
自动售货机商务数据分析与应用

摘要

自动售货机以线上经营的理念，给予顾客线下接触实惠高品质商品的便利服务，是零售经营的又一大突破。自动售货机的供给频率，种类选择，供给量，站点选择等都是自动售货机盈利的重要关键因素。本文就 2017 年 A、B、C、D、E 地的自动售货机的数据进行分析，呈现不同地方不同时间的自动售货机的各项指标，为自动售货机的种类选择，营销策略，站点选择等提供一定的帮助。

针对任务一，主要利用 python 进行数据预处理和统计分析，先对数据进行一定的预处理分析后再简单的进行部分指标的统计分析，以表格形式呈现并分析。

针对任务二，主要利用 python 和 Excel 相结合，生成各种直观的图形并加以分析，给予了更多对种类和站点选择等有帮助的分析建议。

针对任务三，利用 python 和 Excel 相结合，生成直观的自动售货机画像即词云图，使得能够直观看出各个地方顾客对商品的喜好与选择，对种类选择方面有很大帮助。

针对任务四，利用灰色预测的方法进行对 2018 年 1 月的预测，尽管方法可行，但数据来源不够，需要在提供相关数据就能加以分析。

关键词：python, 灰色预测, Excel, 种类选择, 站点选择

Abstract

It is another breakthrough in retail operation that vending machines provide convenience services for customers to contact affordable and high-quality goods offline with the concept of online operation. The supply frequency, type selection, supply quantity and site selection of vending machines are all important factors for vending machines to make profits. In this paper, the data of vending machines in A, B, C, D and E in 2017 are analyzed, and various indexes of vending machines in different places and at different times are presented, which will provide some help for the selection of vending machines, marketing strategies and site selection.

For Task 1, Python is mainly used for data preprocessing and statistical analysis. Firstly, some data are pre-processed and analyzed, then some indexes are simply

analyzed and presented in tabular form.

For Task 2, we mainly use Python and Excel to generate various intuitive graphics and analyze them, and give more helpful analysis suggestions for types and site selection.

For Task 3, the use of Python and Excel combined to generate intuitive vending machine portraits, that is, word clouds, so that we can intuitively see the preferences and choices of customers in various places, which is very helpful in the selection of types.

For Task 4, the grey forecasting method is used to forecast January 2018. Although the method is feasible, the data sources are not enough, and it needs to be analyzed by providing relevant data.

Key words: python, grey prediction, Excel, category selection, site selection

目录

一、背景介绍.....	4
1.行业背景.....	4
2.分析目标.....	4
二、数据预处理.....	4
1.缺失值处理.....	4
2.重复值处理.....	5
3. 数据标准化处理.....	5
4.异常值处理.....	5
三、统计分析.....	6
1.预处理命名文件.....	6
2. 每台售货机 2017 年 5 月交易总额和订单量和全年的交易总额和订单量	6
3. 每台自动售货机每月的每单平均交易额和日订单量	8
4. 月均销售次数和月均销售金额.....	10
5.总体分析.....	11
四、数据可视化.....	11
1. 绘制销售量前十和销售量最后二十位的柱状图	11
2.分别绘制五台机器的销量前十和销量后十	12
3.绘制 2017 年 6 月份销售前十柱状图.....	15
4.绘制每台售货机每月总交易额折线图	15
5.绘制月环比增长率.....	20
6. 绘制每台售货机毛利润占总毛利润比例的饼图	23
7. 绘制每月交易额均值气泡图.....	24
8. 绘制售货机 C 6、7、8 月三个月订单量的热力图	25
五、生成各地自动售货机画像.....	26
1.数据预处理.....	26
2.生成自动售货机画像.....	27
六、业务预测.....	31
1. 灰色预测模型理论.....	31
2.灰色系统的特点.....	31
3.灰色生成.....	31
4. 累加生成简介.....	32
5. GM（11）模型.....	32
6.预测值的求解.....	34
7. GM(1,1)模型精度检验	34
8.精度检验等级参照表.....	35
9.具体实践.....	35
七、小结.....	36

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、站点选择等都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，掌握商业需求量，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

(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。

二、数据预处理

1.缺失值处理

(通过对整个附件 1 的数据进行缺失值查询，结果如下，代码见附件)结果如下图，发现并没有缺失值，因此不用进行缺失值填补。)

		Total	Percent	
提现	0	提现	0	0.0
状态	0	状态	0	0.0
地点	0	地点	0	0.0
支付时间	0	支付时间	0	0.0
商品	0	商品	0	0.0
实际金额	0	实际金额	0	0.0
应付金额	0	应付金额	0	0.0
设备ID	0	设备ID	0	0.0
订单号	0	订单号	0	0.0
dtype: int64				

2.重复值处理

（通过代码发现没有重复值，总的的数据量总共是 70680 条记录，结果如下图）

```
print(a.duplicated().value_counts())
```

```
False      70680  
dtype: int64
```

3. 数据标准化处理

（将应付金额和实际金额标准化，运用的方法是离差标准化方法，即 Min-max 标准化，使得其能够进行比较，不过后期数据仍然使用原先数据，具体代码见附件）

标准化之前的数据是：

	应付金额	实际金额
0	4.5	4.5
1	3.0	3.0
2	5.5	5.5
3	5.0	5.0
4	3.0	3.0

标准化之后的数据是：

	应付金额	实际金额
0	0.036	0.036
1	0.024	0.024
2	0.044	0.044
3	0.040	0.040
4	0.024	0.024

4.异常值处理

（1）将数据标准化后的数据中应付金额和实际金额进行异常值的处理，首认为实际金额和销售金额小于 0 的数据无效数据，进行一次筛选，第二次计算标准差，标准差的绝对值越大说明数据越有可能异常，在这里将阈值设置为 2，标准差大于 2 时被认为异常值。结果部分如下图，总共有 1707 条数据记录为异常数据，应该将其进行删除（具体代码见附件）

	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
54	DD201708167493001227609686693	E43A6E078A04228	16.5	16.5	统一来一桶香辣老坛酸菜牛肉面	2017/1/2 16:59	C	已出货未退款	已提现
72	DD20170816749300216167858379323	E43A6E078A06874	11.0	11.0	250ml红牛	2017/1/2 21:25	E	已出货未退款	已提现
76	DD20170816749300219460569744911	E43A6E078A06874	16.0	16.0	益力多100g*5瓶	2017/1/2 23:01	E	已出货未退款	已提现
89	DD201708167493662320583621934	E43A6E078A04228	12.0	12.0	250ml维他柠檬茶	2017/1/3 12:28	C	已出货未退款	已提现
128	DD201708167493310994276352579	E43A6E078A06874	112.5	112.5	100g*5瓶益力多	2017/1/4 0:42	E	已出货未退款	已提现
186	DD201708167493318471937072216	E43A6E078A06874	11.0	11.0	100g*5瓶益力多	2017/1/7 13:02	E	已出货未退款	已提现
188	DD20170816749300679580989029282	E43A6E078A04228	34.0	34.0	188g澳门永辉鸡仔饼	2017/1/7 14:02	C	已出货未退款	已提现
207	DD2017060120342069B41A466BA7E	E43A6E078A04172	25.0	25.0	芙蓉王	2017/1/8 19:57	A	已出货未退款	已提现
242	DD201708167493847144280155851	E43A6E078A04228	11.8	11.8	37.5g费列罗榛果威化巧克力3颗	2017/1/8 20:24	C	已出货未退款	已提现
313	DD201708167493324092467752323	E43A6E078A06874	13.0	13.0	250ml红牛	2017/1/8 21:45	E	已出货未退款	已提现
324	DD201708167493530494747287176	E43A6E078A04134	12.0	12.0	巧克力奶油蛋糕	2017/1/8 21:57	B	已出货未退款	已提现
329	DD201708167493851137095860395	E43A6E078A04172	18.0	18.0	260g马来西亚进口荣蒂丝口口香饼干	2017/1/8 22:42	A	已出货未退款	已提现
346	DD2017060208293796BA4B4BDCBB5	E43A6E078A06874	25.0	25.0	芙蓉王	2017/1/8 23:26	E	已出货未退款	已提现
377	DD201708167493855206704688743	E43A6E078A07631	12.0	12.0	330ml伊利畅馨乳酸菌原味	2017/1/9 3:11	D	已出货未退款	已提现
388	DD201708167493856598539837215	E43A6E078A04228	14.0	14.0	维他原味豆奶	2017/1/9 5:03	C	已出货未退款	已提现
392	DD201708167493303992205263012	E43A6E078A06874	26.0	26.0	250ml红牛	2017/1/9 7:21	E	已出货未退款	已提现
461	DD201708167493530861468619957	E43A6E078A04172	51.0	51.0	188g澳门永辉鸡仔饼	2017/1/9 12:04	A	已出货未退款	已提现
521	DD201708167493276805059500327	E43A6E078A06874	12.5	12.5	250ml维他柠檬茶	2017/1/9 14:05	E	已出货未退款	已提现

(2) 将异常值去掉之后留下的数据如下图（代码详解见附件）留下来的数据可以导出为文件之后在进行接下来的分析。

```
In [10]: d=a[df['实际金额'].abs()<2]
```

```
In [11]: print(d)
```

	订单号	设备ID	应付金额	实际金额	\
0	DD201708167493663618499909784	E43A6E078A07631	4.5	4.5	
1	DD20170816749366355814061164	E43A6E078A04172	3.0	3.0	
2	DD201708167493578526890939886	E43A6E078A06874	5.5	5.5	
3	DD201708167493683507186615837	E43A6E078A04228	5.0	5.0	
4	DD201708167493759548618252006	E43A6E078A04134	3.0	3.0	
5	DD2017081016294251D0FA5D314F1	E43A6E078A04134	4.5	4.5	
6	DD201708167493663534589050871	E43A6E078A04228	7.0	7.0	
7	DD201708167493663526093267894	E43A6E078A04228	8.0	8.0	
8	DD20170526144916178D2B0429C26	E43A6E078A04134	3.0	3.0	
9	DD20170810162914447BA58EF3127	E43A6E078A04228	2.0	2.0	
10	DD2017081674930611792937303584	E43A6E078A07631	5.0	5.0	
11	DD2017081674930612078044876854	E43A6E078A04228	3.0	3.0	
12	DD2017083114403518C7B2A1F9252	E43A6E078A04228	3.0	3.0	
13	DD201708167493596679223858182	E43A6E078A06874	5.0	5.0	
14	DD201708167493596702065938611	E43A6E078A06874	5.0	5.0	
15	DD201708167493596819993334758	E43A6E078A06874	5.0	5.0	
16	DD201708167493663388737478695	E43A6E078A04172	5.8	5.8	
17	DD201708167493663381827746988	E43A6E078A04172	0.8	0.8	
18	DD201708167493596828748693834	E43A6E078A06874	5.0	5.0	

三、统计分析

1.预处理命名文件

将进行预处理后的数据命名为“新数据.csv”（具体代码见附件），并将其中数据按地区 A, B, C, D, E 分别分成 5 个文件并命名为：“A.csv”，“B.csv”……以便于下面进行统计分析。（A 文件即 task1-1A 文件等等）

2. 每台售货机 2017 年 5 月交易总额和订单量和全年的交易总额和订单量

(1) 每台售货机 2017 年 5 月的交易额和订单量（利用 python 读取文件后进行

筛选日期后得出结果并绘制成表格，具体代码见附件）

	A	B	C	D	E	合计
2017 年 5 月订单量 (个)	728	843	750	539	1262	4122
2017 年 5 月交易总额 (元)	2866.3	3236.3	3042.1	2001.6	5184.2	16330.5

