

目 录

一、	背景介绍	2
二、	数据预处理和分析	3
1、	数据预处理.....	3
2、	数据简单统计分析.....	3
三、	各售货机交易额及销量分析	5
1.	2017 年 6 月销量前五的商品.....	5
2.	各售货机每月总交易额及月环比增长率情况.....	6
3.	各售货机毛利润占总毛利润的比例.....	8
4.	每月交易额均值气泡图.....	9
5.	售货机 C 6、7、8 月订单量的热力图.....	9
四、	各自动售货机饮料类商品销量分析	12
1.	自动售货机饮料类商品的图像绘制及营销意见.....	12
五、	业务预测	15
1.	预测未来销售额的原理.....	15
2.	时间序列模型的介绍.....	15
3.	预测的具体步骤.....	16
4.	预测过程.....	16
六、	小结	20

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

- (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，绘制相应图表及预测每台售货机 2018 年 1 月的商品销量。
- (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。

3. 数据来源

某商场在不同地点安放了 5 台自动售货机，编号分别为 A、B、C、D、E 。所有数据均从这 5 台自动售货机采集而来。

4. 数据采集

收集得来的数据包括:订单号、设备 ID、应付金额、实际金额、商品、支付时间、地点、状态、提现状态。

5. 数据量

商品信息：70680 条记录。

二、数据预处理和分析

1、数据预处理

首先利用 pandas 库将数据导入 python 中，检查是否存在重复值、缺失值及异常值，最终发现存在一行日期错误的数据。将异常数据清洗后，对数据进行简单的统计分析。

2、数据简单统计分析

(1) 计算每台售货机 **2017 年 5 月份**的交易额、订单量及所有售货机总额和订单总量。

售货机	交易额（元）	订单量
A	3385.1	756
B	3681.2	869
C	3729.4	789
D	2392.1	564
E	5699	1292
SUM	18886.8	4270

(2) 计算每台售货机每月的日平均交易额与日均订单量

A 售货机			B 售货机		
	日均交易额	日均订单量		日均交易额	日均订单量
Jan	50.32	11	Jan	45.79	12
Feb	14.68	4	Feb	20.08	6
Mar	30.48	9	Mar	31.93	9
Apr	60.15	15	Apr	81.91	20
May	112.84	25	May	122.71	29
June	225.17	56	June	251.68	62
July	65.02	16	July	50.62	12
Aug	74.56	22	Aug	117.20	33
Sep	149.32	35	Sep	240.24	58
Oct	209.75	52	Oct	277.72	68
Nov	172.90	39	Nov	289.00	68
Dec	252.90	67	Dec	270.14	74

售货机 C

	日均交易额	日均订单量
Jan	54.68	13
Feb	26.40	7
Mar	33.05	9
Apr	107.74	24
May	124.31	26
June	282.41	63
July	101.57	25
Aug	164.24	42
Sep	247.63	56
Oct	315.66	74
Nov	281.89	65
Dec	312.68	79

售货机 D

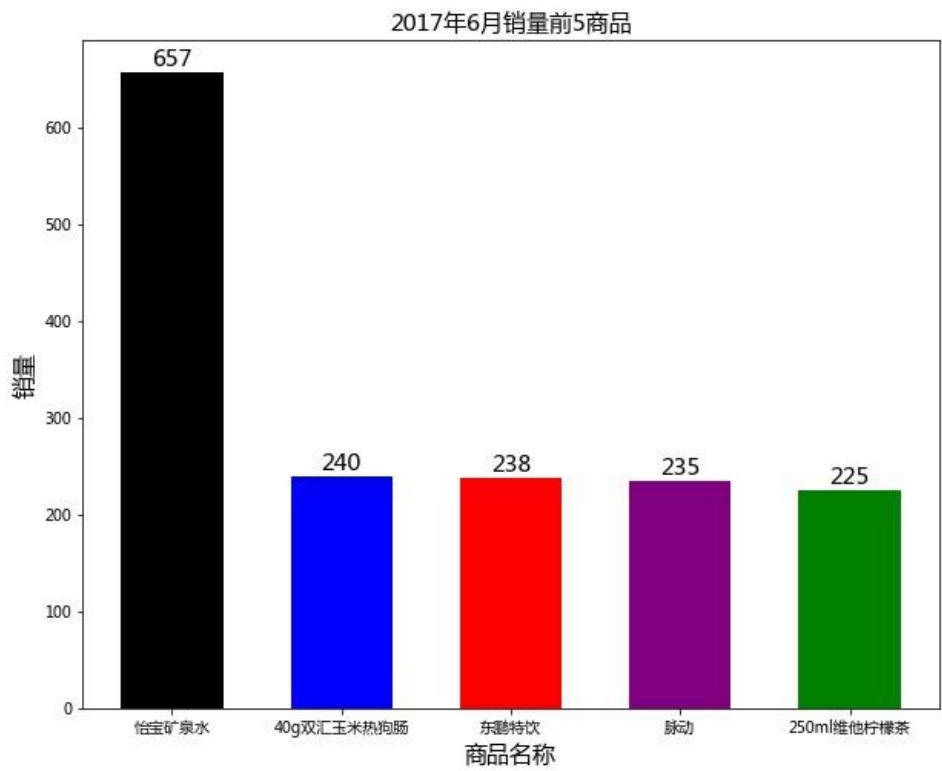
	日均交易额	日均订单量
Jan	31.88	9
Feb	14.52	5
Mar	27.56	6
Apr	55.97	15
May	79.74	19
June	139.57	35
July	44.69	11
Aug	79.04	24
Sep	127.77	33
Oct	153.56	40
Nov	155.78	40
Dec	198.04	55

售货机 E

	日均交易额	日均订单量
Jan	55.23	12
Feb	31.29	9
Mar	50.23	12
Apr	124.10	30
May	189.97	43
June	329.99	86
July	106.21	27
Aug	224.08	59
Sep	568.48	138
Oct	340.29	93
Nov	716.73	167
Dec	451.92	108

