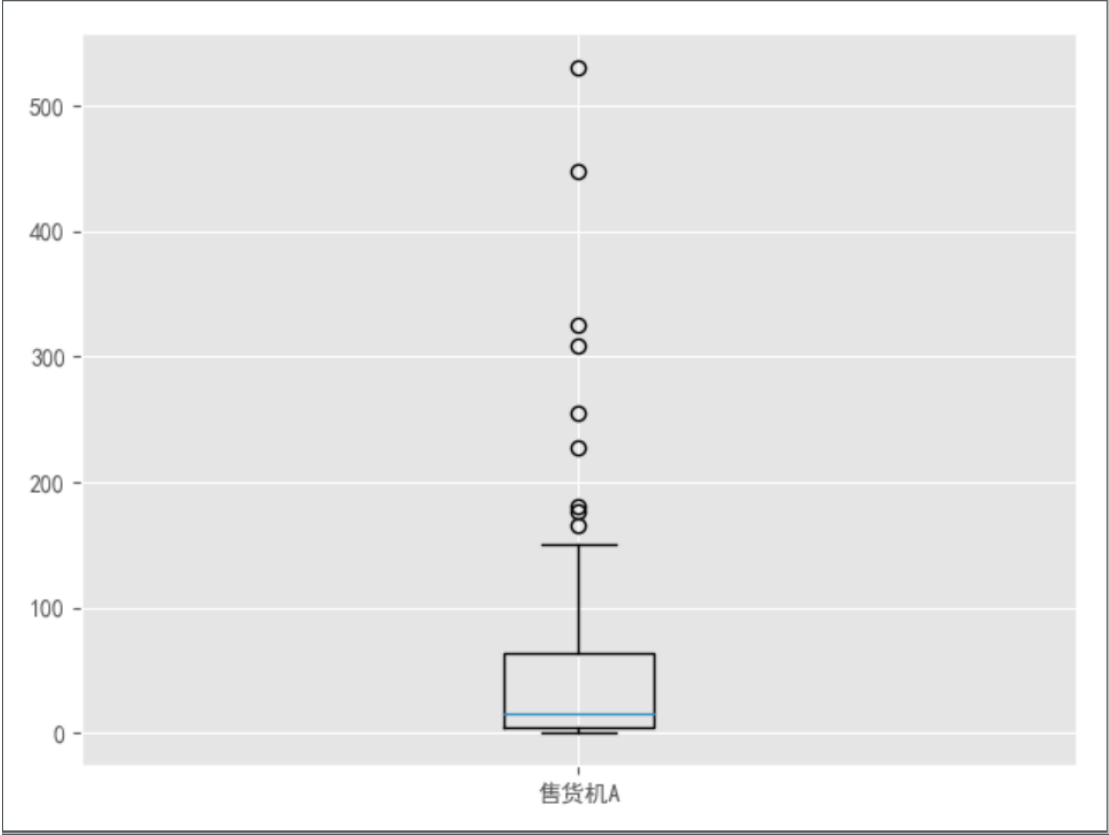
	A	B	C	D
1		商品	大类	二级类
2	0	100g*5瓶益力多	饮料	乳制品
3	9	145ml旺仔牛奶罐装	饮料	乳制品
4	10	145ml旺仔牛奶盒装	饮料	乳制品
5	12	150g健能酸奶原味	饮料	乳制品
6	23	250ML东鹏特饮	饮料	功能饮料
7	24	250ml红牛	饮料	功能饮料
8	25	250ml维他奶巧克力味	饮料	植物蛋白
9	26	250ml维他柠檬茶	饮料	茶饮料
10	27	250ml维他原味豆奶	饮料	植物蛋白
11	28	250ml燕塘红枣牛奶	饮料	乳制品
12	29	250ml燕塘甜牛奶	饮料	乳制品
13	30	250ml燕塘原味酸奶	饮料	乳制品
14	31	268ml雀巢咖啡丝滑拿铁	饮料	咖啡
15	35	330ml伊利畅意乳酸菌原味	饮料	乳制品
16	36	340ml蒙牛优益C乳酸菌	饮料	乳制品
17	44	450ml美汁源果粒橙	饮料	果蔬饮料
18	45	480ml小茗同学冷泡溜溜哒茶	饮料	茶饮料
19	46	480ml小茗同学冷泡青柠红茶	饮料	茶饮料
20	47	500ML宝矿力	饮料	功能饮料

