

新零售—无人智能售货机商务数据分析报告

报告人： 吴逸阳

报告时间： 2019-8-22

目录

- 一. 背景介绍.....2
 - 1. 报告背景.....2
 - 2. 分析目标.....2
- 二. 数据预处理.....2
 - 1. 数据来源.....2
 - 2. 数据处理 (Task 1-1).....2
- 三. 品牌销量分析.....3
 - Task 1.....3
 - Task 1-2.....3
 - Task 1-3.....3
 - Task 2.....4
 - Task 2-1.....4
 - Task 2-2.....4
 - Task 2-3.....7
 - Task 2-4.....9
 - Task 2-5.....9
- 四. 用户情感分析.....11
 - 1. 商品售货机标签画像 (Task 3)11
 - 2. 销量预测 (Task 4)13

一. 背景介绍

1. 行业背景

随着时代的发展和科技的创新，突破了时间和空间局限性的无人售货机越来越受人们的欢迎。自动售货机结合了线上和线下销售的经营模式，提供了便利便民的服务，通过自助的销售模式节约了人工成本和时间成本，成为当下零售经营的又一主流模式。

对于自动售货机运营者而言，自动售货机里面商品的供给频率、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，我们需要通过数据分析来帮助经营者掌握了解用户需求和商品需求量等等，为用户提供精准贴心的服务，是掌握经营方向的重要手段，对自动售货机这以营销模式的发展有着非常重要的意义。

2. 分析目标

题目中所给出的数据里包括了商场在不同地点安放的 5 台自动售货机（编号分别为 A、B、C、D、E）。在这两份数据中，包括了每台售货机中每件商品的支付金额和实际金额、支付时间等等，以及每类商品的分类依据。结合两者的数据，我们需要从以下几个方面进行分析：

- 1) 对数据进行处理，提取出商品的销售数据和市场需求进行分析。
- 2) 统计每一台售货机的商品销售数据并绘制所需图表。
- 3) 进行用户情感分析，对每一台售货机绘制相应的标签画像。
- 4) 对统计完的数据进行拟合，预测来年的销售量情况。

二. 数据预处理

1. 数据来源

题目中所给的数据分为两为附件，格式为 csv，分别为:附件 1、附件 2。

第一个附件包括了商场中所有售货机的订单信息，包括了支付时间、实际金额等等。（总共有 70679 条信息）

第二个附件包括了每一样商品所属类别，第一大类（饮料和非饮料）和二级品。（总共有 315 条商品信息的分类）

2. 数据处理

数据处理方法：python（包括 pandas、numpy、matplotlib 库）、excel

数据预处理过程：(Task 1-1)

1. 利用 python 将附件 1 导入到 dataframe 中，并用 pandas 库对数据进行异常值和去重处理。（此处剔除了 C 售货机一条支付时间格式异常的数据）将所有的支付时间格式转化成 python 可读取模式。
2. 利用 pandas 中的 groupby 函数对售货机的地点进行分类，导出以‘task1-1+地点’为名称的 csv 文件中。
3. 利用 excel 文件的 vlookup 函数对每个商品进行大类、二级品的划分。前期数据预处理完成。

三. 品牌销量分析

Task 1

Task 1-2

运用 python 软件统计得出的结果如图所示。

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。

	交易额订单量	交易金额
A	756	3385.1
B	869	3681.2
C	789	3729.1
D	564	2392.1
E	1292	5699
所有售货机	4270	18886.8

2017 年 5 月每台售货机交易额订单量数据

Task 1-3

运用 python 软件统计得出的结果如图所示。

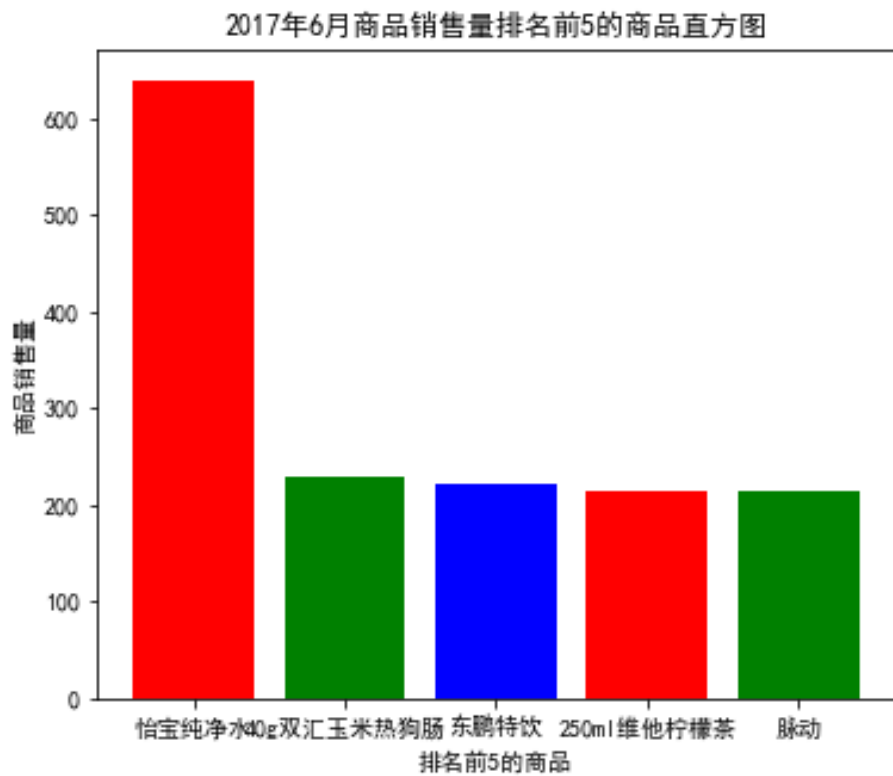
以下是每台售货机每月的每单平均交易额与日均订单量（用黄色底标出）。

	A	B	C	D	E
1 月	4.569	3.798	4.434	3.804	4.723
2 月	3.870	3.268	3.823	3.089	3.640
3 月	3.375	3.595	3.756	3.437	4.300
4 月	4.034	4.117	4.371	3.791	4.1654
5 月	4.478	4.236	4.727	4.241	4.411
6 月	4.088	4.079	4.544	4.025	3.814
7 月	4.119	4.402	3.979	4.230	3.907
8 月	3.360	3.604	3.916	3.328	3.808
9 月	4.320	4.184	4.406	3.828	4.127
10 月	4.017	4.111	4.267	3.888	3.647
11 月	4.476	4.273	4.356	3.865	4.288
12 月	3.784	3.661	3.946	3.586	4.186
日均订单量	28.73	36.94	39.71	23.87	64.40