三、各售货机交易额及销量分析

1. 2017 年 6 月销量前五的商品

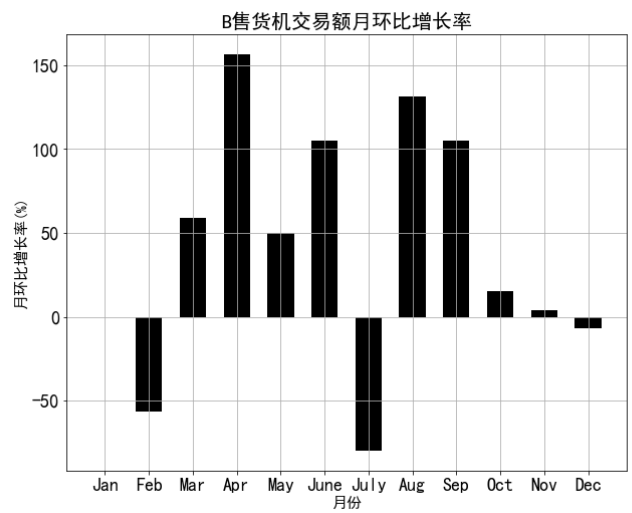
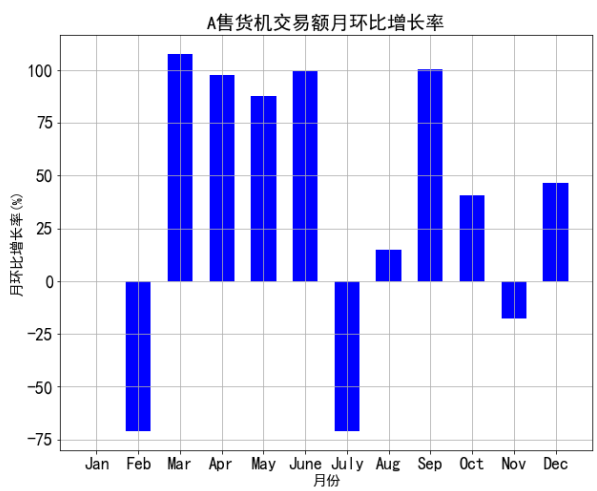
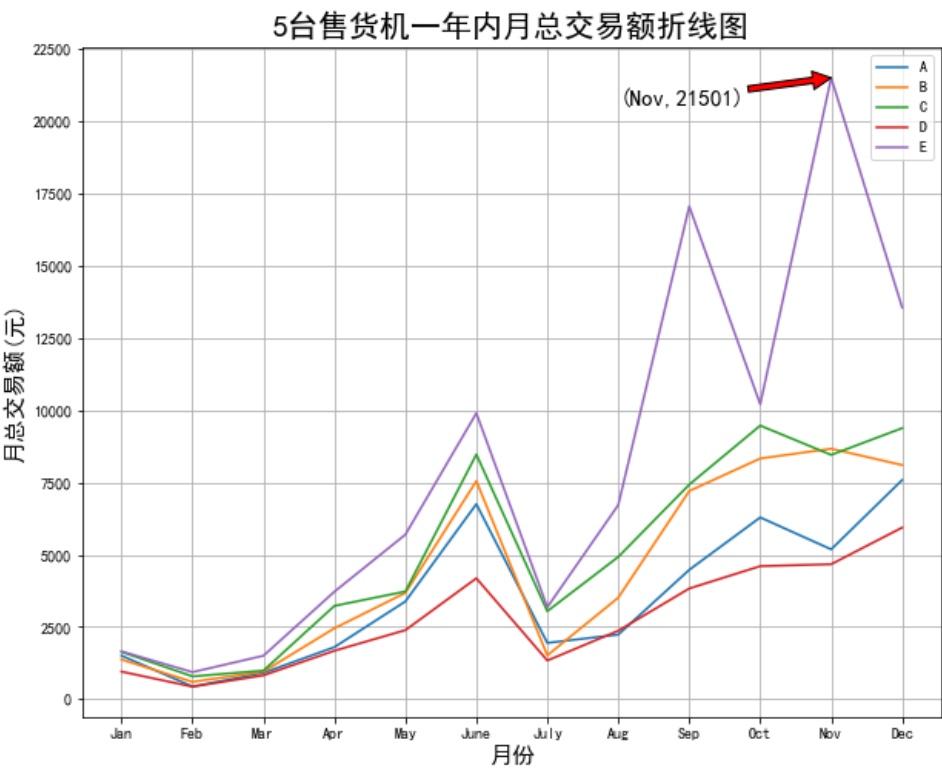