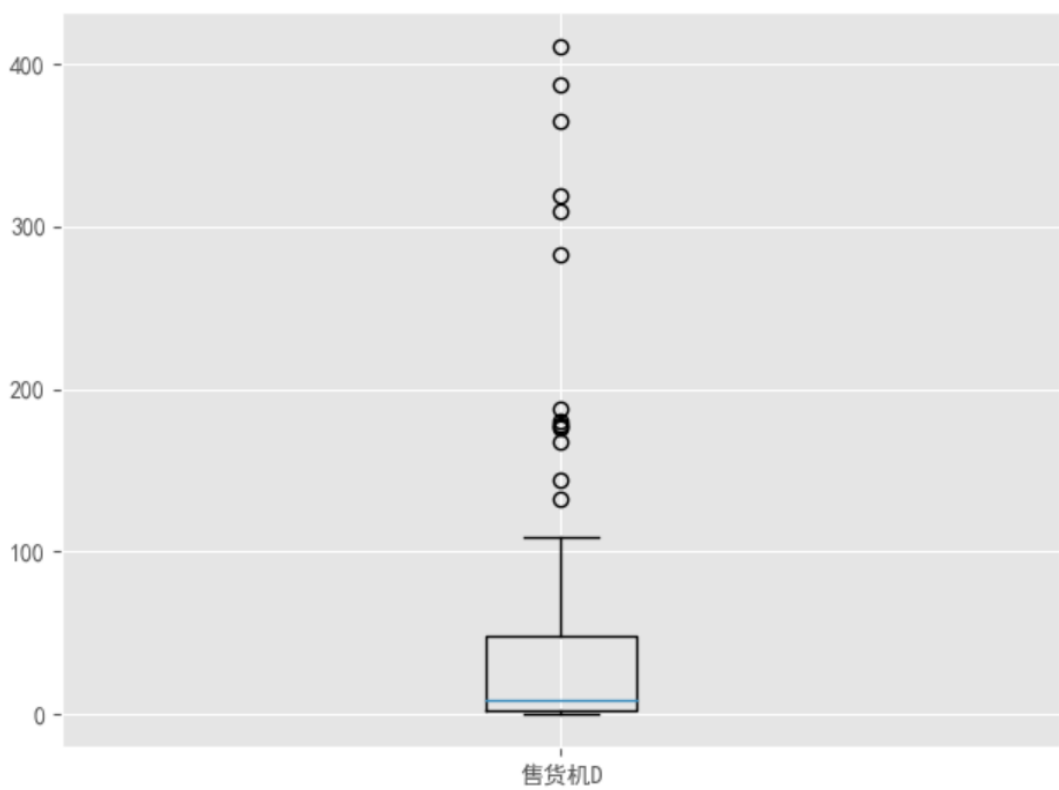
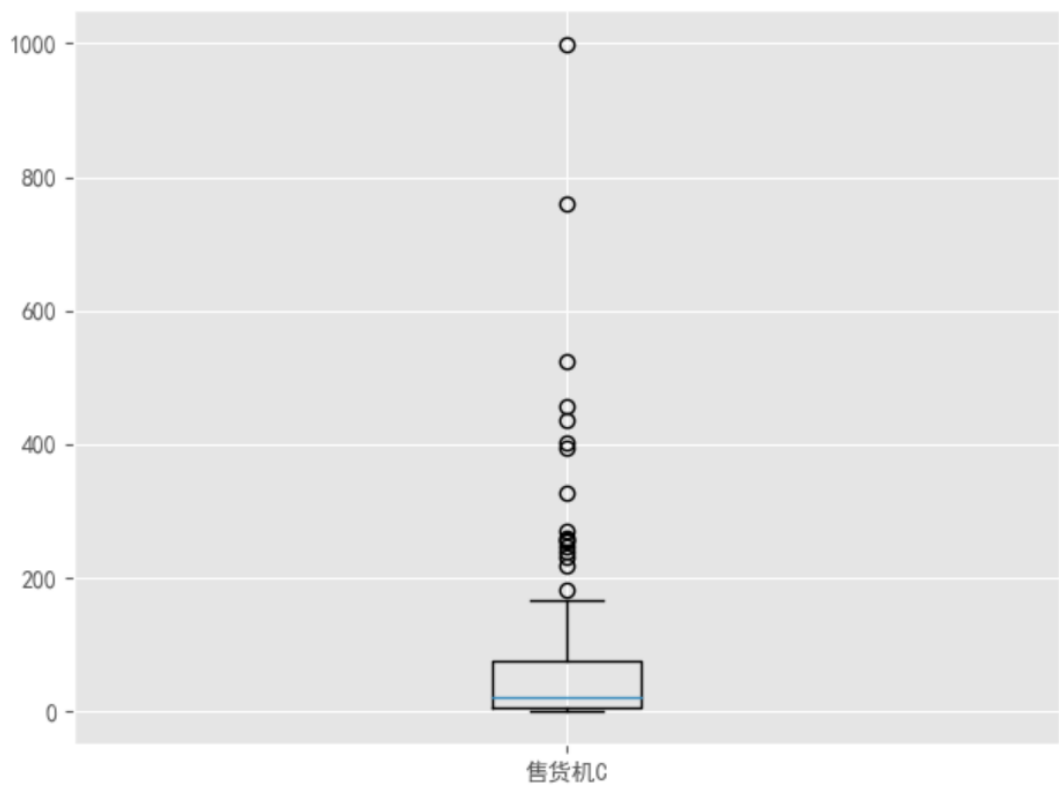
然后将各个售货机的各饮料商品销售情况汇总得出，其中有些商品销售数量为零，要进行缺失值处理。结果如下图所示。