2017 每台售货机平均交易额和日均订单量

Task 2

Task 2-1: 绘制 2017 年 6 月销售量前 5 的商品销量柱状图。



Task 2-2 绘制每台售货机每月总交易额折线图



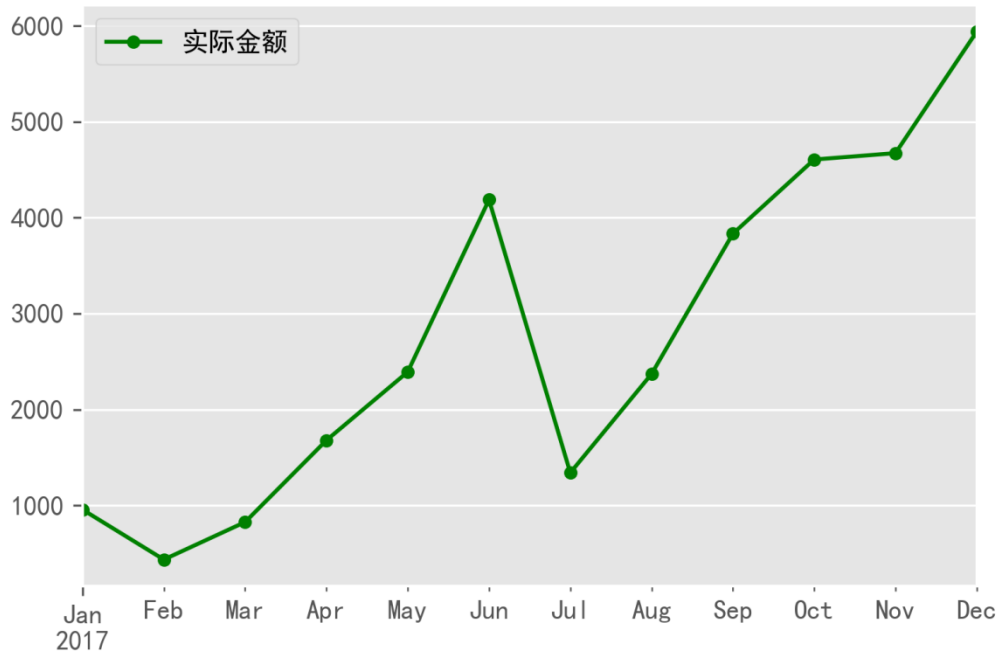
A 售货机



B 售货机

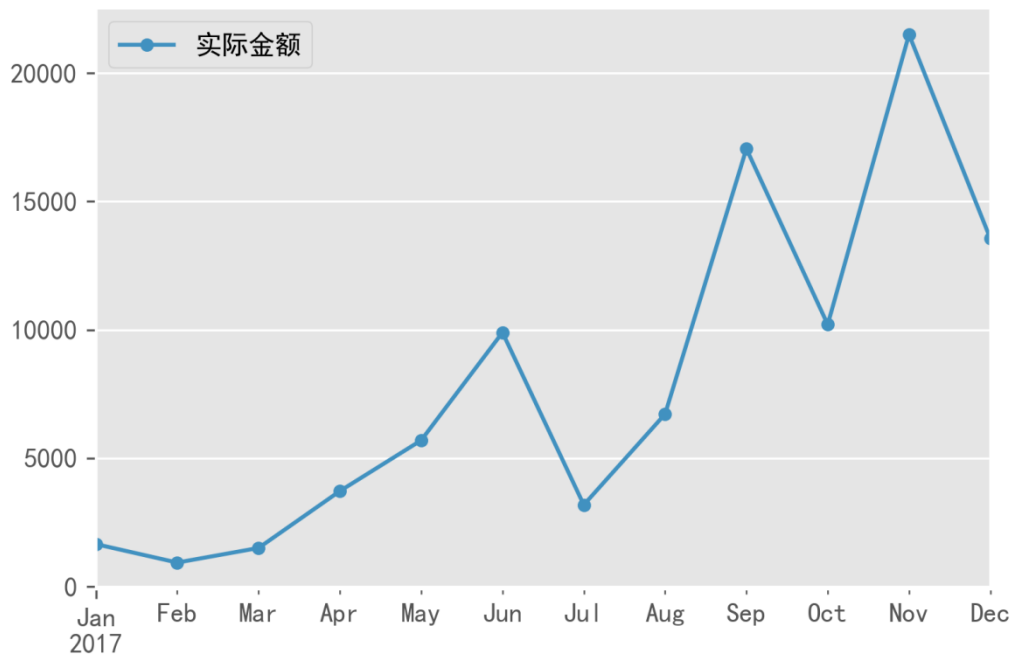


C 售货机



支付时间

D 售货机

