目录

[一、背景介绍 1](#_Toc19068)

[1.商务背景 1](#_Toc4257)

[2.分析目标 1](#_Toc29115)

[二、数据说明 2](#_Toc2753)

[1.数据来源 2](#_Toc17886)

[2.数据情况说明 2](#_Toc4198)

[三、营销数据分析 2](#_Toc20967)

[1.数据预处理与分析 2](#_Toc11276)

[2.数据分析与可视化 7](#_Toc31697)

[四、商品标签设置与画像分析 12](#_Toc849)

[1.标签设置 12](#_Toc5141)

[2.画像分析 13](#_Toc2290)

[3.小结 18](#_Toc7696)

[五、建模分析 20](#_Toc2051)

[1.预测未来销售额的原理 20](#_Toc245)

[2.模型建立与分析 21](#_Toc6578)

# 一、背景介绍

## 1.商务背景

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，按要求绘制对应图表，分析每台售货机2018年1月商品销售额的预测问题。

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、数据说明

## 1.数据来源

数据来源：附件1、附件2

起止时间：2017年1月--2017年12月

## 2.数据情况说明

附件1：提供了A、B、C、D、E五台自动售货机在2017年1月-2017年12月期间每一订单的订单号、设备ID、应付金额、实际金额、商品、支付时间等数据；

附件2：对商品所属的大类、二级类进行划分。

# 三、营销数据分析

## 1.数据预处理与分析

对附件1中的数据去除脏数据，剔除不合理数据，除去了时间异常和实际金额为0的数据；再提取每台售货机对应的销售数据，对所需要的销量，交易额，利润等进行计算。

#### 5月份销售数据分析

表1 2017年5月各售货机销售数据情况

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E | 汇总 |
| 订单量 | 750 | 860 | 782 | 553 | 1287 | 4232 |
| 交易额 | 3385.1 | 3681.2 | 3729.4 | 2392.1 | 5699.0 | 18886.8 |

图1-1 2017年5月各售货机订单量直方图

图1-2 2017年5月各售货机销售额直方图

2017年5月，5台自动售货机的总订单量为4232，总交易额为18886.8。其中，E自动售货机的订单量和销售额均最多，分别为1287，5699，均达到总数的1/3以上；D自动售货机的订单量和销售额均最少，分别为553，2392.1。订单量的多少直接影响销售额，因此E自动售货机的高订单量使其也拥有高销售额。

此外，自动售货机A，B，C的订单量和销售额相差不大，其中，自动售货机B的订单量位居第二，而自动售货机C的销售额位居第二。由于自动售货机每单的金额不一，所以也可能存在订单量高，销售额却略低的情况。

#### 各月销售数据分析

1. 月日均订单量分析

表2 2017年各售货机每月日均订单量

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 | 7月 | 8月 | 9月 | 10月 | 11月 | 12月 | 均值 |
| A | 11 | 4 | 9 | 15 | 25 | 55 | 16 | 22 | 35 | 52 | 39 | 67 | 29 |
| B | 12 | 6 | 9 | 20 | 29 | 62 | 11 | 33 | 58 | 67 | 68 | 74 | 37 |
| C | 13 | 7 | 9 | 24 | 26 | 63 | 25 | 42 | 56 | 74 | 65 | 79 | 40 |
| D | 9 | 5 | 6 | 15 | 18 | 34 | 11 | 24 | 33 | 39 | 40 | 55 | 24 |
| E | 12 | 9 | 12 | 30 | 43 | 86 | 27 | 59 | 137 | 92 | 167 | 108 | 65 |

图2 2017年各售货机每月的日均订单量折线图

注：1-12月每月按30天计算

2017年，各自动售货机在1月-12月的日均订单量总体趋势一致，都是在1-6月日均订单量逐渐增加，7月日均订单量减少，之后日均订单量再次增加。其中，1-3月，各售货机的订单量相差不大；8月-12月，订单量相差较为明显。

除了1月，E自动售货机在其他各月的的日均订单量均最高，且在9，11，12三个月均超过100；而D自动售货机在各月的日均订单量均最低。

1. 月每单均销售额分析

表3 2017年各售货机每月的每单平均交易额

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 | 7月 | 8月 | 9月 | 10月 | 11月 | 12月 |
| A | 4.52 | 3.86 | 3.59 | 4.06 | 4.51 | 4.07 | 4.11 | 3.36 | 4.31 | 4.03 | 4.48 | 3.80 |
| B | 3.75 | 3.26 | 3.61 | 4.09 | 4.28 | 4.08 | 4.41 | 3.58 | 4.14 | 4.12 | 4.28 | 3.67 |
| C | 4.36 | 3.83 | 3.77 | 4.42 | 4.77 | 4.52 | 4.00 | 3.91 | 4.44 | 4.29 | 4.36 | 3.95 |
| D | 3.74 | 3.09 | 4.31 | 3.82 | 4.33 | 4.06 | 4.26 | 3.32 | 3.91 | 3.90 | 3.86 | 3.58 |
| E | 4.68 | 3.64 | 4.31 | 4.16 | 4.43 | 3.84 | 3.93 | 3.80 | 4.14 | 3.68 | 4.29 | 4.17 |