分析：从表格中可以看出，5 月份各地的销量和交易额以 E 地的自动售货机居高，而 D 地的自动售货机销量和交易额偏低，其他三地的交易额和销量比较相近。总的来说，可以看出以 5 月份来说，E 地的销量和交易额偏高，可以加重商品的输出量，而 A 地 B 地 C 地也可以继续增加商品的输出量，D 地的销量不太好，可能的原因在于所处位置过于偏僻或者顾客不太愿意使用自动售货机等原因，应该采取适当的促销活动比如降价等提高销量。

(2) 计算所有售货机交易总额和订单总量（利用 python 读取文件后进行操作得出结果，具体代码见附件）

	A	B	C	D	E	合计
所有订单量 (个)	10192	13166	14054	8532	23028	68974
所有交易总额(元)	37435.7	47643.4	53128.6	30176.1	86931.6	255315.4

分析：根据总交易总额和订单总量可以看出，总的销售情况跟（1）分析的 2017 年的不多，但是总体差距偏大，E 地的销售量和交易额都远远的偏高于其他四地，可以继续保持地方的优势并加大商品的输出，甚至可以在 E 地多建自动售货机。相对来说，A 地和 D 地在总的交易额和销售量比较相近，属于五地中偏低的地方，应该适当的进行全年的促销活动，在销量偏低的月份尤其进行促销活动，

以提高整体的交易额和销售水平，B 地和 C 地也可以适当增加促销活动，在原有的优势上在提高销量。

3. 每台自动售货机每月的每单平均交易额和日订单量

(1) 每台售货机每月的每单平均交易额（每月的销售总量除以每月的订单量）
（具体代码见附件）

每月的每单平均销售额/元	A	B	C	D	E	总月平均销售额/元
一月	3.70	3.57	3.80	3.63	3.95	3.73
二月	3.51	3.20	3.58	3.00	3.43	3.344
三月	3.31	3.39	3.60	3.56	4.21	3.614
四月	3.90	3.60	3.92	3.69	3.71	3.764
五月	3.94	3.84	4.10	3.71	4.10	3.938
六月	3.60	3.64	3.80	3.77	3.59	3.68
七月	3.67	3.89	3.64	3.63	3.59	3.684
八月	3.19	3.45	3.69	3.30	3.66	3.458
九月	3.93	3.63	3.81	3.55	3.77	3.738
十月	3.65	3.72	3.86	3.60	3.56	3.678
十一月	3.95	3.69	3.78	3.56	3.89	3.774
十二月	3.53	3.46	3.66	3.34	3.90	3.578
每台机器平均销售额/元	3.656667	3.59	3.77	3.528333	3.78	3.665

从机器平均销售额来说，每台自动售货机平均销售额以 C 地和 E 地偏高，这两个地方可以继续保持商品的输出，而 D 地和 B 地还有 A 地都偏低，这三个地方应该提高交易总额以提高平均销售额，在售价较高的商品上多入手，多做促销，可以提高平均销售额。

从月份平均销售额来说，以二月、八月、十二月偏低，以五月、四月、九月、

十一月偏高，在月平均销售额偏低的月份应该采取促销的活动，在相应拉低该月份的平均销售额的地方多采取促销活动，提高该月的月均销售额。而其他占优势的月份应该继续加大商品的输出量，将月份的优势和地方的优势结合起来。

总体来说，根据这份表格的数据，我们应该根据每台机器每个月份的平均销售额，对比总体的平均销售额来说，如果高了就继续提高商品量的输出，如果低了就应该就这台机器这个月份进行促销活动，特别是地方和月份的影响都要考虑，两者综合提高整体平均销售。（比如 D 机器的二月份就是偏低，应该实行促销活动等，E 机器的 5 月份偏高，应该继续保持优势并加大商品输出）

（2）每台售货机每月的日均订单量（数据都四舍五入取整，具体代码见附件）

每月的日均订单量 (取整)/个	A	B	C	D	E	总日平均销售量/个
一月	10	11	12	8	11	10
二月	4	7	7	5	9	6
三月	8	8	8	6	11	8
四月	15	20	23	15	29	20
五月	23	27	24	17	41	26
六月	54	60	60	34	85	59
七月	14	11	24	10	27	17
八月	21	31	40	23	57	34
九月	33	57	54	32	134	62
十月	48	64	70	37	89	62
十一月	37	65	63	39	163	73
十二月	63	70	75	53	103	73
每台机器日均订单量/个	28	36	38	23	63	38

根据每台机器日均订单量来看，E地的自动售货机还是普遍的偏高，可以继续加大E地的商品输出，而其他四地的日均订单量还是低于平均值，应该相应的进行促销策略或者降价营销等，特别A地和D地的自动售货机，日均订单量偏低，需要进行整改或者商品的连续促销。

根据日均销售量来看，一月到五月和七月的日均销售量较低，九月到十二月的日均销售量较高，可以继续增加九月到十二月的商品，在一月到五月和七月时特别增加促销活动等，增加各地的日均销售量。

总体来看，一月到八月的日均订单量都比较低，九月到十二月的日均订单量偏高，D地和A地的日均销售量一年下来都偏低，应该就D地和A地的一月到八月着重进行整改，进行较大的促销活动提高顾客量，对于其他地点的自动售货机从一到八月和A地和D地的九月到十二月进行小范围促销活动等，其他时间和地点可以尽量提高商品输出，保持时间和地点的优势。

4. 月均销售次数和月均销售金额

（四舍五入，必要时取两位小数）

	A	B	C	D	E	总
月均销售 次数/次	849	1097	1171	711	1919	5747
月均销售 额/元	3119.7	3970.3	4427.4	2514.7	7244.3	21276.4
平均每单 销售 额 / 元	3.67	3.61	3.78	3.53	3.78	3.70

分析：月均销售次数和月均销售额还是以A地和D地偏少，以E地偏高，B地和C地居中，平均每单销售额以B地和D地偏低，需要着重对A地和D地进行整改和促销，B地和C地进行适当的促销，E地继续保持地方的优势，扩大顾客量。

5.总体分析

总体来说，A 地和 D 地在各种指标上都属于偏低，应该特别着重对于 A 地和 D 地的自动售货机进行促销活动或者整改，以避免因为地方的影响导致销量和交易额都偏低，然后 B 地和 C 地较为均衡，也可以进行适当的促销，而 E 地在各项指标中都处于首位，可见地方很受顾客的欢迎，可以适当的在 E 地增加自动售货机或者增加商品的输出量等，在原有优势上进一步扩大交易额和订单量。

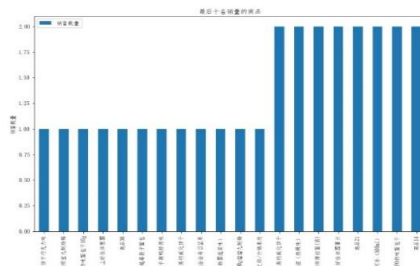
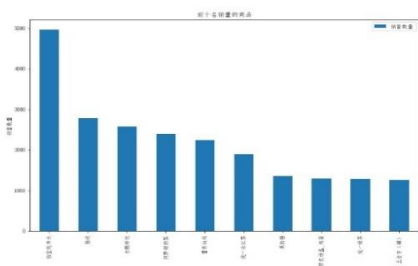
就一年的时间而言，上半年各地日均订单量普遍偏低，下半年日均订单量普遍偏高，而日均交易额普遍相平，但以二月、八月、十二月偏低，以五月、四月、九月、十一月偏高，对此应该就销量和交易额的月份普遍偏低的着重都进行促销等（比如说二月普遍偏低）而对于普遍偏高的（比如十一月）进行商品输出增加的操作。

不同地方的不同月份的各项指标都不同，对于相对集中薄弱的应该大力整改，对于相对集中强势的地方应该继续加强商品输出，从而达到整体的各项指标的增加。

四、数据可视化

1. 绘制销售量前十和销售量最后二十位的柱状图

注：涉及到销售数量，数据中默认为 1，则在数据中最后一栏加一列数值为 1 的数据，并命名为销售数量，运行 python 中的代码，代码详见附件，得到下面结果，销量前十的比较有比较的效果，销量后二十的柱状图比较难以看出比较，大多数集中在 1，2 销售量左右。（图片看不清可以看附件的文件夹）

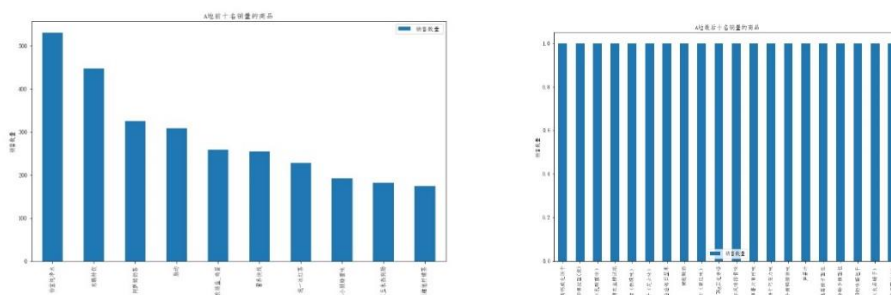


分析：总体来看，怡宝纯净水的销量居于首位，其次是脉动、东鹏特饮、阿萨姆奶茶、营养快线等销量处于前十，65g 格力高百力滋饼干巧克力味，90g 顺宝九制杨梅等的销量处于销量最后，总体看到销量较多的大多数是饮料类，特别是怡宝纯净水，销量处于最后大多数是非饮料类，特别是面包，杨梅，饼干等，销量最后的商品可以考虑不进货，对于销量较多的可以适当加大商品的入货。

2.分别绘制五台机器的销量前十和销量后十

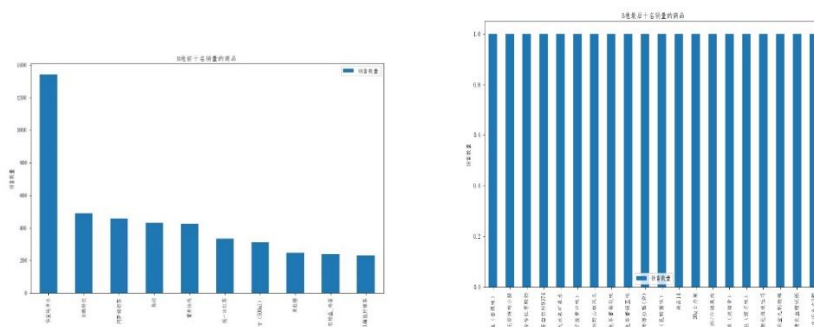
注：分别在原来的数据上都加上销售数量这一列，并且值全为 1，命名仍然不变。（图片看不清可以看附件的文件夹）