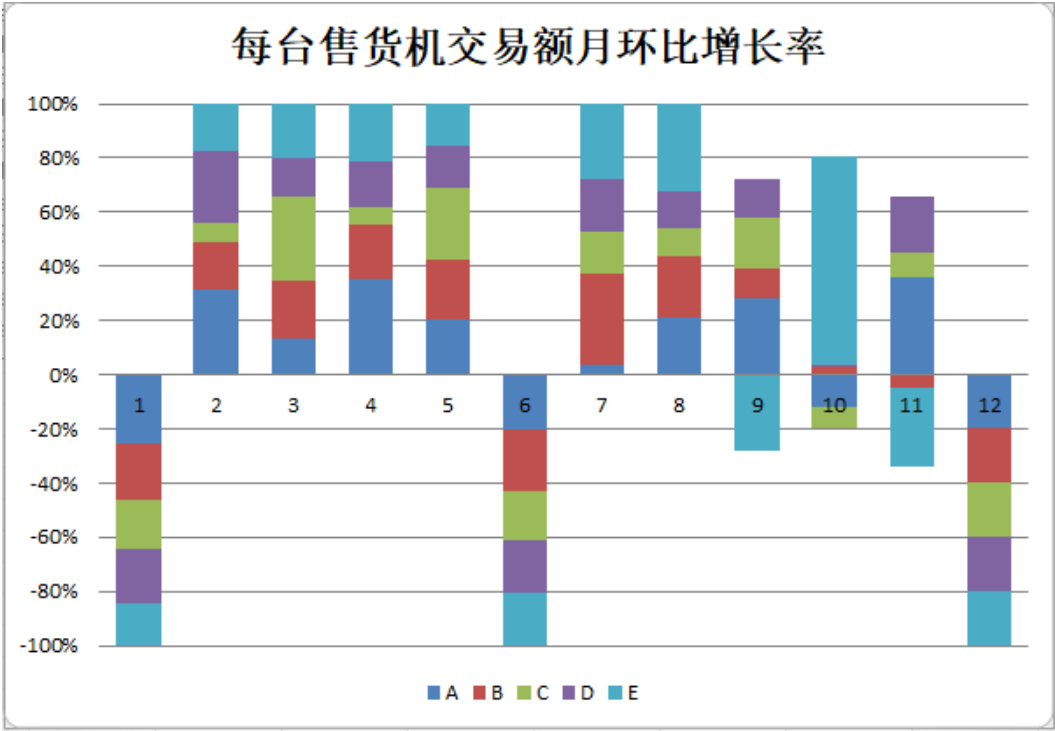


支付时间

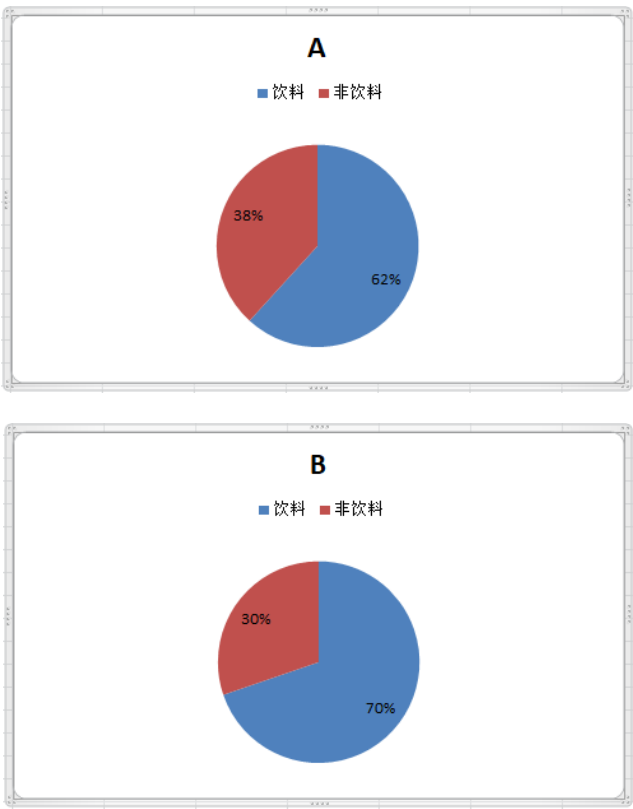
E 售货机

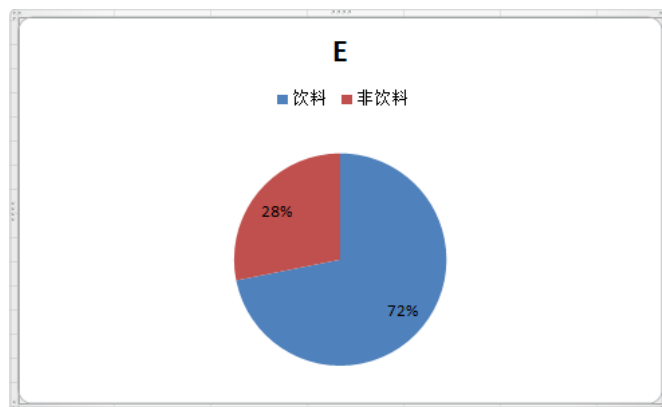
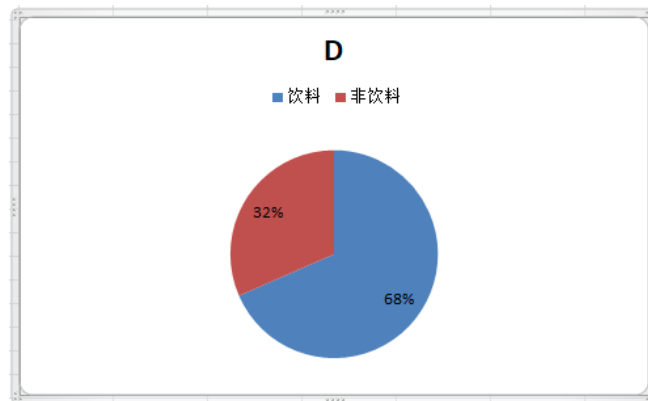
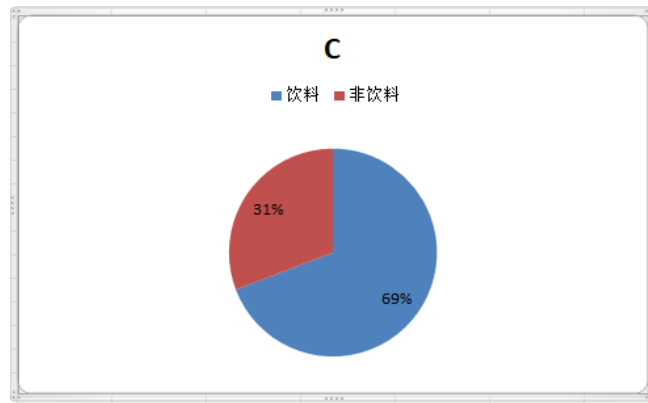
过程: python
详见 task2 附件