	A	B
1		售货机A
2	100g*5瓶益力多	40
3	13g雀巢咖啡1+2特浓	4
4	145ml旺仔牛奶盒装	131
5	145ml旺仔牛奶罐装	2
6	150g健能酸奶原味	28
7	180ml雀巢咖啡罐装	22
8	200ml汇源橙汁100%	0
9	205g伊利亚慕希原味	53
10	205ml安慕希蓝莓味	7
11	250ML东鹏特饮	16
12	250ml燕塘原味酸奶	112
13	250ml燕塘甜牛奶	126
14	250ml燕塘红枣牛奶	128
15	250ml王老吉盒装	22

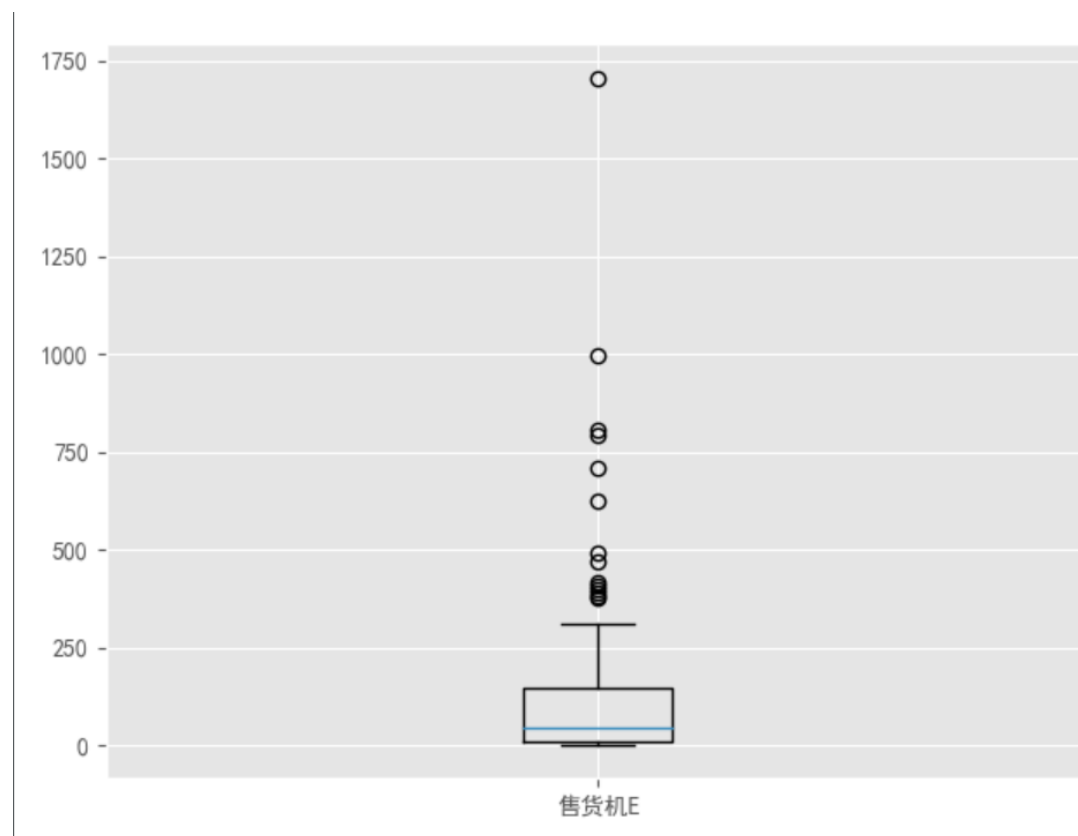
为给每台售货机各饮料类商品的分别贴上热销、正常、滞销的标签，需要根据总体的销售情况制定一个标准。