(1) A 售货机的销量前十和销量后二十



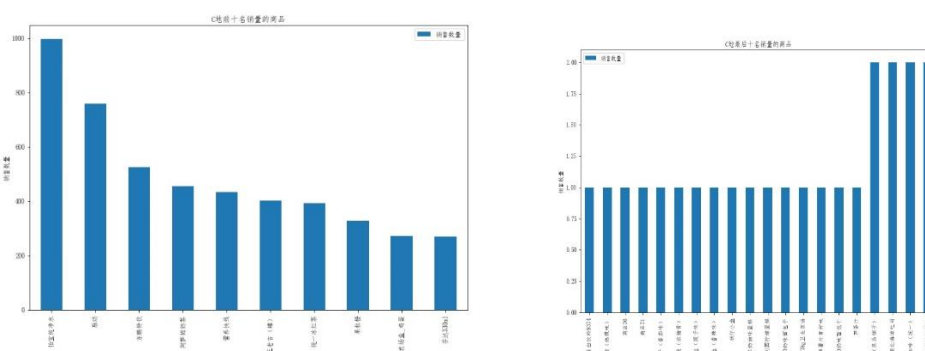
分析：总体来看，销量前几的仍然是怡宝，脉动，东鹏特饮等饮料类，不过 A 地的前十多了无穷农场盐-鸡蛋，玉米肠，无穷烤鸡小腿（蜂蜜味）等，可以适当的增加饮料类的时候，也应该迎合当地顾客的风格，多入些鸡蛋，玉米肠，鸡小腿等的产品。而销量最后的二十位也一样大多数是非饮料类，以面包，坚果，杨梅，饼干等居首，进货时可以考虑下架或者更改商品。

(2) B 售货机的销量前十和销量后二十



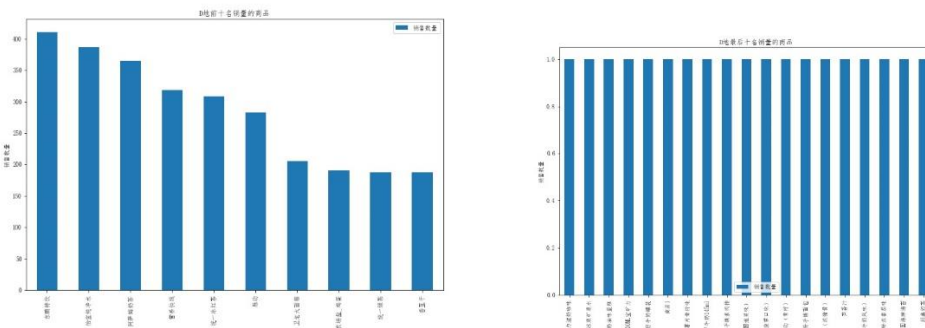
分析：总体来看，销量前十的仍然是怡宝，脉动，东鹏特饮等饮料，也有无穷农场盐-鸡蛋，果粒橙，柠檬茶等处于销量前十，其中怡宝的销量遥遥领先，进货时也应该往 B 地多增加销量前十的产品，多以饮料类进货，特别是怡宝纯净水的进货。销量后二十中，尽管大部分还是非饮料，但也有恒大冰泉矿泉水，无穷鸡小腿处于销量后二十，在进货时，也应该适当的减少该种商品的进货。

(3) C 售货机的销量前十和销量后二十



分析：总体来看，C 地的销量前十还是怡宝、脉动、东鹏特饮等饮料，以怡宝纯净水和脉动居高，还有芬达也处于销量前十，可以进货时多考虑销量前十的。销量后二十的除了面包，饼干等，也有蛋糕，咖啡和芦荟汁等不受顾客的欢迎，可以考虑下架或者更换产品。

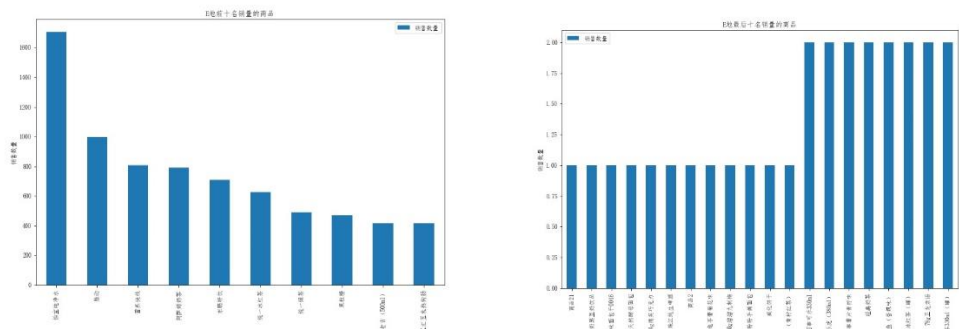
(4) D 售货机的销量前十和销量后二十



分析：总体来看，D 地的销量前十仍然以东鹏特饮，怡宝纯净水，脉动等在前面，不过增加了卫龙大面筋，无穷农场盐-鸡蛋，香豆干等在前十，进货时可以考虑多进这些前十的商品，而处于销量后二十的除了面包，芦荟汁，蛋糕之外，

也有脉动（青柠味），宝矿力，旺仔牛仔等产品，在进货时可以考虑下架或更换口味，比如脉动可以更换口味。

(5) E 售货机的销量前十和销量后二十



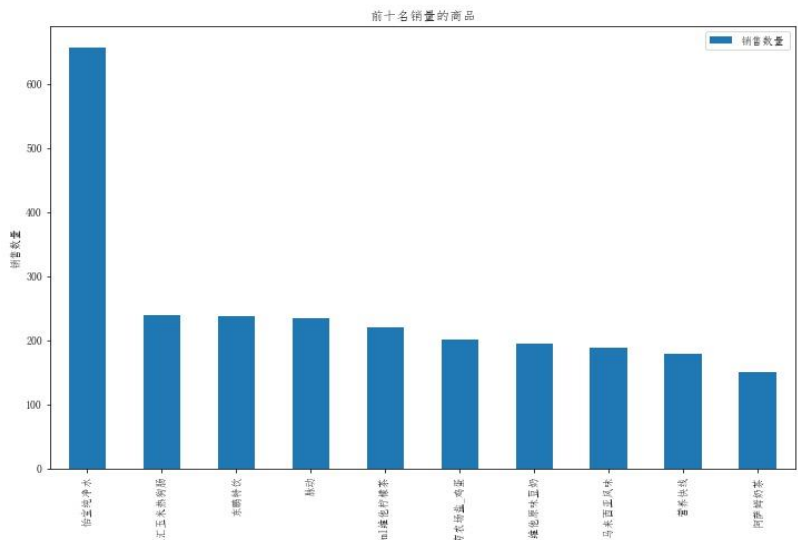
分析：总体来看，E 地的总体销量都是偏高，其中怡宝纯净水的销量遥遥领先，此外脉动，东鹏特饮等饮料都处于前十，此外，E 地的销量前十还有双汇玉米热狗肠等非饮料，可以在进货时适当增加，E 地作为销量很好的地方，更应该注重销量前几商品的进货。而处于后二十的商品除了面包巧克力饼干之外，还有百事可乐冰红茶啤酒等处于后二十，可以考虑下架或者更换饮料品牌味道等。

(6) 总体分析

总的来看，各个地方和总体的都以怡宝纯净水、东鹏特饮、脉动、阿萨姆奶茶等饮料类处于销量前几，但是不同地方也有不同的销量前十，不如无穷农场盐-鸡蛋，热狗肠、香豆干、卫龙大面筋和芬兰等碳酸饮料居于前几，但是后二十除了普遍的面包、饼干、杨梅、蛋糕等非饮料产品，不同地方也有类似恒大冰泉水，脉动（青柠味），啤酒，冰红茶，旺仔牛奶，百事可乐等饮料的出现，所以在进货时，更应该根据各地的销售量情况来绝地产品的进货量，不同地方的顾客风格不同，应该进行不同进货和下架选择。（比如前几的怡宝纯净水、脉动、东鹏特饮等各地可以保持相同，但是类似碳酸饮料 E 地不宜进货，宝矿力、旺仔牛奶等 D 地不宜进货等选择）

3.绘制 2017 年 6 月份销售前十柱状图

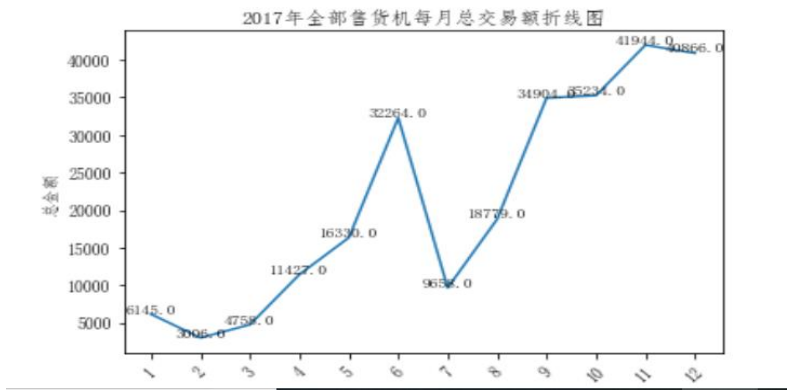
从上述分析可以看到六月份的销售额普遍偏高，可以就 2017 年 6 月份分析销售前十的数据进行柱状图的描绘（具体见代码，本次只分析总数据）结果如下图



分析：总体来看，饮料类商品还是属于销量前十的重要商品，特别对于 6 月份夏季来说，饮料类的销量也普遍偏高，其中怡宝矿泉水也遥遥领先，还有脉动，东鹏特饮等补充能量的饮料可以考虑多进货，而玉米肠，无穷农场盐一鸡蛋作为非饮料类也可以选择多进货，奶类饮料豆奶和阿萨姆奶茶相对于前几的饮料可以尽量减少入货，但相对于其他还是入货要增加。

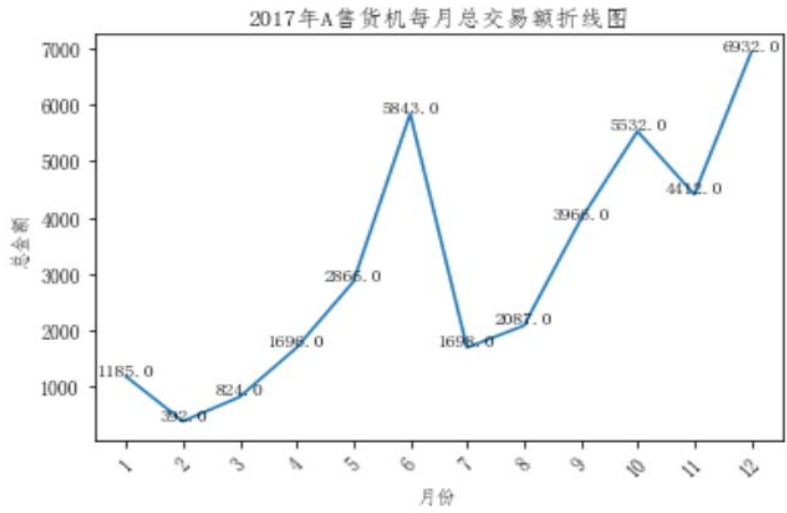
4.绘制每台售货机每月总交易额折线图

(1) 2017 年全部售货机每月总交易额



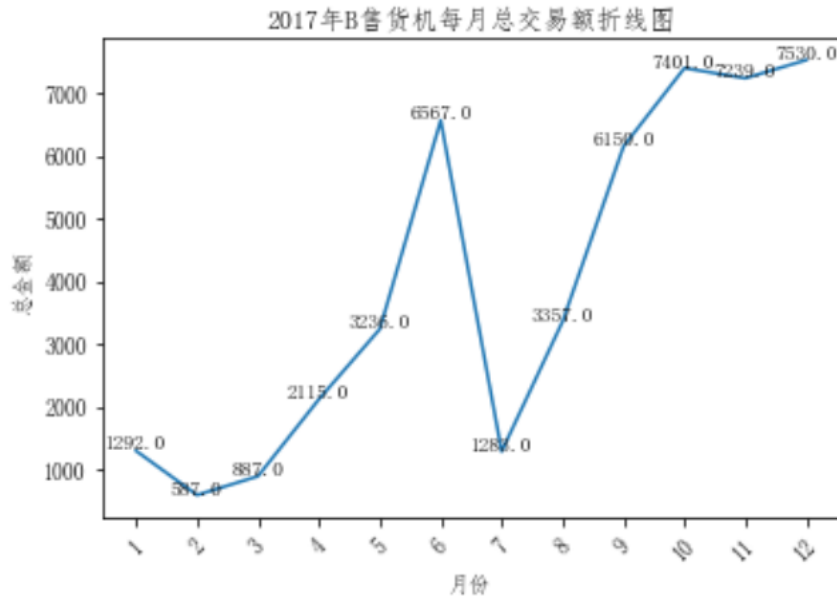
分析：根据折线图可以看到，对于所有售货机总和来说，2月的总交易额偏低，其次是1月、3月、和7月，而6月、9月到12月的总交易额偏高，可以在1月到3月和7月进行商品的促销活动，比如商品的降价或者打折等来吸引顾客。而6月、9月到12月可以适当的进行商品的投入增加活动，适当小规模促销活动等。可以看出，从二月到6月成上升趋势，六月到七月又呈下降趋势，且下降幅度较大，可以进行促销等，七月到十二月又呈现较快的上升趋势，可以继续保持该上升趋势，比如不断提高商品投入等。