Task 2-2 交易额月环比增长率柱状图



Task 2-3 绘制每台售货机毛利润占总毛利润比例的饼图（分为饮料类和非饮料类）

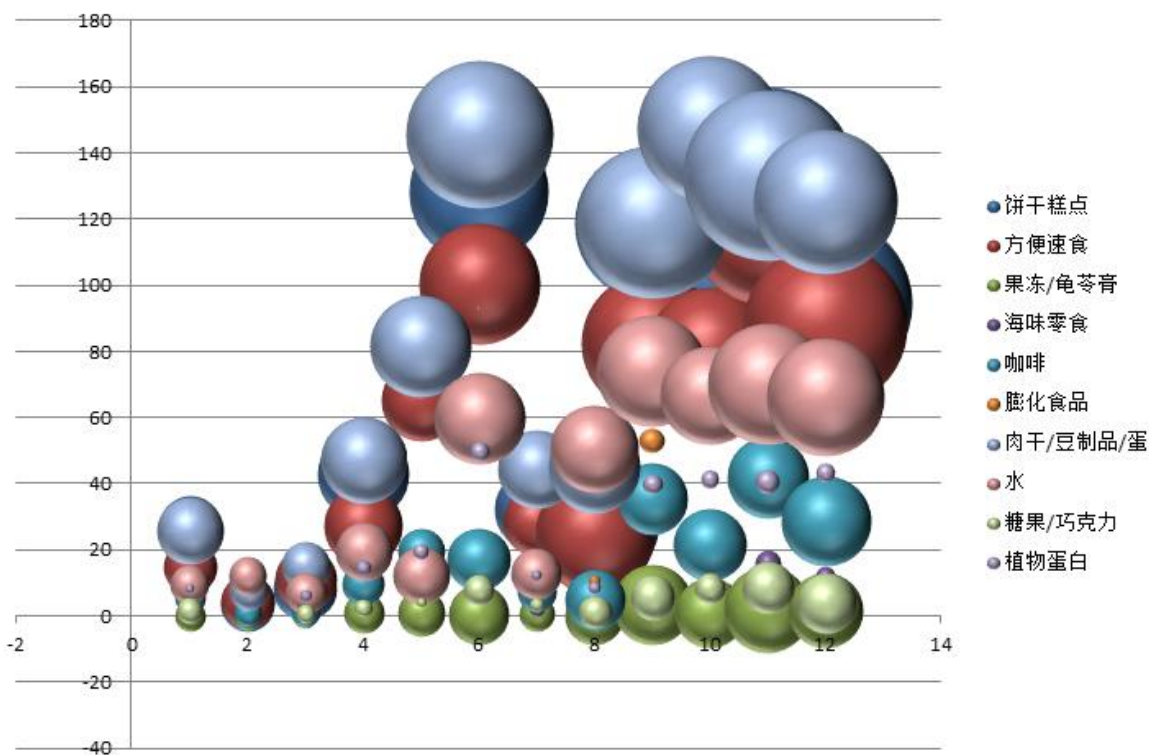




使用工具：excel

制作过程：将数据运用筛选进行分类后，选取公式计算毛利率。最后利用 excel 自带的饼图完成。

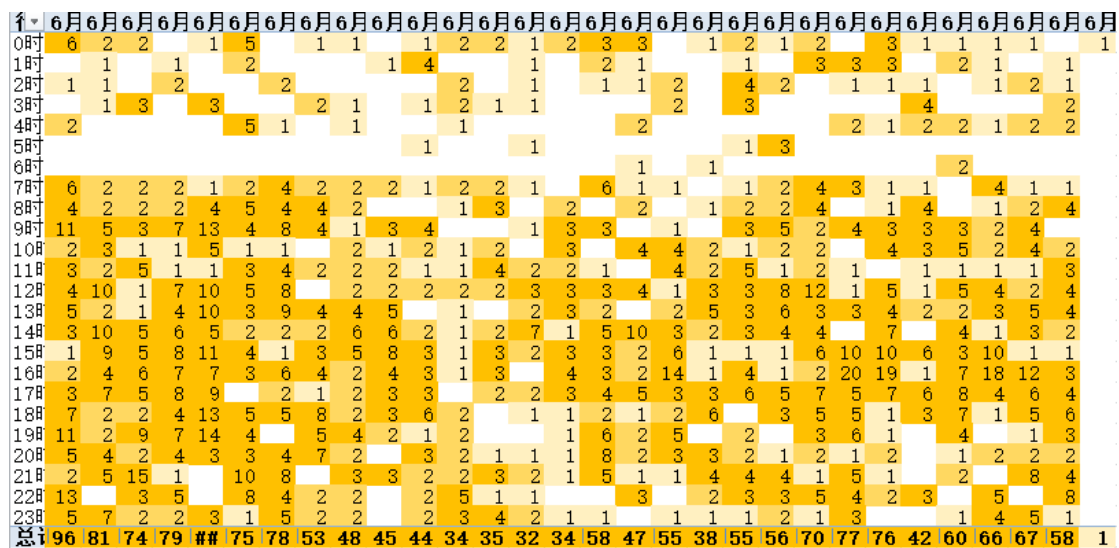
Task 2-4 绘制每个月交易额均值的气泡图



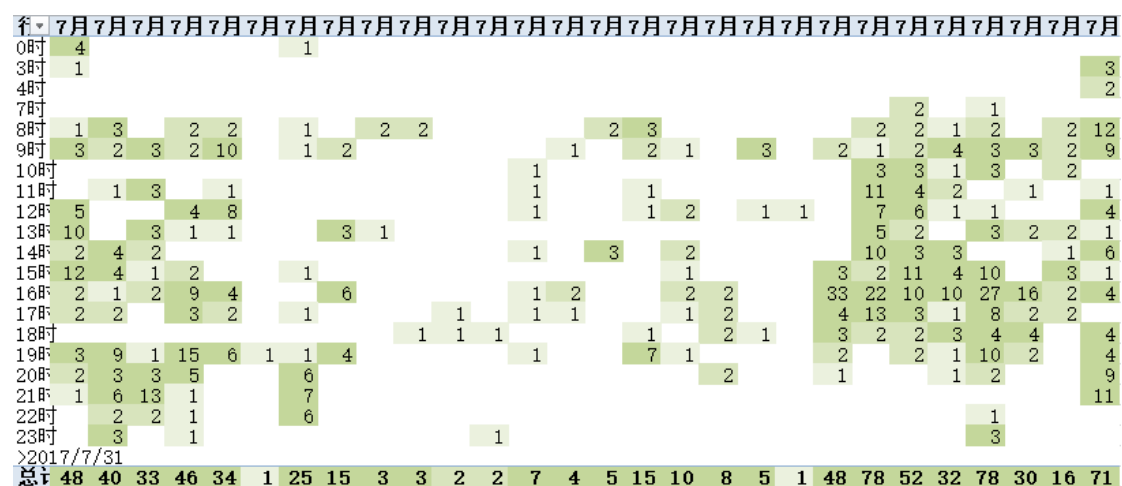
制作工具: excel

制作过程: 利用筛选对二级品进行分类。运用公式对数据进行统计, 用 excel 自带的气泡图进行制作。

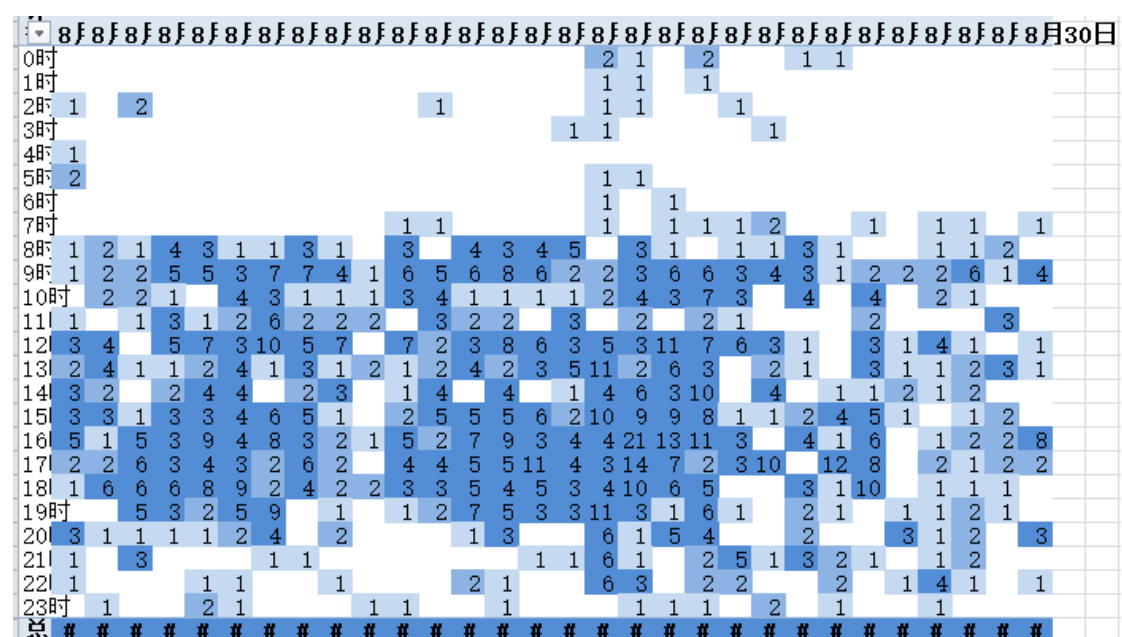
Task 2-5 绘制售货机 C 6、7、8 月订单量的热力图



售货机 C6 月份热力图



售货机 C7 月份热力图



售货机 C8 月份热力图

制作工具: excel

制作过程: 运用透视表对数据进行分组, 运用条件规则设置色阶的规则, 填充色阶得到热力图。

得到结论:

- 1) 对于售货机 C 而言, 6、7、8 月份的商品售卖中, 7 月份售卖的数量是最低的, 6 月份销售量介于二者之间, 8 月份的销售量最多。由此可见, 八月份正当酷暑季节, 人们前来购买的数量会相较于往常而言更高。而 7 月份可能由于各种原因导致销售量遭遇下滑。
- 2) 同时对于一天 24 小时的情况来说, 下午到傍晚时间段的销售量是最多的, 上午的销售量虽然不敌下午, 但是还是有正常足够的销售量。而半夜到凌晨购买的人数是最少的。
- 3) 对于售货机 C, 我们可以在中午时间进货, 这样下午就有充足的商品数量供消费者购买, 同时八月份可以考虑进更多的货、上新产品等等, 这样可以满足消费者购买量大的需求。

三. 用户情感分析

1. 商品售货机标签画像 (Task 3)

运用软件: excel

分类过程: 对所有的商品标签进行整合归合并

- 1) 对商品所有的名称进行规范。

统一同一种商品的名称, 如合并 100g*5 瓶益力多和益力多 100g*5 瓶。

统一同一种商品的不同规格, 并作大类。如 140ml 和 90ml 的销售量都算到该商品的总的销售量中。

对于同一种商品的不同口味, 算入总的商品销量中。