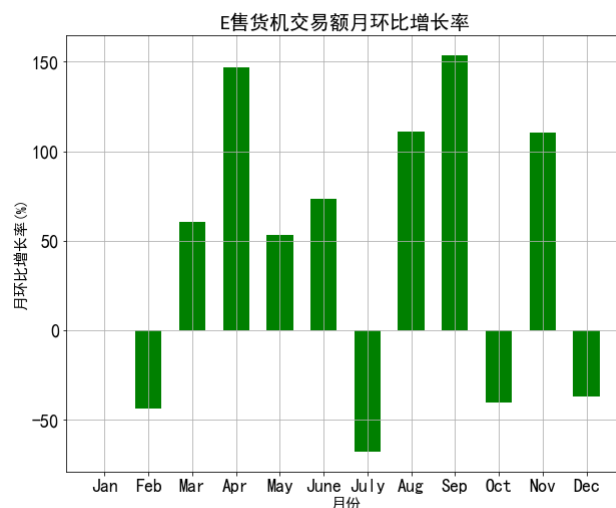
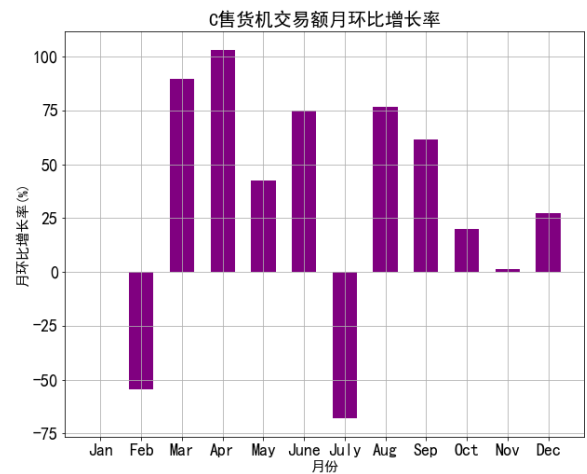
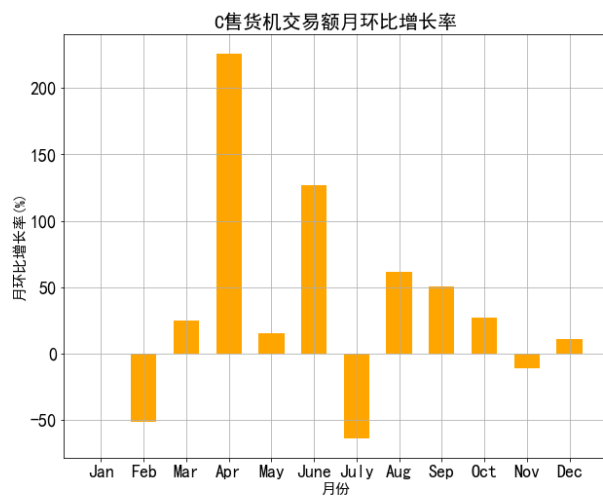
商品名称	怡宝矿泉水	40g 双汇玉米热狗肠	东鹏特饮	脉动	250ml 维他柠檬茶
销量（件）	657	240	238	235	225

表 1 售货机销量前五商品数据

2. 各售货机每月总交易额及月环比增长率情况

(1)绘制各售货机每月交易额折线图及交易额月环比增长率柱状图。

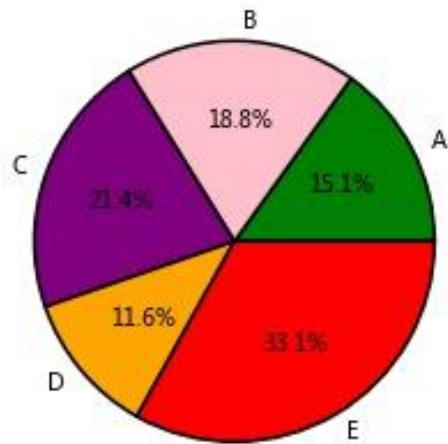




(2) 5 台售货机的月总交易额在一年内都呈相同趋势，且都在 6 月到 7 月之间有比较大的下降。由交易额月环比增长率图可知，每台售货机 7 月的交易额对比 6 月都下降了超过 50%。但 E 售货机在 7 月份过后交易额上升十分迅速，曾在 11 月份达到最高月交易额 21501 元，并在 12 月份以 13557 元月交易额遥遥领先于其他 4 台售货机。D 售货机的月交易额一直处于垫底状态，而其他 3 台售货机的月交易额则相差不大。

3. 各售货机毛利润占总毛利润的比例

每台售货机毛利润占总毛利润饼图

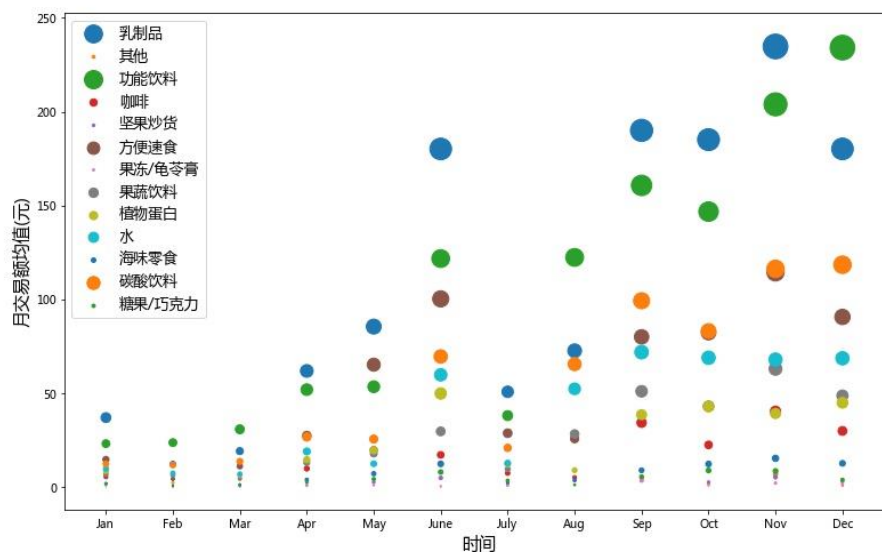


售货机	饮品交易额	非饮品交易额	售货机总交易额	毛利润(饮品+非饮品)
A	23981.3	18561.3	42542.6	9436.585
B	34995.2	18975.1	53970.3	11742.815
C	39554.9	22013.2	61568.1	13414.28
D	21093.6	12149.7	33243.3	7256.145
E	64203	31452.4	95655.4	20703.7
SUM	183828	103151.7	286979.7	62553.525