(2) 2017年A售货机每月总交易额



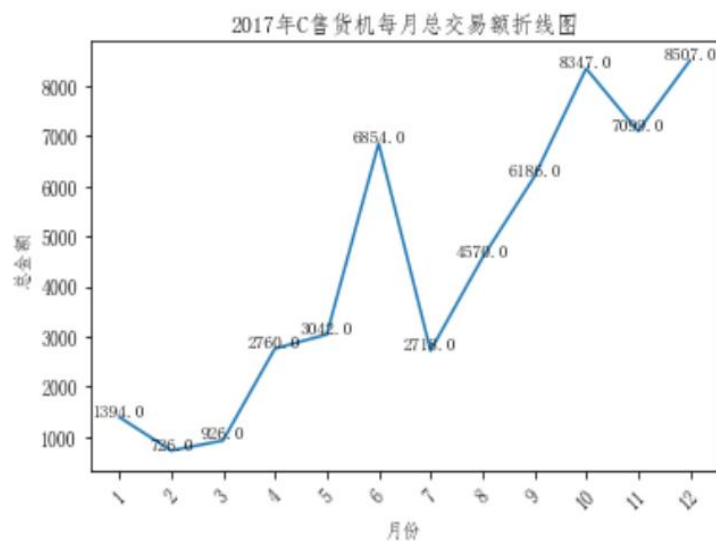
分析：根据折线图可以看到，对于A售货机总和来说，2月的总交易额偏低，其次是1月、3月、4月、7月和8月，而6月、9月到12月的总交易额偏高，可以在1月到4月和7月到八月进行商品的促销活动，比如打折或者降价来提高商品的销量从而吸引顾客。而6月、9月到12月可以适当的进行商品的投入增加活动，适当小规模促销活动等。可以看出，从2月到6月成上升趋势，6月到7月又呈下降趋势，且下降幅度较大，可以进行促销等，7月到12月又呈现较快的上升趋势，中间有小幅下降，但可以继续保持该上升趋势，比如不断提高商品投入等。

(3) 2017年B售货机每月总交易额



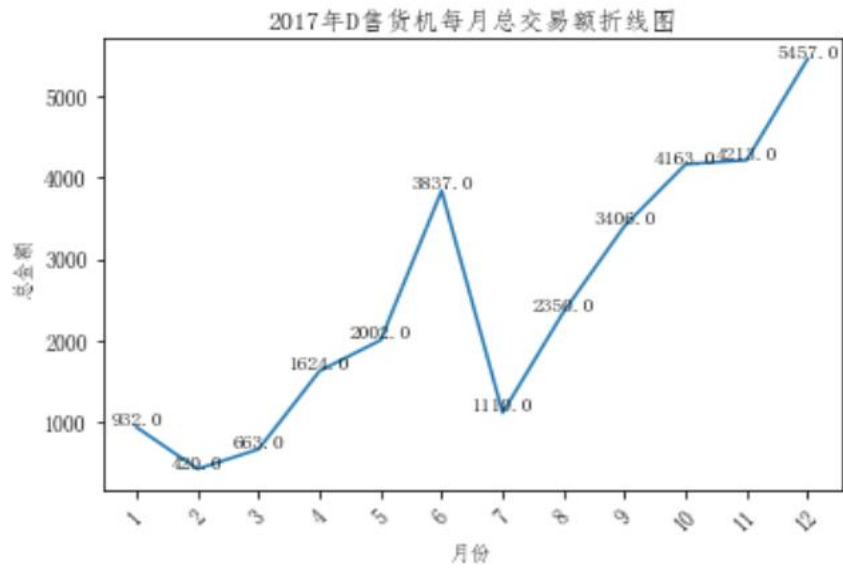
分析：根据折线图可以看到，对于B售货机总和来说，1月到3月和7月的总交易额偏低，而6月、9月到12月的总交易额偏高，可以在1月到3月和7月进行商品的促销活动，比如打折或者降价来提高商品的销量从而吸引顾客。而6月、9月到12月可以适当的进行商品的投入增加活动，适当小规模促销活动等。可以看出，从2月到6月成上升趋势，6月到7月又呈下降趋势，且下降幅度较大，可以进行促销等，7月到12月又呈现较快的上升趋势，可以继续保持该上升趋势，比如不断提高商品投入等。

(4) 2017年C售货机每月总交易额



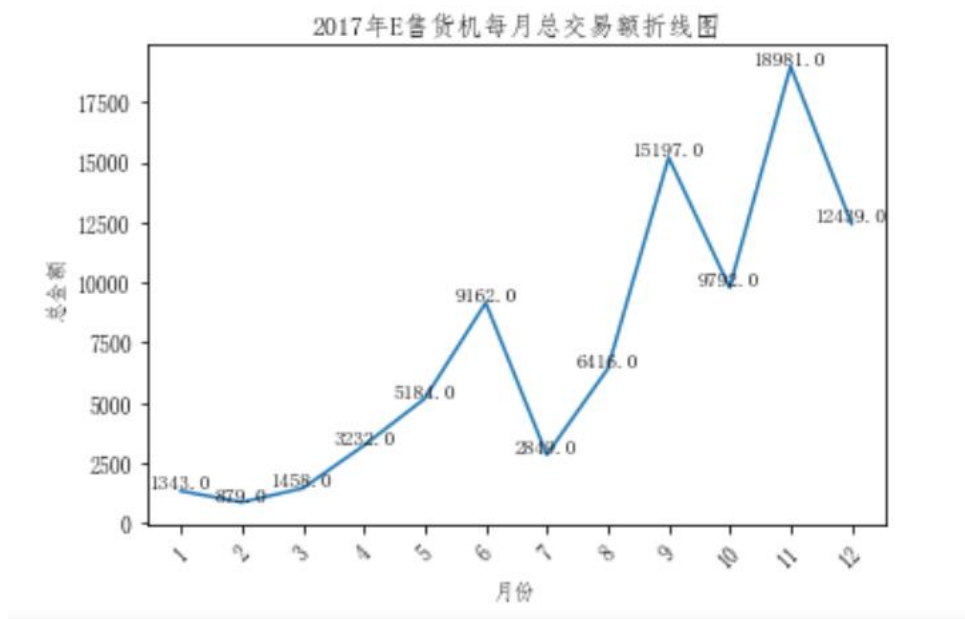
分析：根据折线图可以看到，对于 C 售货机总和来说，1 月到 3 月总交易额偏低，而 6 月、9 月到 12 月的总交易额偏高，可以在 1 月到 3 月进行商品的促销活动，比如打折或者降价来提高商品的销量从而吸引顾客。而 6 月、9 月到 12 月可以适当的进行商品的投入增加活动，适当小规模促销活动等。可以看出，从 2 月到 6 月成上升趋势，6 月到 7 月又呈下降趋势，且下降幅度较大，可以进行促销等，7 月到 12 月又呈现较快的上升趋势，中间有小幅下降，但可以继续保持该上升趋势，比如不断提高商品投入等。

(5) 2017 年 D 售货机每月总交易额



分析：根据折线图可以看到，对于 D 售货机总和来说，1 月到 4 月总交易额偏低，而 6 月、9 月到 12 月的总交易额偏高，可以在 1 月到 4 月进行商品的促销活动，比如打折或者降价来提高商品的销量从而吸引顾客。而 6 月、9 月到 12 月可以适当的进行商品的投入增加活动，适当小规模促销活动等。可以看出，从 2 月到 6 月成上升趋势，6 月到 7 月又呈下降趋势，且下降幅度较大，可以进行促销等，7 月到 12 月又呈现较快的上升趋势，可以继续保持该上升趋势，比如不断提高商品投入等。

(6) 2017 年 E 售货机每月总交易额



分析：分析：根据折线图可以看到，对于E售货机总和来说，1月到4月总交易额偏低，而6月、9月到12月的总交易额偏高，可以在1月到4月进行商品的促销活动，比如打折或者降价来提高商品的销量从而吸引顾客。而6月、9月到12月可以适当的进行商品的投入增加活动，适当小规模促销活动等。可以看出，从2月到6月成上升趋势，6月到7月又呈下降趋势，且下降幅度较大，可以进行促销等，7月到12月又呈现较快的上升趋势，中间有两次较大幅度的下降，可以在该段时间进行商品的促销等活动，总体为向上的趋势，可以继续保持该上升趋势，比如不断提高商品投入等。

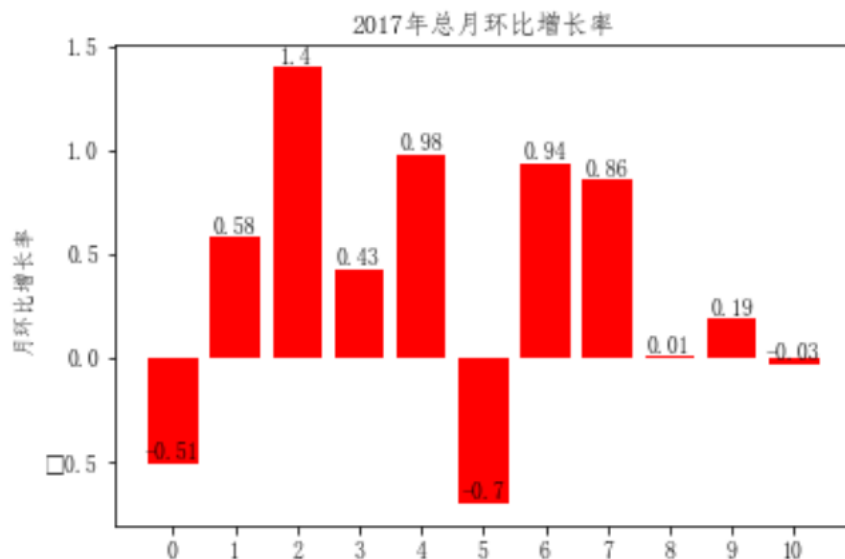
（7）总体分析

总的来说，各地的自动售货机都是以2月为最低，1月到3月还有4月以及7月都是销量偏低的时间段，且1月到6月是小幅度上升时期，6月销量也普遍偏高，而9月到12月都是销量普遍偏高且持续上升阶段，中间偶尔会有小幅度的下降，而6月到7月普遍会有一次大幅度的下降。因此，总体来说，应该采取促销活动提高1月到4月的总体销量，适当进行营销策略避免6月到7月的下降，同时加大商品的投入或者小幅度的促销避免9月到12月的小范围下降。