- 2) 利用数据透视表统计归类对售货机的地点和商品分类, 计算订单量的多少。得到热力图。(如图所示)

行标签	A	B	C	D	E
UFO炒面	7	4	20	6	32
阿萨姆奶茶	418	560	624	417	999
安慕希	62	82	60	39	104
奥利奥夹心小饼干	79	73	111	87	203
澳门永辉鸡仔饼	11	13	26		13
百事可口可乐	40	106	17	37	16
宝丰龙龟苓膏	4	12	4	13	41
宝矿力	4	7	4	1	7
冰红茶	234	340	404	311	628
冰糖雪梨	39	59	59	16	143
出前一丁桶面酱香牛肉王	12	3	4	3	13
达利园蔓越莓提子面包	1				
达利园柠檬蛋糕			1		2
德芙巧克力	2				1
东鹏特饮	464	515	545	436	758
多芙利香草奶油味蛋糕	6	2	1	1	6
多力多玉米片爆香	7	5	4		14
多力多滋	1		2		
费列罗榛果威化巧克力	23	10	36	9	34
芬达	47	198	271	9	382
凤爪	9	21	19	15	16
芙蓉王	20	43	34	8	6
富甲一方紫薯仔	39	34	23	25	70
钙芝奶酪味高钙威化饼干	3				
格力高百力滋饼干	1				
公仔迷你杯	12	18	20	4	34
帝豪苹果数据线	1			1	

部分数据如图所示

(颜色越深表示订单量越多)

- 3) 对每一个商品的所计算出来的销售量观察分析并定义。以 A 售货机为例, 下相同。对数据进行横向和纵向观察比较对于同一台售货机的不同商品而言订单量最小值 0, 最大值 418。
- 4) 同时横向比较 ABCDE 五台售货机, 如果销售量相比于舞台售货机同类商品平

均销售量高，可定义为热销。以此类推。（例如：C 售货机中的珠江啤酒在同类中销售最高，故定义为热销。）

- 5) 因此得到总的定义标签，数据小于 30 定义为滞销，50-150 定义为正常，150 以上定义为热销。
- 6) 保存为 csv 文件。（详见附件）

得到部分售货机画像如图所示

序号	饮料类商品	标签
1	阿萨姆奶茶	热销
2	安慕希	正常
3	百事可口可乐	滞销
4	宝矿力	滞销
5	冰红茶	热销
6	冰糖雪梨	正常
7	东鹏特饮	热销
8	芬达	热销
9	果粒橙	热销
10	恒大冰泉矿泉水	滞销
11	红牛	热销
12	加多宝	热销
13	尖叫	正常
14	健能酸奶	正常
15	经典奶茶	滞销
16	康师傅冰糖雪梨	滞销
17	康师傅矿泉水	正常