所以我们首先根据各个售货机各商品的销售数据分别绘制箱型图，显示数据分布情况资料，具体结果如下。









```

A
count 119.000000
mean  50.966387
std   85.217064
min   0.000000
25%   4.000000
50%   16.000000
75%   63.500000
max   531.000000

B
count 119.000000
mean  74.067227
std   154.895412
min   0.000000
25%   4.000000
50%   15.000000
75%   93.000000
max   1342.000000

C
count 119.000000
mean  80.563025
std   150.256822
min   0.000000
25%   5.000000
50%   22.000000
75%   74.500000
max   999.000000

D
count 119.000000
mean  45.848739
std   83.626048
min   0.000000
25%   2.000000
50%   9.000000
75%   48.500000
max   411.000000

E
count 119.000000
mean  132.084034
std   233.314677
min   0.000000
25%   9.000000
50%   44.000000
75%   147.000000
max   1705.000000

```

同时具体数据分布情况如上。我们可以从中得知每个售货机的下四分位数、中位数、上四分位数、上限、下限、极上限和极下限。

我们选取上四分位数作为区分热销和正常的标准，同时选取下四分位数作为

区分滞销和正常的标准，销量大于上四分位的为热销，小于下四分位数的为滞销，小于上四分位数且大于下四分位数的为正常。并根据此给出每台售货机饮料类商品的标签，分别保存为“task3-1A”，“task3-1B”等文件，结果见下图。

	A	B
1		标签
2	100g*5瓶益力多	正常
3	13g雀巢咖啡1+2特浓	正常
4	145ml旺仔牛奶盒装	热销
5	145ml旺仔牛奶罐装	滞销
6	150g健能酸奶原味	正常
7	180ml雀巢咖啡罐装	正常
8	200ml汇源橙汁100%	滞销
9	205g伊利安慕希原味	正常
10	205ml安慕希蓝莓味	正常
11	250ML东鹏特饮	正常
12	250ml燕塘原味酸奶	热销
13	250ml燕塘甜牛奶	热销
14	250ml燕塘红枣牛奶	热销
15	250ml王老吉盒装	正常

### 任务 3.2

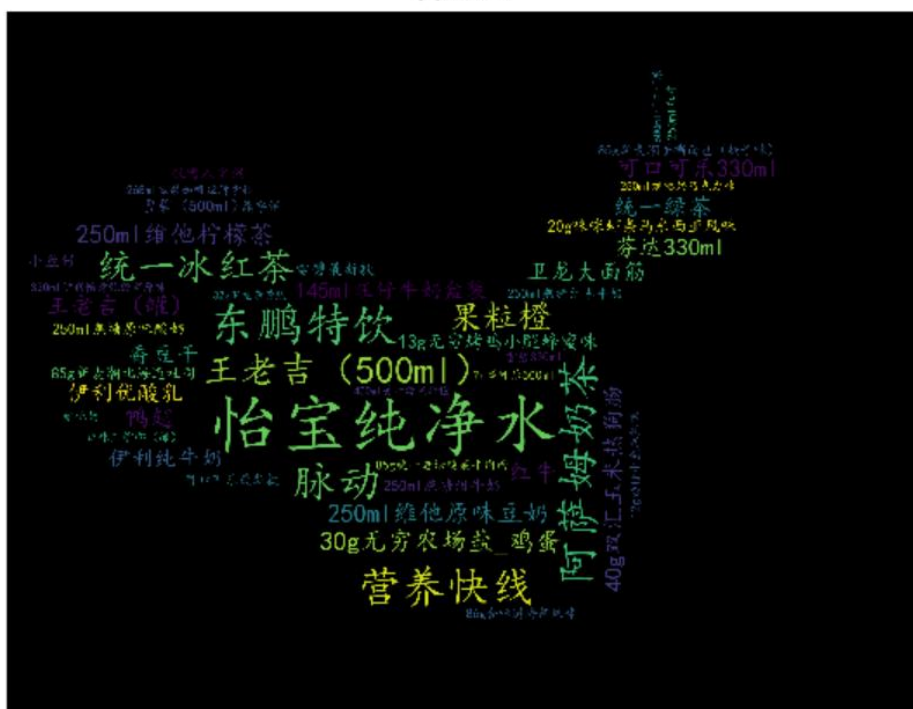
在上一个小任务的基础上，用同样的方法对商品进行划分标签分别“热销”、“正常”、“滞销”，得到售货机所有商品的标签情况，分别保存为“task3-2A.csv”、“task3-2B.csv”、“task3-2E.csv”文件。同时根据商品的销量情况作为依据，生成每台售货机的词云画像，其中没有销售过的商品没出现在附件一中，销量情况进行缺失值处理为零。