5.绘制月环比增长率

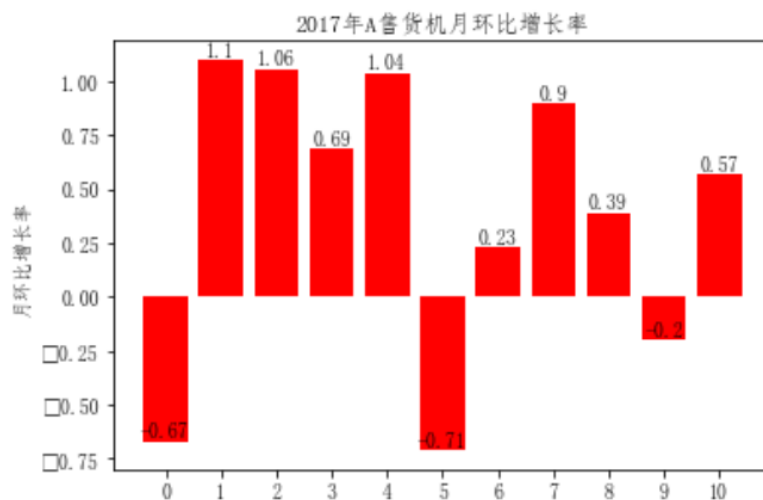
注：总数据和分开五个地方都画柱状图，具体代码见附件（月环比增长率即下月相对于上月增加或者减少的总金额相对于上月所占的百分比）

（1）2017 年总月环比增长率



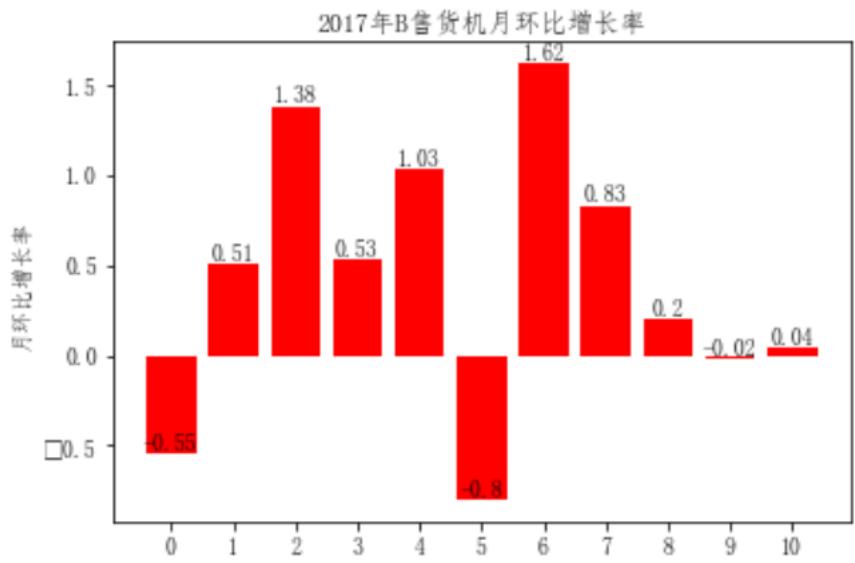
分析：总体来看，2月和7月的月环比增长率呈负且下降较大，4月、6月、8月和9月的月环比增长率呈正且上升较大，可以对2月和7月份采取一定的促销活动，吸引顾客的眼球，避免交易额下降太多，而4月、6月、8月和9月可以着重增加商品投入，增加总交易额。

（2）2017 年 A 售货机月环比增长率



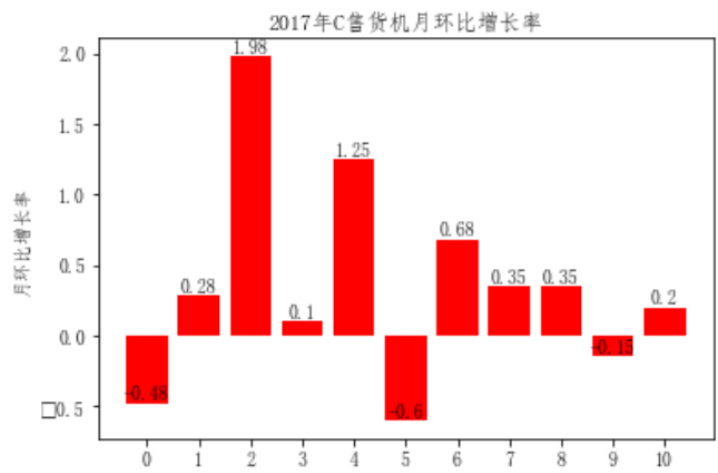
分析：总体来看，2月和7月的月环比增长率呈负且下降较大，3月到6月、9月和12月的月环比增长率呈正且上升较大，可以对2月和7月份采取一定的促销活动，吸引顾客的眼球，避免交易额下降太多，而3月到6月、9月和12月可以加大商品投入，增加总交易额，着重增加3月到6月的商品投入。

(3) 2017年B售货机月环比增长率



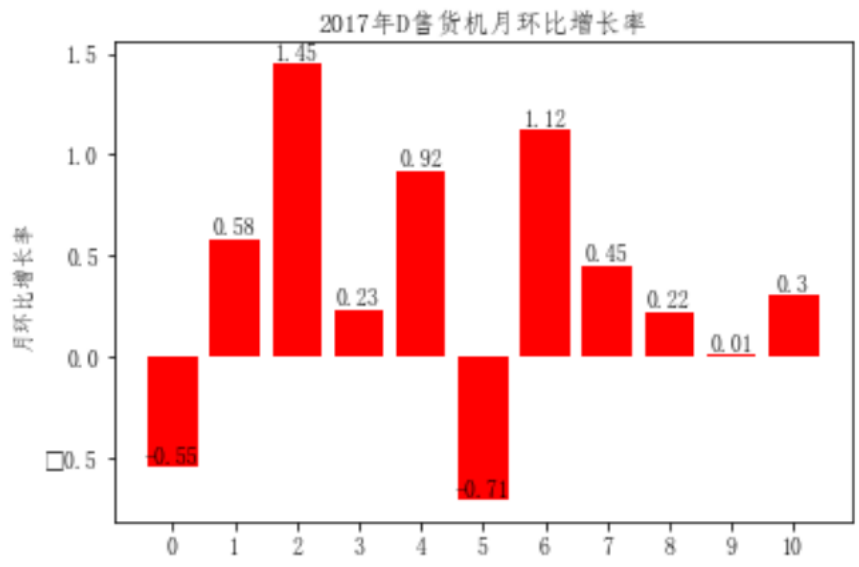
分析：总体来看，2月和7月的月环比增长率呈负且下降较大，3月到6月、9月和10月的月环比增长率呈正且上升较大，可以对2月和7月份采取一定的促销活动，吸引顾客的眼球，避免交易额下降太多，而3月到6月、9月和10月可以加大商品投入，增加总交易额，着重增加4月、6月和8月的商品投入。

(4) 2017年C售货机月环比增长率



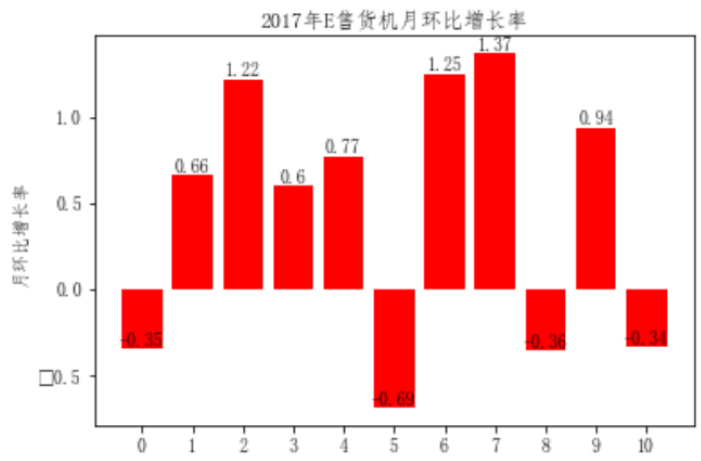
分析：总体来看，2月和7月的月环比增长率呈负且下降较大，4月、6月和9月的月环比增长率呈正且上升较大，可以对2月和7月份采取一定的促销活动，吸引顾客的眼球，避免交易额下降太多，而4月、6月和9月可以加大商品投入，增加总交易额，着重增加4月和6月的商品投入。

(5)2017年D售货机月环比增长率



分析：总体来看，2月和7月的月环比增长率呈负且下降较大，4月、6月和8月的月环比增长率呈正且上升较大，可以对2月和7月份采取一定的促销活动，吸引顾客的眼球，避免交易额下降太多，而4月、6月和8月可以加大商品投入，增加总交易额，着重增加4月和8月的商品投入。

(6)2017年E售货机月环比增长率



分析：总体来看，2月和7月的月环比增长率呈负且下降较大，其次10月和12月都有一定程度的下降，而3月到6月、8月到9月的月环比增长率呈正且上升较大，可以对2月和7月份采取较大的促销活动，吸引顾客的眼球，避免交易额下降太多，对10月和12月进行一定的促销，减少损失，而对3月到6月、8月到9月期间可以加大商品投入，增加总交易额，着重增加4月、8月和9月的商品投入。

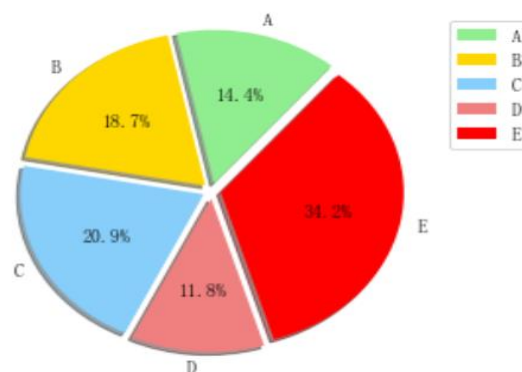
（7）总体分析

总的来说，不同地方的月环比增长率有所差异，但普遍来说，2月和7月是月环比增长率为负且数值较大的两个月，因此在这两个月期间各地都应该要进行促销活动或者整改，吸引顾客的眼球以加大销售量，以达到销售额的增加，而普遍月环比增长率为正且数值较大的为4月、6月、8月和9月，这四个月份各地的优势都很大，可以借此多增加商品投入，增加营业额。对于其他或多或少的增长下降，都可以适当的调整以增加总交易额

6. 绘制每台售货机毛利润占总毛利润比例的饼图

（1）先用 EXCEL 表格的 VLOOKUP 函数将新数据文件中加上大类一列，大类的来源为附件二的数据）加上之后在进行分析。再依次将“A.csv”“B.csv”...文件中都加上大类一栏，依然按照上述 VLOOKUP 的方法。（具体代码见附件）