表 2 各售货机饮品交易额与毛利润数据

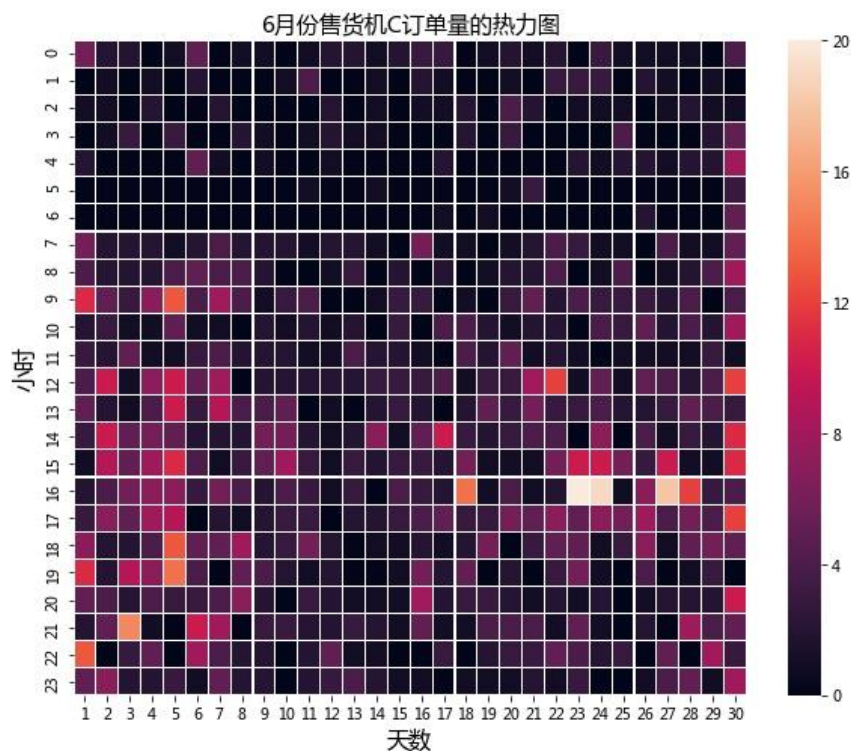
由上图可知，5 台售货机的总毛利润为 **62553.525** 元，而售货机 E 和售货机 C 的毛利润相对较多，分别占总毛利润的 **33.1%**和 **21.4%**。售货机 D 的毛利润占比最少，为 11.6%。

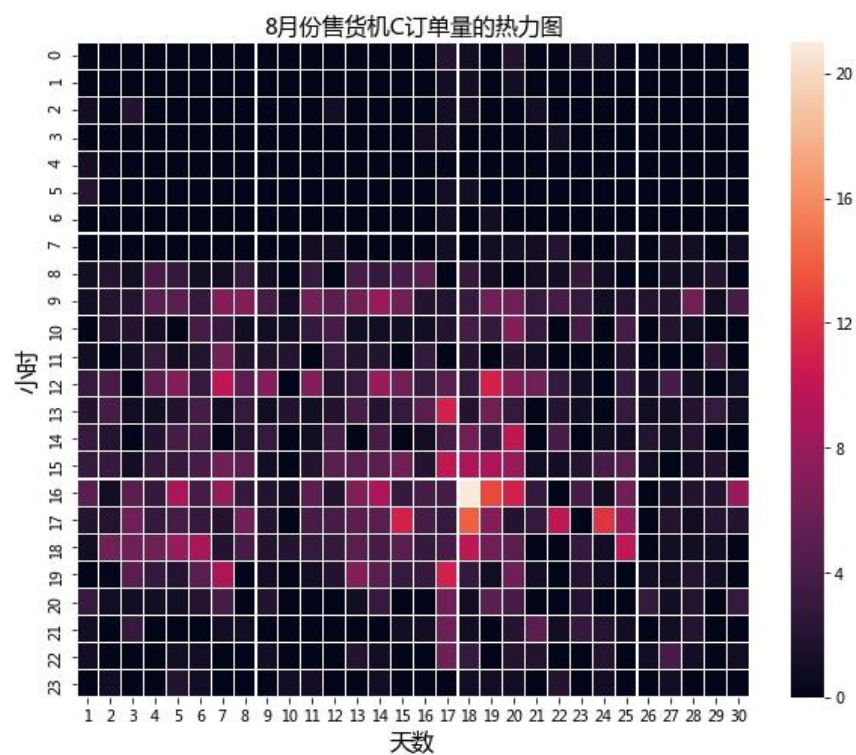
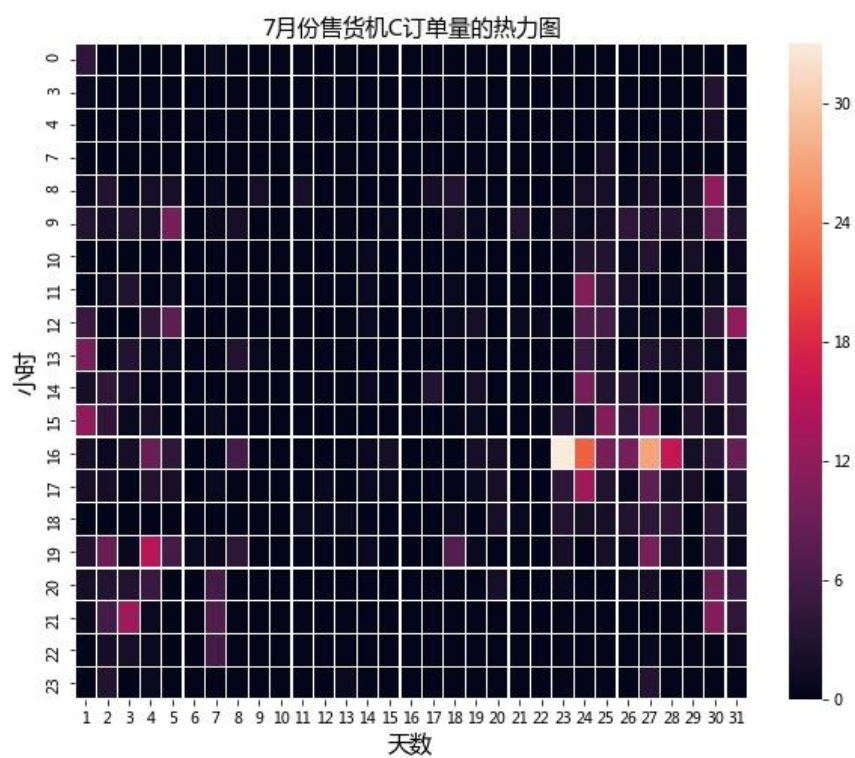
4. 每月交易额均值气泡图