根据画像可以发现，所有售货机的主要销售商品都为怡宝纯净水，其中售货机 A 还主要销售东鹏特饮、阿萨姆奶茶、脉动、营养快线、13g 无穷烤鸡小腿蜂蜜味和 30g 农村盐\_鸡蛋，几乎都为饮料类商品。售货机 B 除了主要销售东鹏特饮、阿萨姆奶茶、脉动、营养快线和 30g 农村盐\_鸡蛋外，还有王老吉（500ml）、250ml 维他原味豆奶、统一冰红茶和果粒橙，也主要为饮料类。售货机 C 除了主要销售东鹏特饮、阿萨姆奶茶、脉动、营养快线、王老吉（500ml）、统一冰红茶和果粒橙和 30g 农村盐\_鸡蛋外，还有 250ml 维他柠檬茶、可口可乐 330ml、红牛、伊利优酸乳和卫龙大面筋。售货机 C 相较于售货机 A、B，销售的东西较多元

销售建议：售货机 A 和 E 可以配置主要几款饮料如怡宝纯净水、东鹏特饮、阿萨姆奶茶、脉动、营养快线和 250ml 维他柠檬茶，配置重心放在上述几款饮料商品。售货机 D 也是如售货机 A 和 E 一般配置，但可以增添多一款卫龙大面筋售卖。售货机 B 和 C 配置重心也是放在饮料类商品，但是售卖的饮料类可以多元化一些。

[illegible]