图3 2017年各售货机每月每单的交易额均值折线图

2017年，各售货机每月每单的交易额均值较为集中，大致在3-5区间内；且整体上看并无明显的差距。

## 2.数据分析与可视化

#### 6月份销量分析

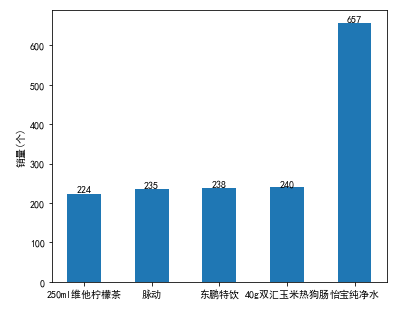


图4 2017年6月销量前5的商品销量直方图

2017年6月销量前5的商品为：怡宝纯净水、40g双汇玉米热狗肠、东鹏特饮、脉动、250ml维他柠檬茶。其中，怡宝纯净水销量最高，为657；其余四种商品销量整体差别不大。此外，5种商品中饮料类占据4种，唯一的非饮料类40g双汇玉米热狗肠销量位居第二。可见，饮料类商品比非饮料类商品更为热销。

#### 月总交易额分析

1.月总交易额折线图

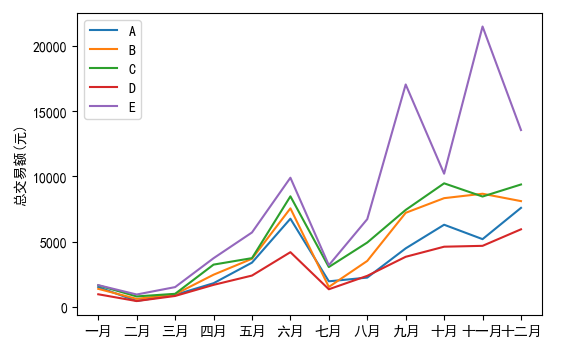


图5 各售货机每月总交易额折线图

2017年，各自动售货机在1月-12月的月总交易额总体趋势一致，1-6月月总交易额逐渐增加，7月月总交易额减少，之后又再次增加。其中，1-3月，各售货机的月总交易额较为接近；8月-12月的月总交易额相差较为明显。

E自动售货机在各月的的月总交易额处于较高状态，且在9月份月总交易额超过15000，11月份甚至超过了20000；而D自动售货机在各月的月总交易额则最低。

与图2联系，自动售货机每月月总交易额和日均订单量趋势一致，日均订单量高的自动售货机，其月总交易额也较高。

2.月交易额环比增长率分析

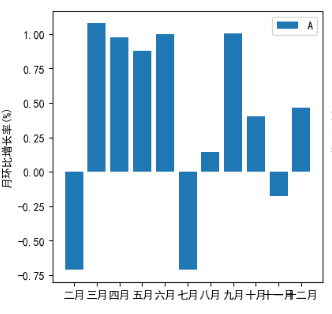


图6-1 A自动售货机月交易额环比增长率柱形图

A自动售货机在2月，7月，11月的月总交易额环比增长率为负，且2月和7月下降较大，接近0.75%；在3-6月，9月总交易额环比增长率比较高，其中，3月总交易额环比增长率最高，超过了1%。

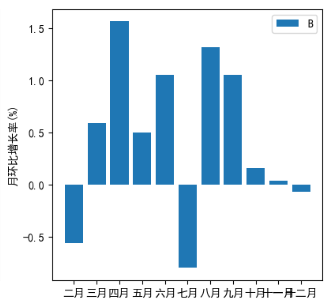


图6-2 B自动售货机月交易额环比增长率柱形图

B自动售货机在2月，7月，12月的月总交易额环比增长率为负，且7月下降最大，接近0.75%；在4、8、9月总交易额环比增长率比较高，其中，4月总交易额环比增长率最高，超过了1.5%。

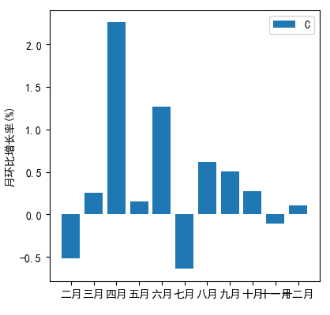


图6-3 C自动售货机月交易额环比增长率柱形图

C自动售货机在2月，7月，11月的月总交易额环比增长率为负，且7月下降最大，超过0.5%；在4、6月总交易额环比增长率比较高，其中，4月总交易额环比增长率最高，超过了2%。