由上图可知，**乳制品**和**功能饮料**每月的交易额都位列所有商品二级类的前二位，证明属于这两类的商品最热销。而坚果炒货、果冻/龟苓膏、糖果/巧克力类食品一年内的交易额都较少，每个月都处于交易额的倒数位置。

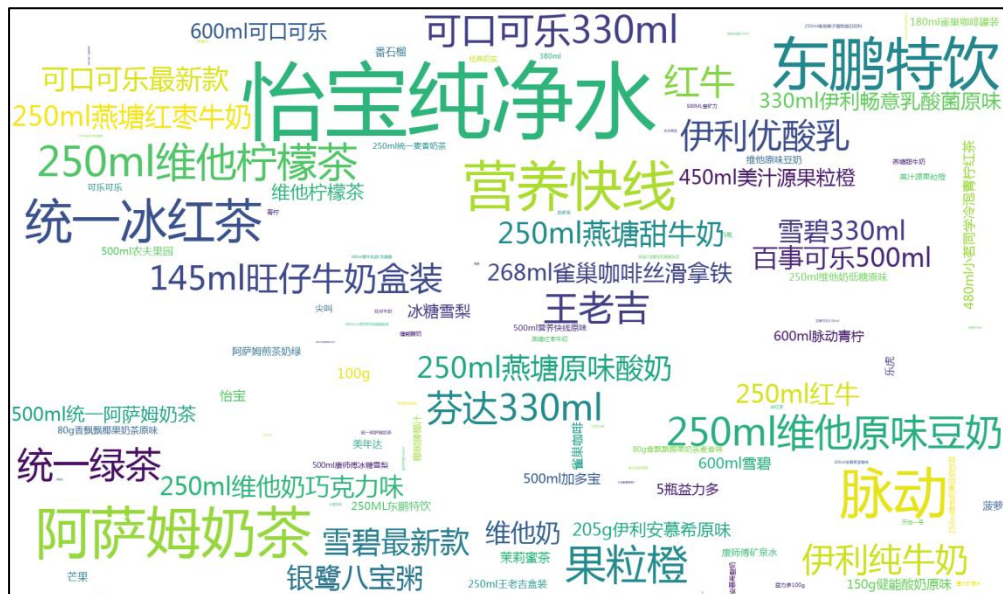
5. 售货机 C6、7、8 月订单量的热力图



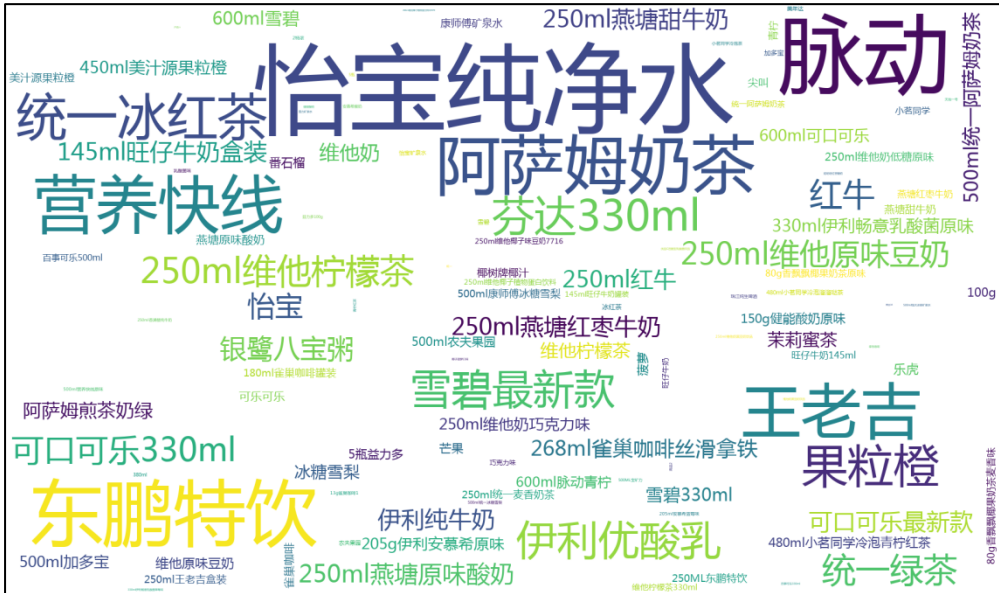


	月销量最大值出现时间	日销量最大值所在区间
6 月	6 月 23 日	15:00 --- 19:00
7 月	7 月 23 日	15:00 --- 17:00
8 月	8 月 18 日	15:00 --- 18:00