销售意见：

- 1) 对于定义为热销的产品，可以尝试多进货，多补货。以免造成供不应求而降低销售量的情况。
- 2) 对于在不同的售货机中定义标签不同的产品，可以尝试采用不同的政策方法。
- 3) 对于定义为滞销的产品，可以缩减为仅有 1-2 两台进行售卖。（可以进行横向比较，选取同类商品中销售量最高的售货机上架）

2. 销量预测（Task 4）

预测未来销售额的原理是首先可以根据之前所拥有的数据进行整合和归类。然后根据得到的数据建立相应的模型，得到所要预测的数据。

在这里我们采用了线性拟合的方法得到每个售货机的相应公式。再带入所要求的月份，就可以得到数据。

使用方法： excel

预测过程： 1) 先得到每台售货机每个大类关于每个月的销量数据统计。

	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	717.3	313.5	557.6	940.3	1353.2	3748.8	1106.8	1586.5	2782.7	3345	3104.8	4424.8
非饮料	792.4	127	356.7	864.2	2031.9	3006.3	843.7	650.4	1696.8	2947.4	2082.2	3162.3

A 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	870.3	506.5	688	1628.6	1977.5	4491.1	5341.5	7207.2	9026.9	6283.4	6087.8	6561.5
非饮料	503.3	95.8	269.9	882.6	1757.5	3059.2	5159.3	5291.1	7162.6	3480.2	4014.1	2974.6

B 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	919.6	624	780.7	2137.2	1709	4783.6	1849.7	3901.7	4653.9	6121.5	5465.9	6608.1
非饮料	720.9	168	210.8	1095.1	2020.4	3688.6	1197.4	1025.5	2775.1	3348.2	2990.8	2772.4

C 售货机

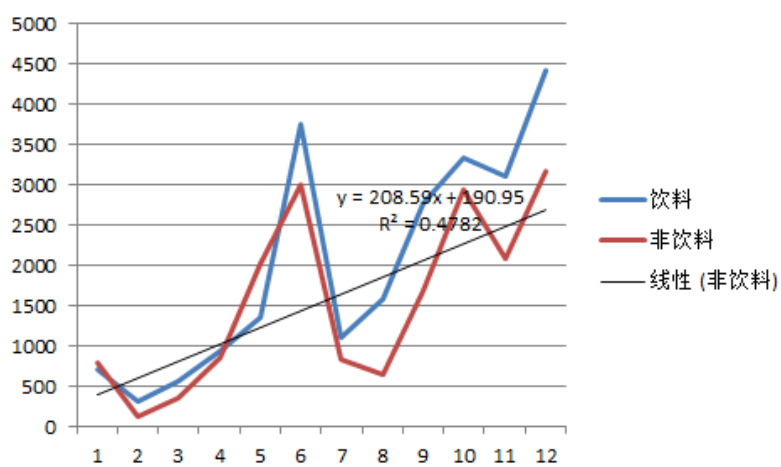
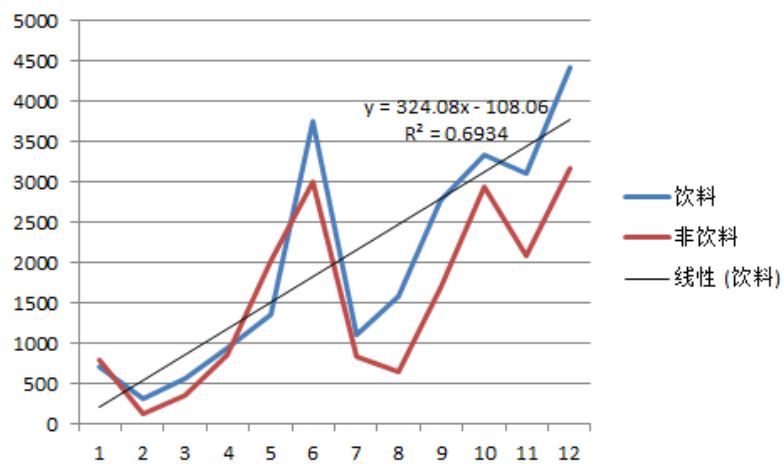
	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
饮料	479.8	361	503.5	1048	1066.8	2100.3	770.8	1854.6	2748.9	2859.2	3333.3	3967.4
非饮料	476.6	74.5	323.2	631.1	1325.3	2086.7	570	516.7	1084.2	1747.5	1340.1	1937.8

D 售货机

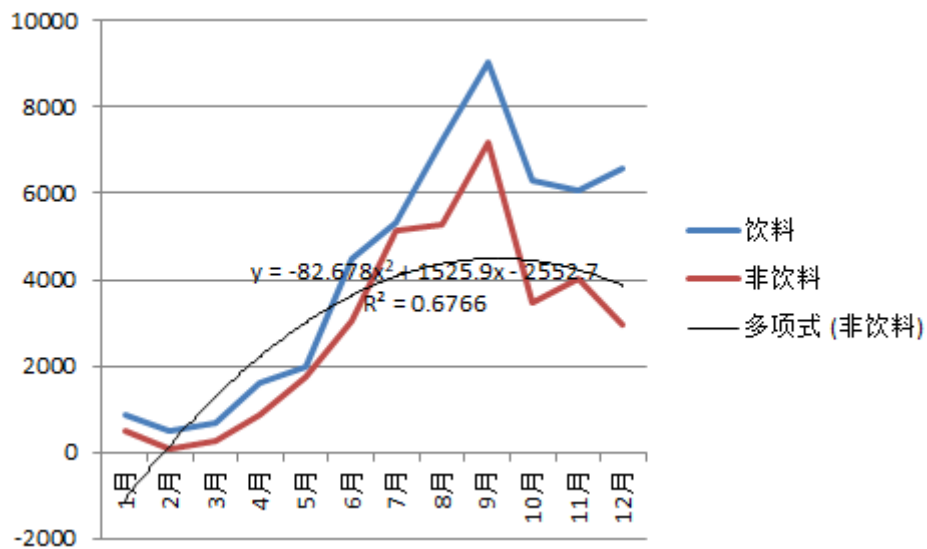
	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
饮料	1201.1	742.3	1107.5	2349.5	3197.2	5873.3	2004.6	5276	11803	6681.7	14650.5	9316.3
非饮料	455.7	196.4	399.5	1373.6	2501.8	4026.4	1181.8	1446.5	5251.3	3526.9	6851.3	4241.2