（2）绘制饼图（具体见代码，饮料类为百分之二十五，非饮料类为百分之二十）（数值如表格）

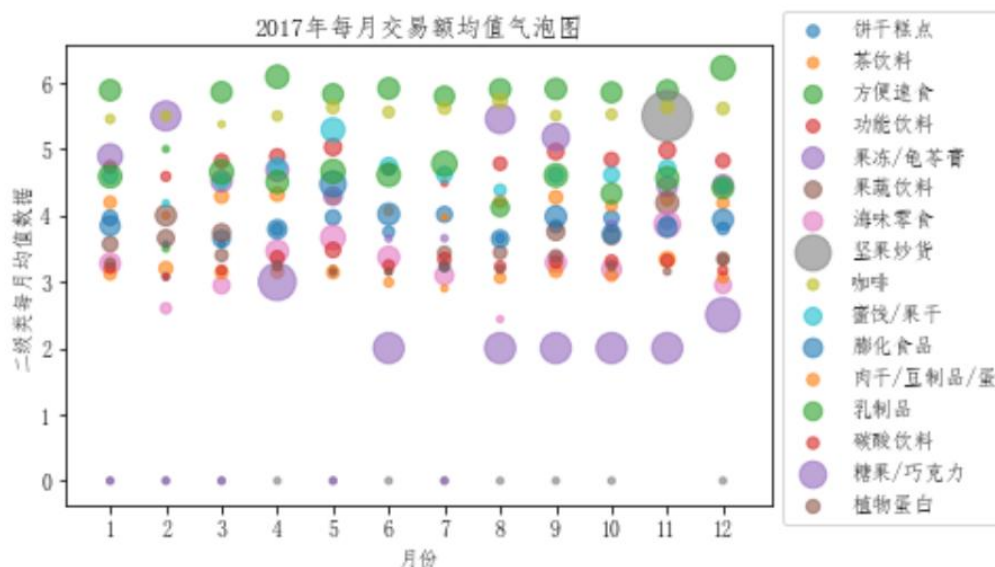


地区	所有地区	A	B	C	D	E
总利润	59586	8583	11157	12453	7009	20382
占比	1	14.4%	18.7%	20.9%	11.8%	34.2%

分析：总体来看，从饼图和表格的数据可以看出，E地的利润占比最大，其次是C地和B地，最后是A地和D地，E地的销售利润占比大离不开销量和交易额的领先，而A地和D地销量和交易额都处于比较低的地位，特别是D地，利润占比也低，因此，就利润占比来说，想要提高总体的利润，首要的是继续保持E地的优势，继续加强商品的投入，其次就是要对A地和D地进行商品的更改或者商品的促销，更好的迎合当地的顾客的口味与风格，就可以就销量来探讨当地的顾客喜欢的商品并加以入货和促销。

7. 绘制每月交易额均值气泡图

注：横轴为月份，纵轴为商品的二级类目。



分析：由气泡图可以看出，方便素食的均值在每个月份都偏高，而坚果炒货类11月份偏高，咖啡类的交易额均值也普遍较高，而糖果/巧克力的交易额均值在一二月份偏高，其他月份都偏低。其他的二级类大多数处于中间，没有因为月份而影响交易额均值，有部分二级类交易额为0。商家进货时就可以就月份交易额均值高的多选择，月份交易额均值少的少进货。

8. 绘制售货机 C 6、7、8 月三个月订单量的热力图

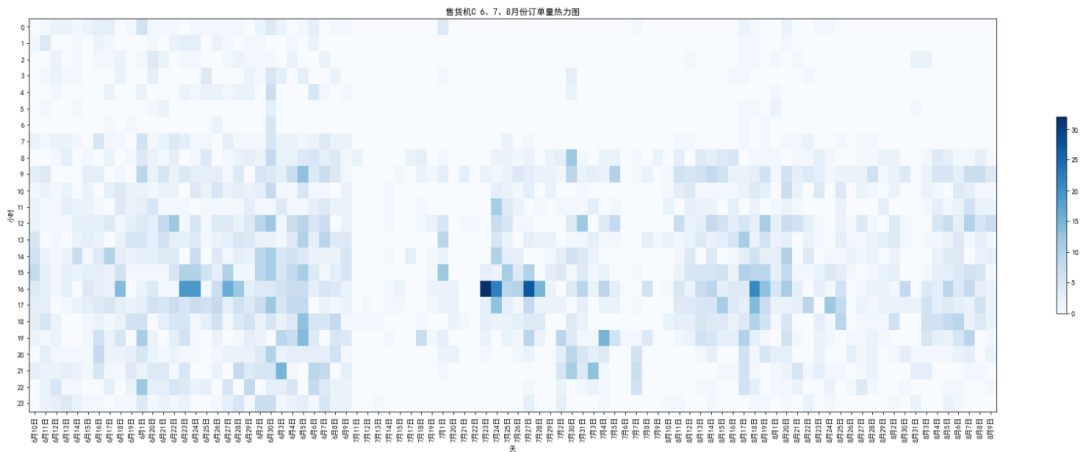
(1) 从文件 ‘C.csv’ 中利用筛选提取售货机 C 6、7、8 月三个月的数据，并将支付时间改为日期和小时的格式，利用 python 统计重复数的功能统计出每一天每一个小时的订单量，并将导出来的数据导出成 ‘CS.xlsx’ 文件。(部分数据如下图所示，左边数据为月份：天数：小时，右边为订单量)

7:23:16	32		
7:27:16	27		
7:24:16	22		
8:18:16	21		
6:24:16	19		
6:23:16	19		
6:27:16	16		
6:03:21	15		
7:28:16	15		
7:04:19	15		
6:05:10	14		

(2) 在 ‘CS.xlsx’ 中先将上述数据左边拆分成月份天数和小时两列，在将所得的三列数据用 EXCEL 的数据透视表的功能绘制成表格，以便下述的热力图绘制，部分数据如下图所示。(横坐标为月份天数，纵坐标为小时，空白的值为 0)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
6月10日		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
6月11日	1		4		1		1		2	1	4	1	2	1	5	6	8	4	3
6月12日	2			2	2	1			2	1		1	1	1	1	1	1	1	3
6月13日	2			1					2	3		2	3	1	2	3	3	3	1
6月14日	1	1	1	1	1		1		1				2	3	2	7	2		2
6月15日	2									1	3	2	2	3	3	1	3	4	3
6月16日	3	2	1						5		3		1	3	2	4	3	3	4
6月17日	3	1	1			2		1	1	2		3		4		10	2	2	5
6月18日			2		2				1		1		4	1	2	3	6	14	3
6月19日	1							1		1		2	2	3	5	2	1		2
6月1日	6		1		2				6	4	9	2	3	4	5	3	1	2	3
6月20日	1	1	4	3		1			1	2	2	1	5	3	3	3	1	4	6
6月21日	1		2				2		2	1	5	2	1	8	6	4	1	1	5
6月22日	1	2							4	3	2	2	1	12	3	4	5	1	7
6月23日		3	1		2				3		4		1	1	3		10	19	5
6月24日	1	3	1		1				1	1	3	4		5	4	7	10	19	7
6月25日	1			4	2				1	4	3	2	1	1	1		6	1	6
6月26日	1	2			2			2			3	5	1	4	2	4	3	7	8
6月27日	1	1	1		1				3	1	1	2	1	4	3	1	10	16	3
6月28日	1		2		2				1	2	4	3	1	2	5	3	1	12	6

(3) 将 CS.xlsx 文件导入 python, 进行缺失值变成 0 和热力图的绘制，代码具体见附件，热力图见下面。(横轴为天，纵轴为时)



(4) 总体分析

总体来看，一天之中销售量相对集中的是从早上 7 点到晚上 23 点左右，其中最集中是在十二点到十八点左右，从图中可以看到颜色比较深，相对销售量集中。而对于从六月到八月来说，六月份相对销售量比较多，七月份中旬之后到八月上旬销售量也偏多，其他时间相对颜色比较浅，销售量比较少。因此，对 C 地销售策略可以就六月份和七月上旬到八月下旬的十二点到十八点进行商品量的投入和促销等，在受顾客欢迎的时间段在加大商品的促销，而对于其他时间段不受顾客欢迎的时间段也可以适当的进行促销，抵消该时间段的不良影响，提高销售量。

五、生成各地自动售货机画像

1. 数据预处理

(1) 先读取每个售货机的文件，利用 python 将商品的重复次数进行一次统计，然后导出分别名为 ‘task3-1A.csv’，‘task3-1B.csv’ … 文件。（具体代码看附件）

(2) 利用 Excel 的 VLOOKUP 函数，将附件二的大类分别对应到导出的文件中，在利用 sumif 函数，利用 ‘A.csv’，‘B.csv’ … 文件分别计算出每种商品的销售总额总量，根据前面所运用的饮料类为百分之二十五，非饮料类为百分之二十的利润，用 if 函数算出每种商品的利润，然后根据前面任务二算出的每台机器的毛利润，算出每台售货机每种商品的毛利润占比，并分别保存，命名不变。在利用 python 将各自的列都加上列名（代码见附件）（部分结果如下，以 A 机器为例）

	商品	销量	大类	总销售额	利润	利润占比
0	怡宝纯净水	531	饮料	1070.5	267.625	0.031181
1	东鹏特饮	448	非饮料	1583.5	316.700	0.036899
2	阿萨姆奶茶	326	非饮料	1307.5	261.500	0.030467
3	脉动	309	饮料	1272.5	318.125	0.037065
4	30g无穷农场盐_鸡蛋	259	非饮料	817.0	163.400	0.019038

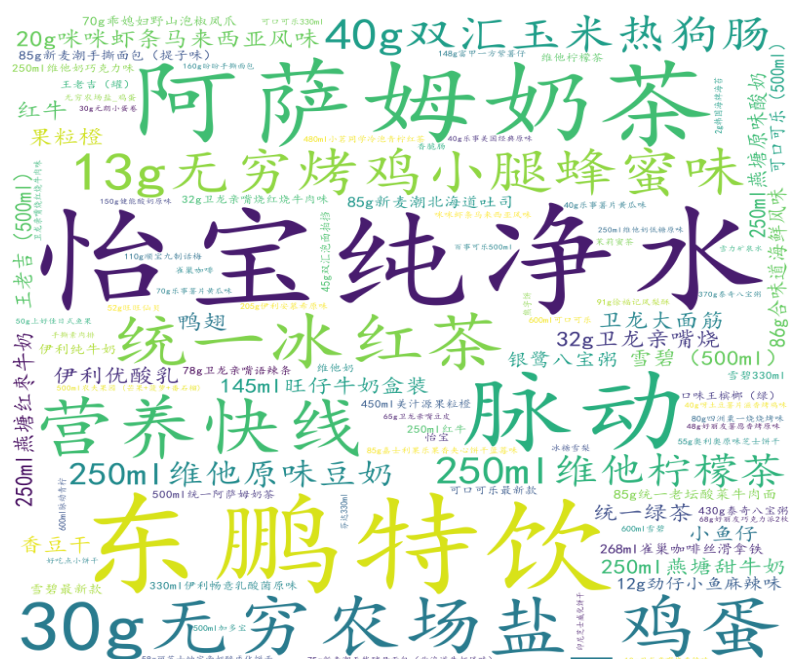
(3) 根据销量和毛利润占比来判断商品的销量情况应该为热销，正常，还是滞销（注：在这里的标签划定范围为标准值与最大值的中间值和标准值与最小值的中间值，大于标准值与最大值的中间值的为热销，处于标准值与最大值的中间值和标准值与最小值的中间值之间的为正常，小于标准值与最小值的中间值的为滞销，满足销量和利润占比的任一情况皆可判定）（具体代码见附件，部分结果如下图，以 A 为例子）最后仍然导出保存成 ‘task3-1A.csv’，‘task3-1B.csv’ … 文件，标签设立完成。