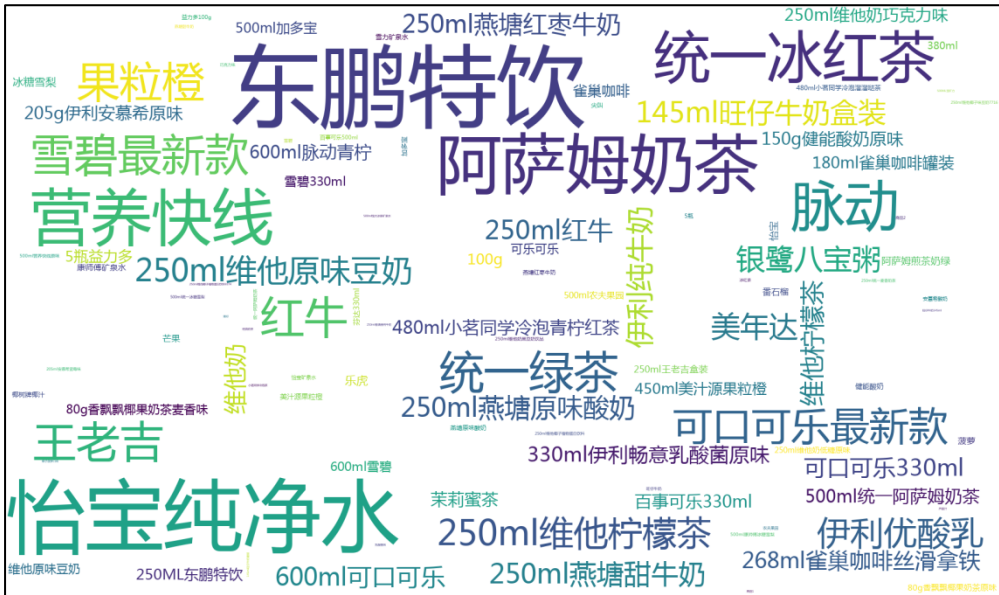
- (1) 在这三个月内，每天订单量的最大值都出现在下午 3 点到 7 点这个时间段，说明这个时间段是消费的高峰时段。
- (2) 每个月销售量的最大值都出现在每个月的 20 号附近，而且都在下午 4 点到 5 点这个时间段。
- (3) 相比于 6 月份，7 月份售货机 C 订单量热力图中黑色的部分有着十分明显的增加，占据了整个图约 85% 的面积，说明 7 月份售货机 C 订单量有了十分显著的下降。这一结论也可从售货机每月总交易额折线图中看出，不仅仅是 C 售货机，所有的售货机从 6 月到 7 月销售量都有十分明显的下降。



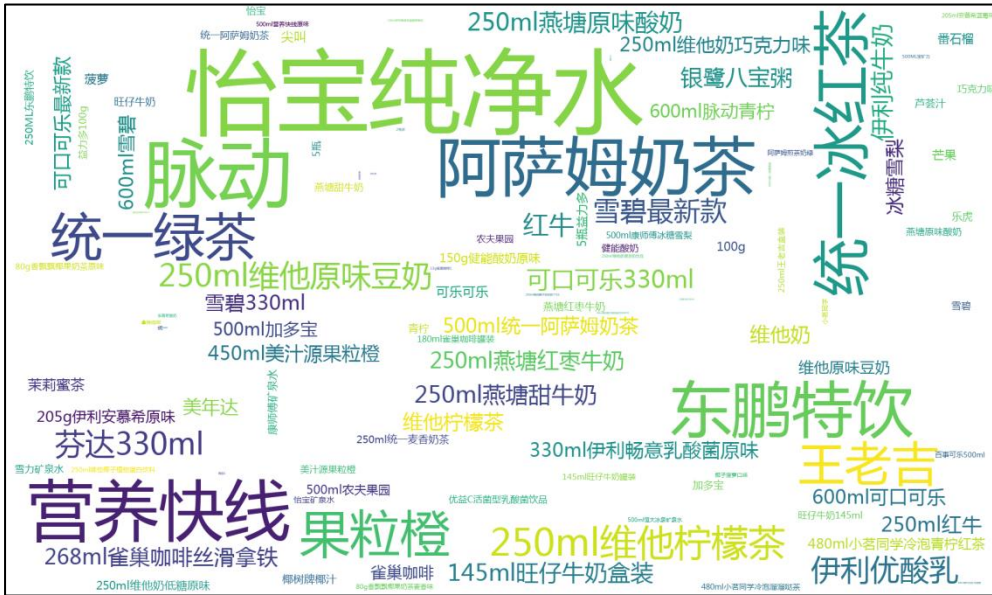
售货机 C



售货机 D



售货机 E



分析以上的 5 张售货机饮料类商品的词云图，给出以下营销意见：

- (1) 维持五种饮料的高库存：**结合商品标签及词云图分析可知，在 5 台售货机中，怡宝纯净水、营养快线、阿萨姆奶茶、脉动、东鹏特饮均为最畅销的 5 种饮料，所以应维持这 5 种饮料的高库存。
- (2) 适当减少碳酸饮料的库存：**一直以来比较受大众欢迎的碳酸饮料（可乐、雪碧等），在 5 台售货机中销量都较少，所以碳酸饮料的库存应适当地减少。
- (3) 增加运动饮料的选择：**在销量最高的 5 类饮品中，脉动、东鹏特饮属于运动饮料，营养快线、阿萨姆奶茶属于乳制品。乳制品的选择很多，相比之下，运动饮料的选择较少，只有脉动、东鹏特饮和红牛。在运动饮料处于热销的情况下，我认为可增加运动饮料的种类，给予消费者更多的选择。

五、 业务预测

1. 预测未来销售额的原理

预测销售额，即寻找商品销售额与商品各个特征之间的因果关系，并将这种关系用数学模型表示出来，并使用该数学模型计算它们之间的相关程度，从而预测未来的销售额。线性回归模型常被运用于解决此类问题。本文曾尝试寻找促销量与售货机交易额之间的关系，但分析数据后发现所有的实际金额都等同于应付金额，即无进行促销活动，所以不能建立其与销售量之间的关系。因此，本文采用时间序列模型中的 AR 模型，MA 模型，以及 ARMA 模型来对无人售货机 2018 年 1 月份的月交易额进行预测。

2. 时间序列模型的介绍

(1) 自相关系数(ACF)

在时间序列分析中，任意时刻的序列值 X_t 都是一个随机变量，每一个随机变量都会有均值和方差，记 X_t 的均值为 μ_t ，方差为 σ_t ；任取 $t, s \in T$ ，定义序列 $\{X_t\}$ 的自协方差函数 $\gamma(t, s) = E[(X_t - \mu_t)(X_s - \mu_s)]$ ，而自相关系数 ρ 的计算方法为 $\rho(t, s) = \frac{\text{cov}(X_t, X_s)}{\sigma_t \sigma_s}$ 。

相关系数度量指的是两个不同事件彼此之间的相互影响程度；而自相关系数度量的是同一事件在两个不同时期之间的相关程度，形象的讲就是度量自己过去的行为对自己现在的影响。

如果时间序列 $\{X_t, t \in T\}$ 在某一常数附近波动且波动范围有限，即有常数均值和常数方差，并且延迟 k 期的序列变量的自协方差和自相关系数是相等的或者说延迟 k 期的序列变量之间的影响程度是一样的，则称 $\{X_t, t \in T\}$ 为平稳序列。

(2)ARMA 模型

具有如下结构的模型被称为自回归移动平均模型，简记为 ARMA(p,q).