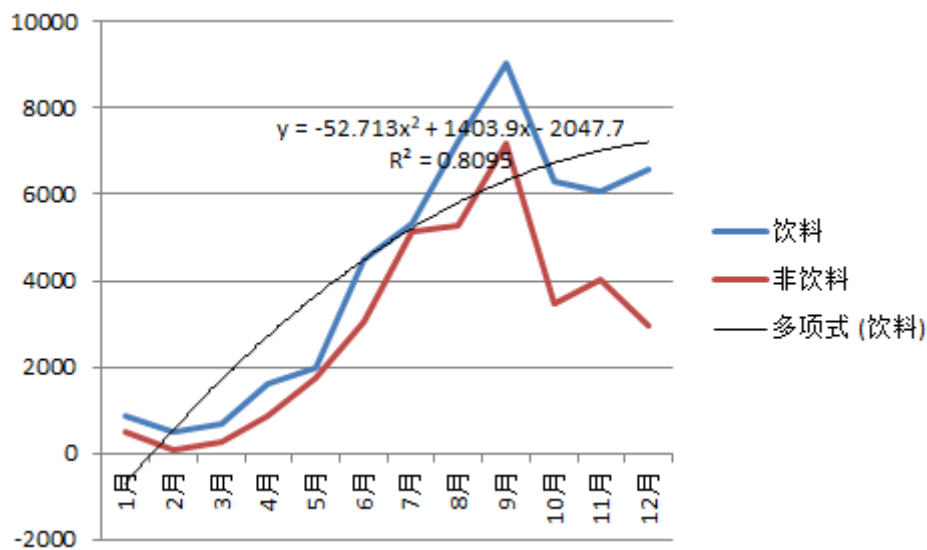
E 售货机

2) 在 excel 中运用线性拟合，得到有关于商品销量的数据公式。

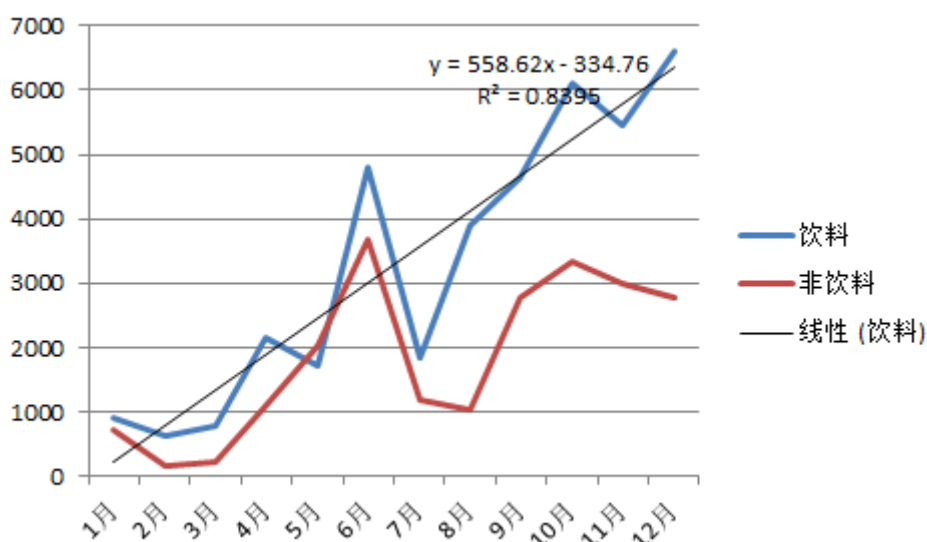


A 售货机

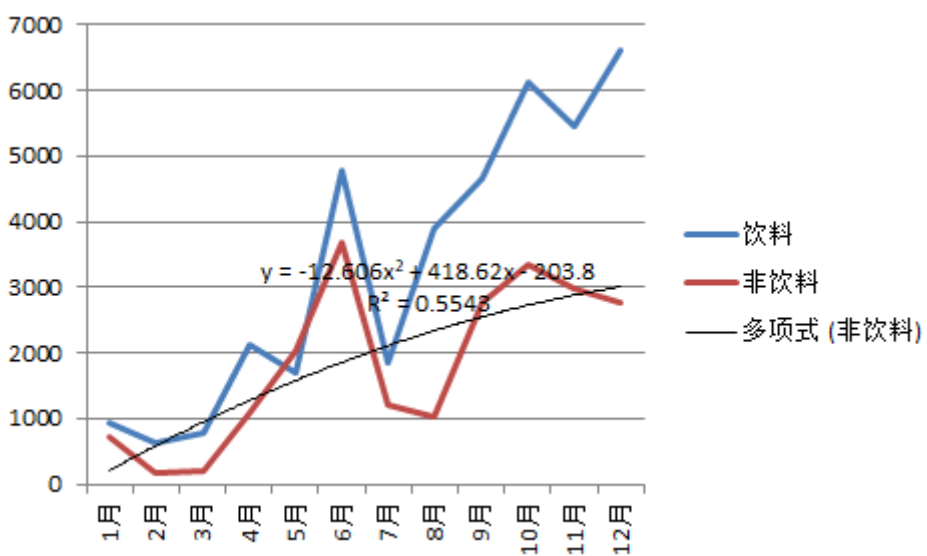


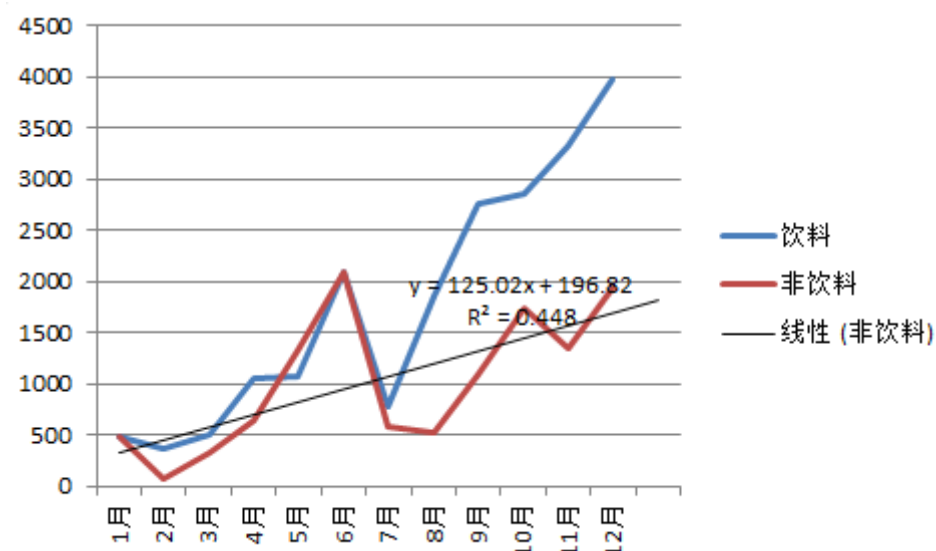
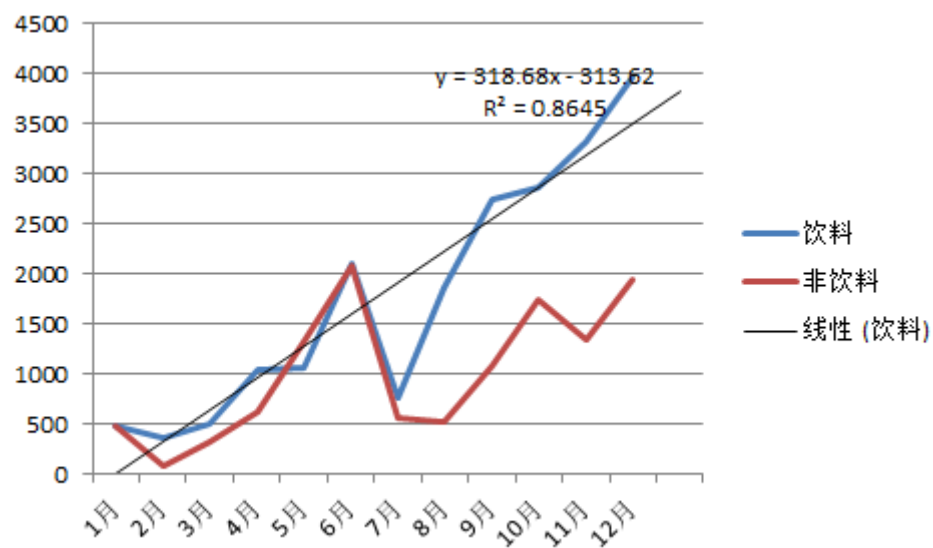


B 售货机