	商品	销量	大类	总销售额	利润	利润占比	标签
0	怡宝纯净水	531	饮料	1070.5	267.625	0.031181	热销
1	东鹏特饮	448	非饮料	1583.5	316.700	0.036899	热销
2	阿萨姆奶茶	326	非饮料	1307.5	261.500	0.030467	热销
3	脉动	309	饮料	1272.5	318.125	0.037065	热销
4	30g无穷农场盐_鸡蛋	259	非饮料	817.0	163.400	0.019038	热销

2.生成自动售货机画像

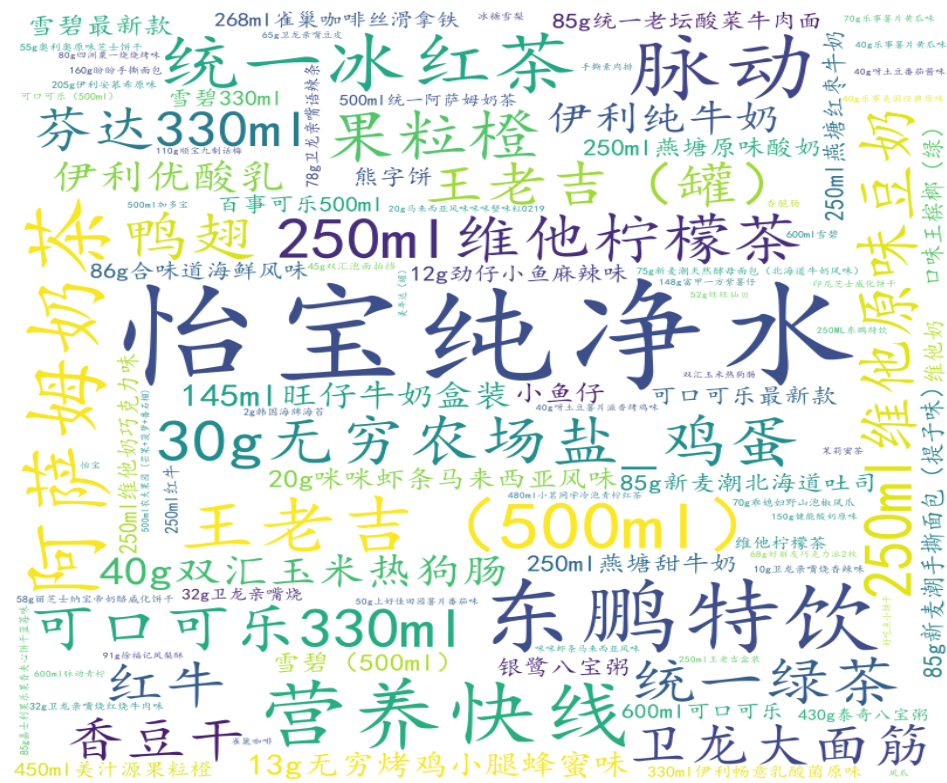
（根据销量生成词云图，让用户对商品的情况可以一目了然，以下给出每个自动销售机的词云图，即每个自动售货机的画像，代码的文件名更改就可以生成各个自动售货机的画像）（具体代码见附件）

(1)A 自动售货机画像



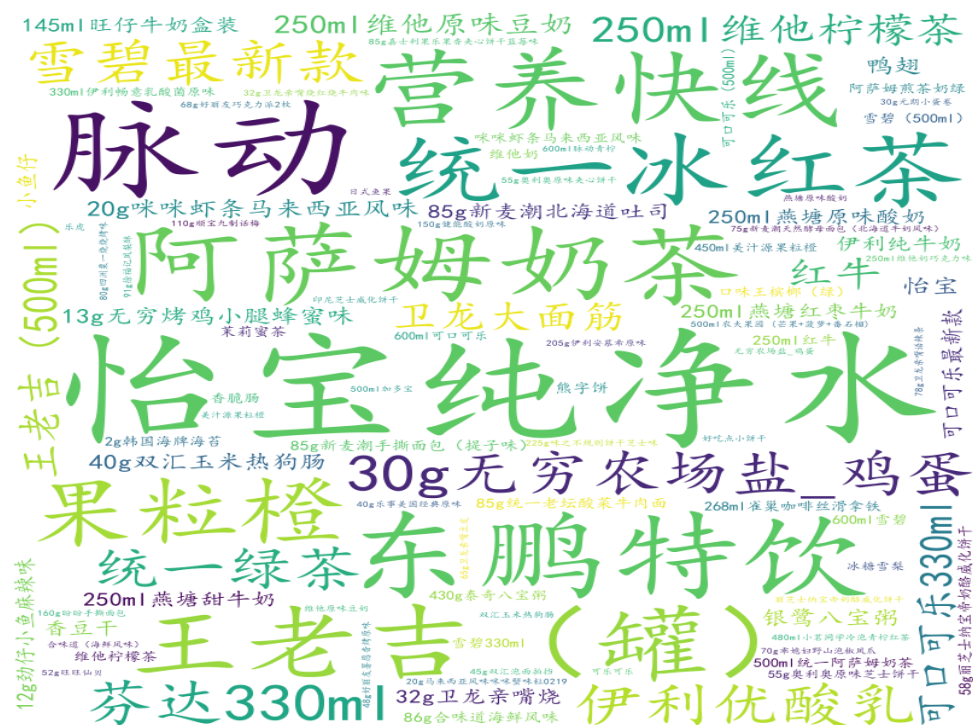
分析; 总体来看, 可以看出 A 地的顾客比较倾向于怡宝纯净水, 阿萨姆奶茶, 脉动, 东鹏特饮等饮料的同时, 也对无穷烧鸡小腿蜂蜜味, 无穷农场盐_鸡蛋, 柠檬茶, 豆奶等有一定的喜好, 而其他商品相对来说销量较少, 甚至有些比如雪碧面包等销量很好, 入货时可以根据当地顾客选择在入货。

(2) B 自动售货机画像



分析: 总体来说, B 地的顾客对于怡宝纯净水, 东鹏特饮, 脉动, 阿萨姆奶茶, 统一冰红茶, 香豆干, 卫龙大面筋, 可口可乐, 王老吉等比较感兴趣, 而相应的也对面包, 雪碧最新款, 拿铁等不太感兴趣, 入货时可以根据当地人的口味入货 (同时大瓶的可口可乐不一定有小瓶的受顾客欢迎)

(3) C 自动售货机画像



分析：总体来说，C 地的顾客比较倾向于怡宝纯净水，东鹏特饮，王老吉，脉动，营养快线，芬达等饮料，但也对面包，燕塘酸奶，柠檬茶，伊利纯牛奶等不感兴趣，入货时可以根据当地顾客选择决定入货。

(4) D 自动售货机画像



分析：总体来说，D 地的顾客对于怡宝纯净水，阿萨姆奶茶，东鹏特饮，营

地的顾客选择决定入货的数量。

六、业务预测

1. 灰色预测模型理论

灰色预测模型（Gray Forecast Model）是通过少量的、不完全的信息，建立数学模型并做出预测的一种预测方法。当我们应用运筹学的思想方法解决实际问题，制定发展战略和政策、进行重大问题的决策时，都必须对未来进行科学的预测。预测是根据客观事物的过去和现在的发展规律，借助于科学的方法对其未来的发展趋势和状况进行描述和分析，并形成科学的假设和判断。

灰色系统理论是研究解决灰色系统分析、建模、预测、决策和控制的理论。灰色预测是对灰色系统所做的预测。目前常用的一些预测方法（如回归分析等），需要较大的样本，若样本较小，常造成较大误差，使预测目标失效。灰色预测模型所需建模信息少，运算方便，建模精度高，在各种预测领域都有着广泛的应用，是处理小样本预测问题的有效工具。

2. 灰色系统的特点

- (1) 用灰色数学处理不确定量，使之量化。
- (2) 充分利用已知信息寻求系统的运动规律。
- (3) 灰色系统理论能处理贫信息系统。

3. 灰色生成

将原始数据列中的数据，按某种要求作数据处理称为生成，尽管表述其行为的数据可能是杂乱无章，然而它必然是有序的，都存在着某种内在规律，不过这些规律被纷繁复杂的现象所很难直接从原始数据中找到，某种内在的对数据的生成就是企图从杂乱无章的现象中去发现，常用的灰色系统生成方式有：累加累减均值生成，级比生成。

4. 累加生成简介

累加生成，即通过数列间各时刻数据的依个累加以得到新的数据与数列。累加前的数列称原始数列，累加后的数列称为生成数列。累加生成是使灰色过程由灰变白的一种方法，它在灰色系统理论中占有极其重要地位，通过累加生成可以看出灰量积累过程的发展态势，使离乱的原始数据中蕴含的积分特性或规律加以显化。累加生成是对原始数据列中各时刻的数据依次累加，从而生成新的序列的一种手段。

式子可以归纳为：

$$x^{(1)}(i) = \left\{ \sum_{j=1}^i x^{(0)}(j) \mid i = 1, 2, L, N \right\}$$

此式子所表示的数据列为原始数据的一次累加生成，简称为一次累加

5. GM (1,1) 模型

(1) 灰色系统是对离散序列建立的微分方程，GM (1,1) 是一阶微分方程模型，其形式为：

$$\frac{dx}{dt} + ax = u \quad GM(1,1)$$

(2) 离散形式和预测公式如下：

$$\Delta^{(1)} \left(x^{(1)}(k+1) \right) + a(x(k+1)) = u$$

$$\hat{x}^{(1)}(k+1) = \left[x^{(1)} - \frac{\hat{u}}{\hat{a}} \right] e^{-\hat{a}k} + \frac{\hat{u}}{\hat{a}}$$

(3) 由导数定义可知：

$$\frac{dx}{dt} = \lim_{\Delta t \rightarrow 0} \frac{x(t + \Delta t) - x(t)}{\Delta t}$$

当 Δt 很小时并且取很小的 1 单位时，则近似地有 $x(t+1) - x(t) = \frac{\Delta x}{\Delta t}$

(4) 写成离散形式：

$$\frac{\Delta \mathbf{x}}{\Delta t} = \mathbf{x}(\mathbf{k} + 1) - \mathbf{x}(\mathbf{k}) = \Delta^{(1)} (\mathbf{x}(\mathbf{k} + 1))$$

(5) 由于 $\frac{\Delta \mathbf{x}^{(1)}}{\Delta t}$ 涉及到累加列 $\mathbf{x}^{(1)}$ 的两个时刻的值，因此， $\mathbf{x}^{(1)}(i)$ 取前后两个时刻的平均代替更为合理，即将 $\mathbf{x}^{(1)}(i)$ 替换成

$$\begin{aligned} & \frac{1}{2} [\mathbf{x}^{(i)}(i) + \mathbf{x}^{(i)}(i-1)], (i = 2, 3, \dots, N) \\ \mathbf{x}^{(i)} &= \frac{1}{2} [\mathbf{x}^{(i)}(i) + \mathbf{x}^{(i)}(i-1)], (i = 2, 3, \dots, N) \\ \mathbf{x}^{(1)}(\mathbf{k} + 1) &= \frac{1}{2} [\mathbf{x}^{(1)}(\mathbf{k} + 1) + \mathbf{x}^{(1)}(\mathbf{k})] \end{aligned}$$