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

即在 t 时刻的随机变量 X_t 的取值 x_t 是前 p 期 $x_{t-1}, x_{t-2}, \cdots, x_{t-p}$ 和前 q 期 $\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_{t-q}$ 的多元线性函数，误差项是当期的随机干扰 ε_t ，为零均值白噪声序列。认为 x_t 主要是受过去 p 期的序列值和过去 q 期的误差项的共同影响。

特别的，当 $q=0$ 时， $AR(p)$ ；当 $p=0$ 时，是 $MA(q)$ 模型。

表 4 – 平稳 ARMA 模型的性质

统 计 量	性 质	统 计 量	性 质
均值	常数均值	自相关系数(ACF)	拖尾
方差	常数方差	偏自相关系数(PACF)	拖尾

3. 预测的具体步骤

(1)时间序列的预处理：平稳性检验以及纯随机性检验

(2)计算时间序列的相关系数：自相关系数(ACF)、偏自相关系数(PACF)，并结合时间序列的相关系数和时间序列模型的性质，给时间序列模型定阶，即选择合适的时间序列模型进行预测。

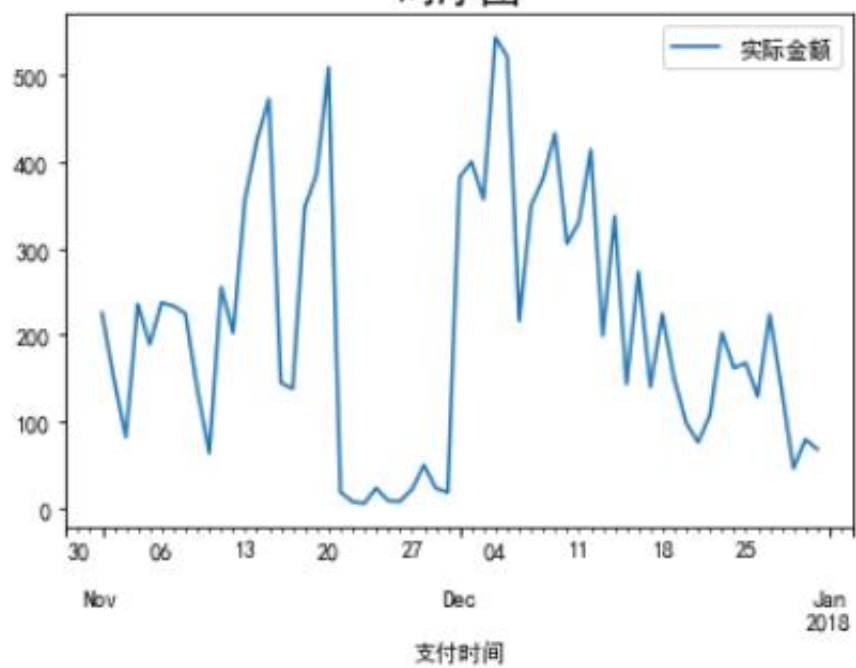
(3)模型可行性检验

(4)模型应用：进行短期预测。

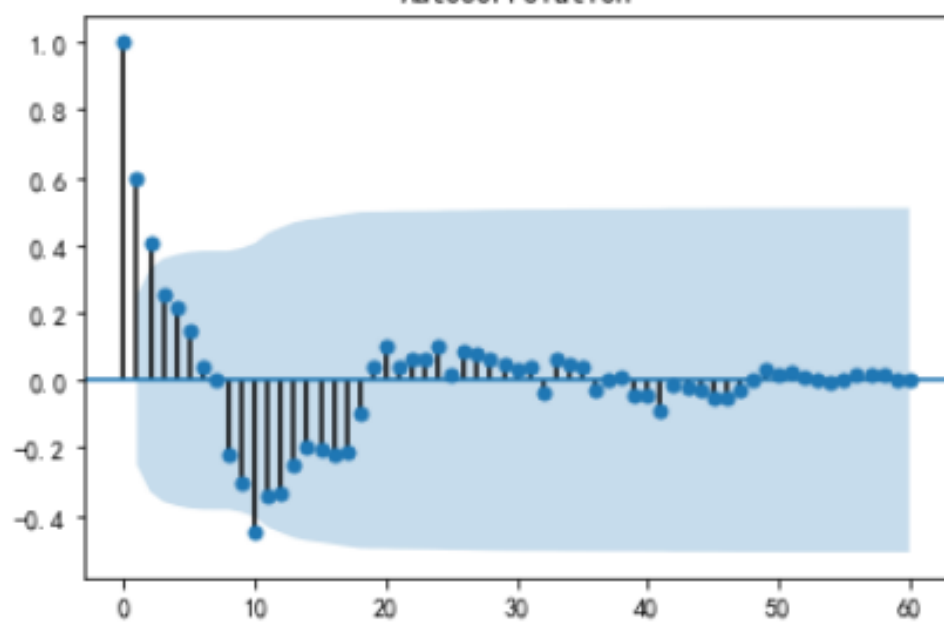
4. 预测过程

(1) 以 A 售货机为例，首先计算 A 售货机 11、12 月的日交易额，并以其为预测的历史数据。然后对该数据进行预处理操作，绘制时序图，并采用单位根检验法来判断其是否平稳非白噪声时间序列，结果如下图：