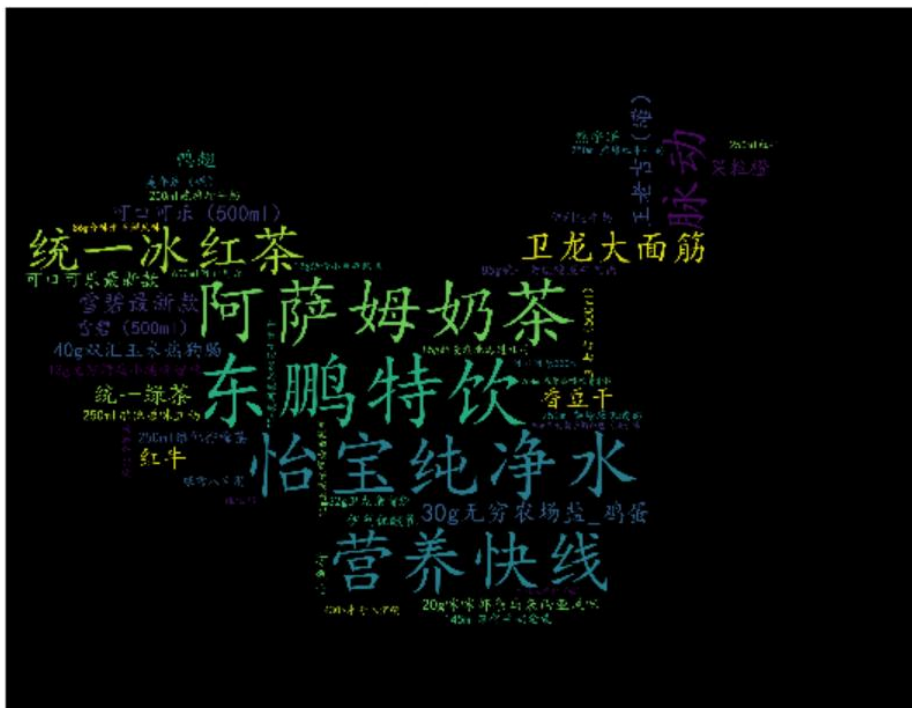
售货机B



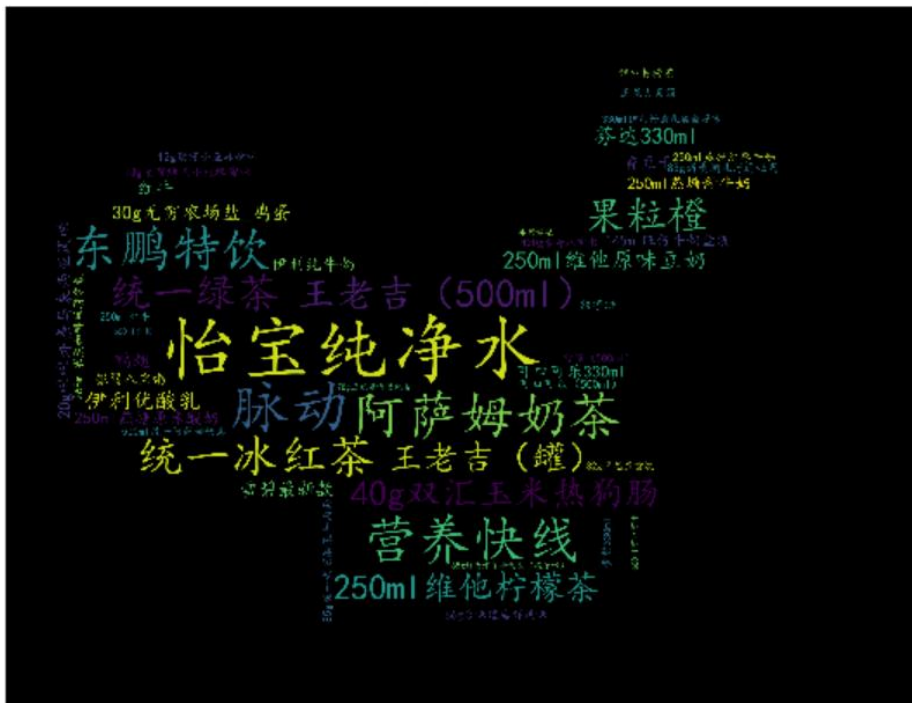
售货机C



售货机D



售货机E



## 六、销售额的预测

2、如果你们认为可以进行预测，请给出过程及结果，如果认为附件所给数据不足以支持预测工作，应该向经营者提出哪些数据方面的要求？

### 任务 4.1

预测未来销售额的原理是虽然未来不可预测，但是基于预计未来人流量、人群购买比例、购买商品种类及其每单平均销售额与过往的相差不多，于是乎可线性或非线性外推到未来，预测未来销售额，本质上属于归纳推理的一种。

### 任务 4.2

具体上不能根据附件提供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，因为根据上述数据可以发现，1 月份每台售货机的交易额都大幅超于其他 11 个月份，其中它的一月份每单平均交易额与其他月份无太大差别，主要是日均订单量较高，于是乎可推测是由于一月份人流量较大，而人群购买比例并没提高，结果使得 2017 年一月份的交易额相对其他月份高出许多。

同样的，也有可能是由于人们有在二月份回家过年的需求，进而人流量并无提高，只是提高了人们购买的欲望才使得每台售货机在 2017 年的售货交易额位于较高水平。

此外也有可能上述提到的方面同时存在，即一月份的交易额相对其他月份高出许多由于一月份人流量较大且人们购买需求同时增大。

同时 2018 年人们购买商品种类及其每单平均销售额相对 2017 年有无变化这方面也有待考究，因为过去数据只有 2017 年这一年份，没有太多可回溯时间资料，不能从单一截面数据进而开始线性外推预测未来年份人们购买商品种类及其每单平均销售额。

除了上述所提及的，还有很重要的一点因素就是根据任务三的自动售货机画像，我们对此已给出总结描述与营销建议，在营销建议的基础之上，未来售货机

的购买商品种类及其每单平均销售额有可能较过往 2017 年发生较大变化，人群购买比例可能也因营销建议的改变而发生变化。

另外商品具体存货和商品补货周期对售货机交易额预测也存在一定的影响。

综上所述，我们不能根据附件提供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。在数据方面，经营者也应尽可能多提供些其他售货机过往年份销售数据、该商场人流量，人群购买比例、人们购买商品种类、每单平均销售额及其存货和商品补货周期以作为预测参考。