(6) 再将下列四个式子：

$$\begin{aligned} \frac{\Delta \mathbf{x}}{\Delta t} &= \mathbf{x}(\mathbf{k} + 1) - \mathbf{x}(\mathbf{k}) = \Delta^{(1)} (\mathbf{x}(\mathbf{k} + 1)) \\ \Delta^{(1)} (\mathbf{x}^{(1)}(\mathbf{k} + 1)) &= \mathbf{x}^{(1)}(\mathbf{k} + 1) - \mathbf{x}^{(1)}(\mathbf{k}) = \mathbf{x}^{(0)}(\mathbf{k} + 1) \\ \mathbf{x}^{(1)}(\mathbf{k} + 1) &= \frac{1}{2} [\mathbf{x}^{(1)}(\mathbf{k} + 1) + \mathbf{x}^{(1)}(\mathbf{k})] \\ \Delta^{(1)} (\mathbf{x}^{(1)}(\mathbf{k} + 1)) + \mathbf{a}(\mathbf{x}(\mathbf{k} + 1)) &= \mathbf{u} \end{aligned}$$

整理成：

$$\mathbf{x}^{(0)}(\mathbf{k} + 1) = \mathbf{a} \left[-\frac{1}{2} (\mathbf{x}^{(1)}(\mathbf{k}) + \mathbf{x}^{(1)}(\mathbf{k} + 1)) \right] + \mathbf{u}$$

(7) 写成矩阵表达式：

$$\begin{aligned} \begin{bmatrix} \mathbf{x}^{(0)}(2) \\ \mathbf{x}^{(0)}(3) \\ \mathbf{M} \\ \mathbf{x}^{(0)}(N) \end{bmatrix} &= \begin{bmatrix} -\frac{1}{2} [\mathbf{x}^{(1)}(2) + \mathbf{x}^{(1)}(1)] & 1 \\ -\frac{1}{2} [\mathbf{x}^{(1)}(3) + \mathbf{x}^{(1)}(2)] & 1 \\ \mathbf{M} & 1 \\ -\frac{1}{2} [\mathbf{x}^{(1)}(N) + \mathbf{x}^{(1)}(N-1)] & 1 \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{u} \end{bmatrix} \\ \text{令 } \mathbf{y} &= (\mathbf{x}^{(0)}(2), \mathbf{x}^{(0)}(3), \mathbf{L}, \mathbf{x}^{(0)}(N))^T \end{aligned}$$

这里的 T 表示转置，令

$$B = \begin{bmatrix} -\frac{1}{2}[x^{(1)}(2) + x^{(1)}(1)] & 1 \\ -\frac{1}{2}[x^{(1)}(3) + x^{(1)}(2)] & 1 \\ n & 1 \\ -\frac{1}{2}[x^{(1)}(N) + x^{(1)}(N-1)] & 1 \end{bmatrix}, U = \begin{bmatrix} a \\ u \end{bmatrix}$$

$$y = BU$$

$$\hat{U} = \begin{bmatrix} \hat{a} \\ \hat{u} \end{bmatrix} = (B^T B)^{-1} B^T y$$

6. 预测值的求解

$$x^{(1)}(k+1) = \left[x^{(1)}(1) - \frac{\hat{u}}{\hat{a}} \right] e^{-\hat{a}k} + \frac{\hat{u}}{\hat{a}}$$

7. GM(1,1)模型精度检验

模型选定之后，，一定要经过检验才能判定其是否合理一定要经过检验才能判定其是否合理，只有通过检验的模型才能用来作预测，灰色模型的精度检验一般有三种方法灰色模型的精度检验一般有三种方法：相对误差大小检验法，关联度检验法和后验差检验法。下面主要介绍后验差检验法：

设按照 GM(1,1)建模法已经求出 $\hat{X}^{(1)}$ ，并将 $\hat{X}^{(1)}$ 做一次累减转化成 $\hat{X}^{(0)}$ ，即

$$\hat{X}^{(0)} = [\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), n, \hat{x}^{(0)}(n)]$$

计算残差：

$$\text{令 } e(k) = x^{(0)}(k) - \hat{x}^{(0)}(k), k = 1, 2, L, n$$

原始序列 $X^{(0)}$ 及残差序列 E 的方差分别为 S_1^2 和 S_2^2 ，则

$$S_1^2 = \frac{1}{n} \sum_{k=1}^n [x^{(0)}(k) - \bar{x}]^2$$

$$S_2^2 = \frac{1}{n} \sum_{k=1}^n [e(k) - \bar{e}]^2$$

$$\text{其中, } \bar{x} = \frac{1}{n} \sum_{k=1}^n x^{(0)}(k), \bar{e} = \frac{1}{n} \sum_{k=1}^n e(k)$$

计算后残差比为： $C = S_2/S_1$

其中指标 C 和 p 是后验差检验的两个重要指标，指标 C 越小越好， C 越小表示 S_1 大而 S_2 小， S_1 大表示原始数据方差大，即原始数据离散程度大。 S_2 小表示残方差小，即残差离散程度小。 C 小就表示原始数据很离散，而模型所的计算值与实际值之差并不太离散。

8.精度检验等级参照表

模型精度等级	均方差比值 C
1 级（好）	$C \leq 0.35$
2 级（合格）	$C \leq 0.5$ 且 $C > 0.35$
3 级（勉强）	$C \leq 0.65$ 且 $C > 0.5$
4 级（不合格）	$C > 0.65$

9.具体实践

（1）先用 python 算出各个自动售货机 2017 年以来每个月饮料和非饮料的交易额（具体见代码）（横轴为月份，纵轴为大类）

①A 售货机

	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
饮料	1184.7	392.5	823.7	1696.0	2866.3	5843.0	1698.5	2087.4	3966.2	5532.4	4412.5	6932.5
非饮料	575.4	79.0	278.1	820.2	1640.1	2525.2	692.7	547.4	1468.7	2482.4	1723.8	2689.7

②B 售货机

	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
饮料	1291.6	587.3	887.4	2114.9	3236.3	6567.2	1282.8	3357.1	6149.6	7400.6	7238.7	7529.9
非饮料	421.3	80.8	199.4	762.7	1462.2	2446.1	570.1	641.0	1988.3	2334.7	2349.4	1819.4

③C 售货机

	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
饮料	1394.2	726.0	926.5	2760.5	3042.1	6853.9	2718.0	4569.6	6185.5	8346.6	7098.6	8507.1
非饮料	512.6	113.0	184.8	903.4	1560.2	2791.7	946.8	732.9	1959.6	2661.1	2084.7	2116.5

④D 售货机

	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
饮料	931.9	420.5	662.6	1623.6	2001.6	3837.2	1109.6	2349.5	3406.0	4163.1	4213.2	5457.3
非饮料	476.6	59.5	231.6	631.1	1198.3	1923.7	483.8	505.7	870.4	1588.9	1066.3	1652.4

⑤E 售货机

	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
饮料	1342.8	879.3	1457.5	3232.3	5184.2	9162.3	2849.4	6415.5	15196.9	9791.8	18980.9	12438.7
非饮料	403.7	137.0	384.5	1100.8	2270.5	3698.6	957.8	1304.0	4344.3	3213.1	5591.5	3597.4

(2) 利用灰色预测法预测 2018 年一月各个自动售货机每个大类的交易额（具体见代码）（模型的优劣相应用好，合格，勉强，不合格代替）（上面是数值，下面是模型的优劣）

	A 售货机	B 售货机	C 售货机	D 售货机	E 售货机
2018 年一月饮料预计销售量	7659.2 合格	10420 好	11149.5 好	6482.6 好	22063.2 好
2018 年一月非饮料预计销售量	2915.9 勉强	2795.2 勉强	2880 勉强	1716.1 勉强	6065.7 合格

分析：总体来说，饮料的灰色预测模型总体上是好的，而非饮料的灰色预测模型总体上并没有那么好，但也没有不合格。但是根据这些数据预测出来 2018 年一月的数据并不是很准确，很大部分源于 9 月到 12 月的数值上升引起的 2018 年 1 月的数据上升，而根据各地 2017 年一月的销售数据来看，2018 年 1 月的销售数据不太可能达到这么高。因此，该灰色预测模型并不能很好的预测 2018 年 1 月的数据，应该提供前几年一月份的数据方可预测 2018 年 1 月的数据。因此，如果想要更好的预测 2018 年 1 月的数据，应该给予前几年 1 月的数据。

七、小结

总的来说，分析可以得出以下结果：

1. 站点选择：根据分析来看，A 地和 D 地的各项指标都偏低，特别是 D 地的指标

严重偏低，B 地和 C 地的各项指标比较均衡，而 E 地的各项指标都很高，属于具有地方优势的站点。对于 A 地和 D 地的自动售货机指标偏低的情况，可能是位置偏僻或者顾客对自动售货机的商品不太感兴趣，因此可以适当的进行商品的更换或者促销策略以吸引顾客的兴趣等。而对于 B 地和 C 地比较均衡的地方，可以适当的取优补短，多入销量好的商品。对于 E 地的自动售货机各项指标都高，说明该地的顾客对自动售货机的商品兴趣较大，应该继续保持优势，甚至可以多保持当地地方优势，扩大自动售货机的建设。

2. 种类选择：根据分析来看，各地的销量前几的种类中或多或少有些不一样，但是大部分相同，主要销售前几的种类还有怡宝纯净水，脉动，东鹏特饮，阿萨姆奶茶，营养快线等饮料，还有无穷鸡小腿，无穷盐鸡蛋，双汇火腿肠等非饮料产品，在入货选择方面可以多入这几种销量前几的商品，饮料类也相对于非饮料盈利和销量都偏高，可以多入饮料类。而销售后几的主要是面包，饼干，咖啡等，相应的地方还有可口可乐，雪碧等不受顾客的欢迎，在入货时可以少入或者下架销量不好的产品，或者实施促销策略进行降价等提高关注度。同时，在饮料容量方面也有要求，部分地方大容量饮料无人问津，小容量饮料很受顾客欢迎，入货时可以进行选择。

3. 时间选择：根据分析来看，各地销量比较高的月份都不太一样，但是总的来说，各项指标偏低的月份主要是 2 月份和 7 月份，因此可以对 2 月份和 7 月份的商品进行促销活动，特别对销量前几的可以多入货，在这两个月多进行促销，提高各地自动售货机的各项指标，在劣势的地方加大促销，可以整体的提高各地自动售货机的各项指标。而相对上升和各项指标偏高的主要是 6 月和 8 月到 12 月，这是自动售货机畅销的几个月，可以加大商品的投入或者进行小幅度的促销策略，扩大时间选择上的优势，大幅度提高整体的各项指标。据热力图可以看到，普遍一天集中销售的时间主要是 12 点到 18 点，因此可以选择 12 点到 18 点进行促销活动，可以在顾客热衷的时间段多增加营业额。

总的来说，应该在受顾客欢迎的时间段、地方和商品上多进行促销和入货，根据

用户情感分析的词云图了解不同地方的顾客对产品的需求,在不同地方不同时间实行不同的营销策略,实现利益最大化。