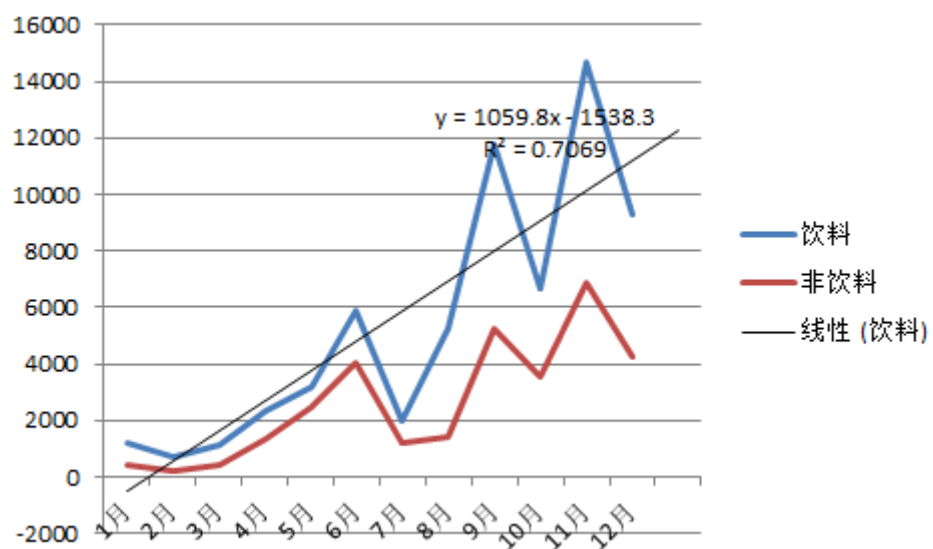


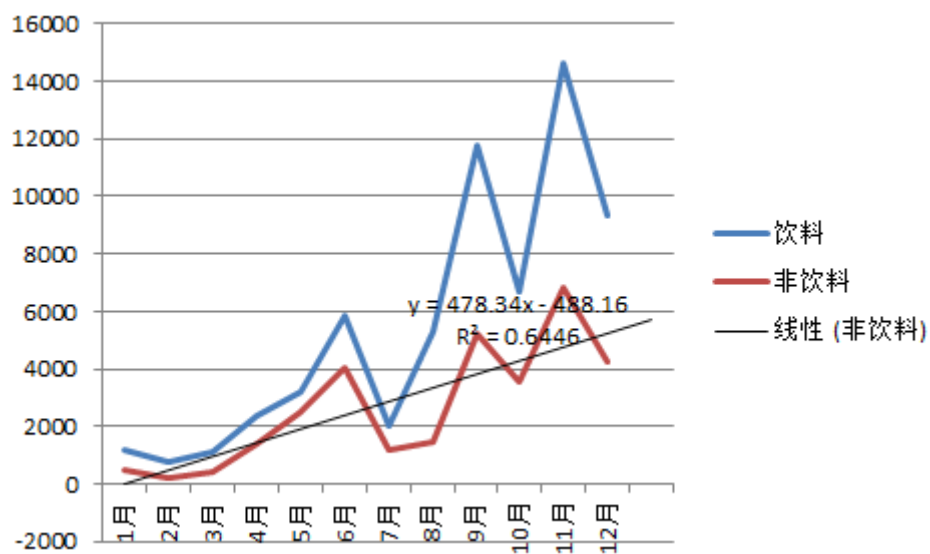
C 售货机





D 售货机





E 售货机

图中在黑色的预测线段中上方的公式就是通过线性拟合得到的预测销售量的公式。带入2018年1月份（转化为自变量为13），可以得到相应的售货机和饮料预测数据。

	A	B	C	D	E
饮料类	4104.98	16043.83	6927.3	3829.2	12239.1
非饮料类	2902.62	15412.31	5049.17	1822.08	5730.26

2018年1月份的销售总额预测