时序图



Autocorrelation



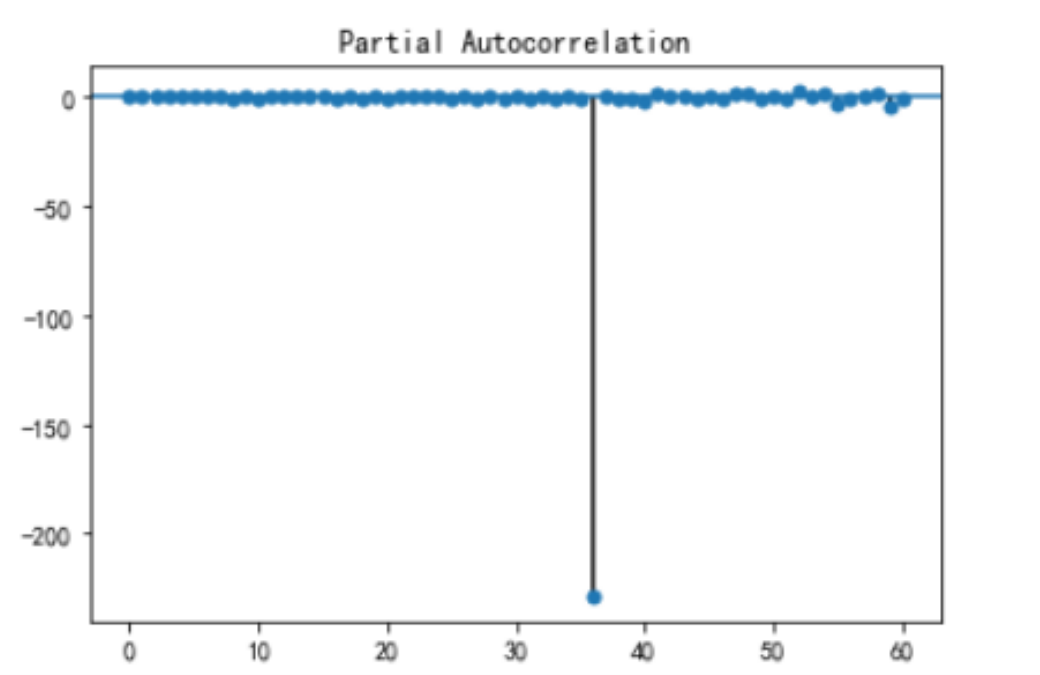


表 5 -原始序列的单位根检验

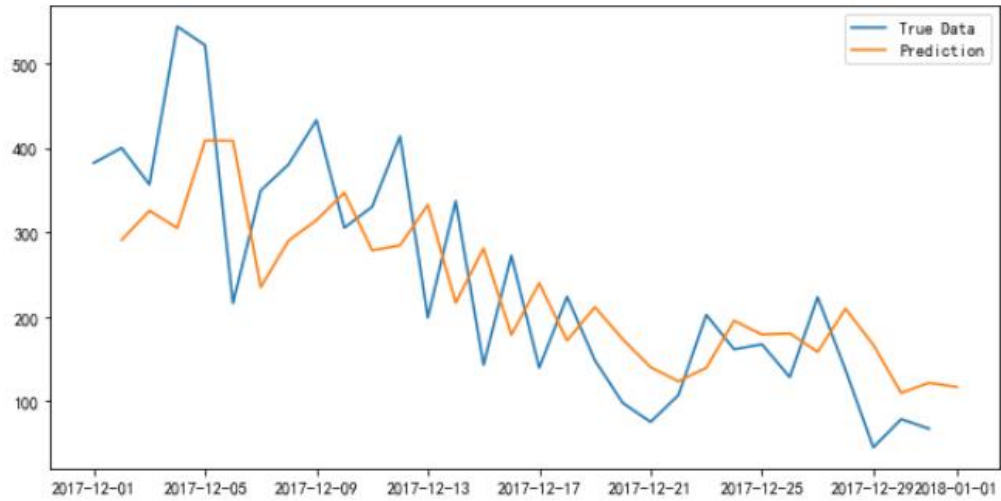
ADF	cValue			p 值
	1%	5%	10%	
-3.70035	-3.55436	-2.91107	-2.59319	0.0041

表 6 – 序列的白噪声检验

stat	P 值
11.304	0.00000156

如上述图表所示，时序图显示该序列没有明显的单调递增或递减的趋势，可大致判断为平稳序列；表 4 中单位根检验统计量对应的 p 值为 0.0041，远小于显著性水平 0.05，即可判断该时间序列为平稳序列。再对该序列进行纯随机性检验。Stat 值为 11.304，p 值远小于显著性水平 0.05，因此得出结论：该序列为平稳非白噪声序列，则可进行下一步步骤。

经计算可知，该时间序列是一个平稳非白噪声序列。接下来需要确定 ARMA 模型中 p 和 q 的值。时间序列的自相关图和偏自相关系数图均显示出拖尾，所以可以对原始序列建立 ARMA(1,1) 模型，将数据呈现为图像来检验模型的可行性，并对接下来一个月的交易额进行预测，计算标准误差。步骤及结果如下图所示。



A 售货机 12 月真实交易额和模型预测交易额的图像

由上图可知，ARMA(1,1)模型预测的交易额与实际的交易额大致相同。基于上述结果，可运用该模型对 2018 年 1 月份各销售机的交易额进行预测。

预测结果如下表。

售货机	预测 2018 年 1 月月交易额（元）
A	5836
B	8218
C	8661
D	7296
E	15680

表 7 各售货机 2018 年 1 月交易额预测值

六、 小结

本文使用 Python 软件，对无人售货机的数据进行分析，得出各售货机的销量及销售额的情况并将其可视化，再绘制出各售货机商品销量的词云图，结合分析，给出营销意见。除此之外，本文通过利用 ARMA 时间序列模型，预测 2018 年 1 月份各售货机的销售额。