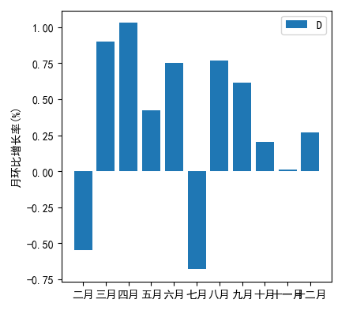


图6-4 D自动售货机月交易额环比增长率柱形图

D自动售货机在2月，7月的月总交易额环比增长率为负，且7月下降较大，超过0.5%；在3、4、6、8月总交易额环比增长率比较高，其中，4月总交易额环比增长率最高，超过了1%。

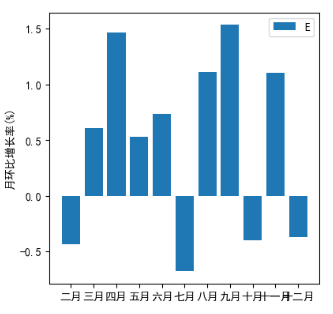


图6-5 E自动售货机月交易额环比增长率柱形图

E自动售货机在2月，7月，10月、12月的月总交易额环比增长率为负，且7月下降最大，超过0.5%；在4、8、9、11月总交易额环比增长率比较高，其中，9月总交易额环比增长率最高，超过了1.5%。

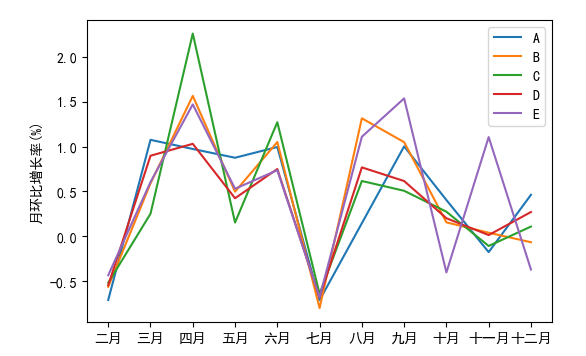


图6-6 月交易额环比增长率折线图

整体上看，在2，7，11月，各自动售货机的月交易额环比增长率较低，为负值。值得注意的是，E自动售货机在10月的月交易额环比增长率为负值，到了11月却达到了1.0%，这使得其在11月份的总交易额比较高；而4月份C自动售货机的月交易额环比增长率超过了2%。

#### 利润分析

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 自动售货机 | A | B | C | D | E |
| 毛利润 | 9707 | 12543 | 14291 | 7703 | 22341 |

表4 各自动售货机毛利润情况表

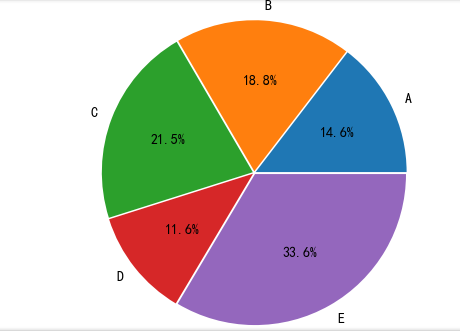


图7 各自动售货机毛利润占总毛利润比例饼图

2017年，E自动售货机所获毛利润最多，占比达到33.6%；D自动售货机所获毛利润最少，占比只有11.6%；C自动售货机毛利润占比位居第二，为21.5%；B、A自动售货机依次次之，为18.8%，14.6%。总体上看，自动售货机所获毛利润与全年的订单量，总交易额成正比，全年的订单量越多，总交易额越大，所获毛利润就越多，这与上面对售货机每月的日均订单量，月总交易额的分析也可以得出。

#### 二级类商品月均交易额分析

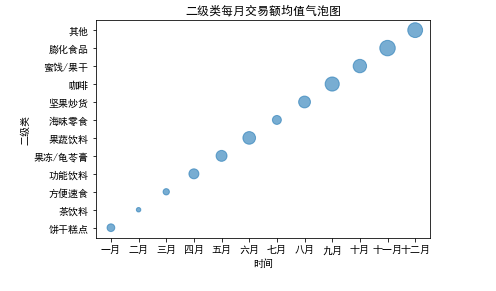


图8 2017年各月交易额均值气泡图

各二级类商品在2017年2月份的交易额均值最少，12月份最大，且在9-12月交易额均值接近。

此外，文末附表给出了二级类商品在2017年1-12月各月的销售额。由数据表可知：饮料类商品如茶饮料，功能饮料、乳制品、水、碳酸饮料在各月的销售额较高，销量较好，而非饮料类商品则是肉干/豆制品/蛋、饼干糕点比较受欢迎；且上述商品大多是在4月或5月之后销售额达到较高水平。最高的是茶饮料，在12月份销售额高达7803。整体来看，自动售货机的主要销售内容是饮料类商品。

#### C自动售货机6-8月订单量分析

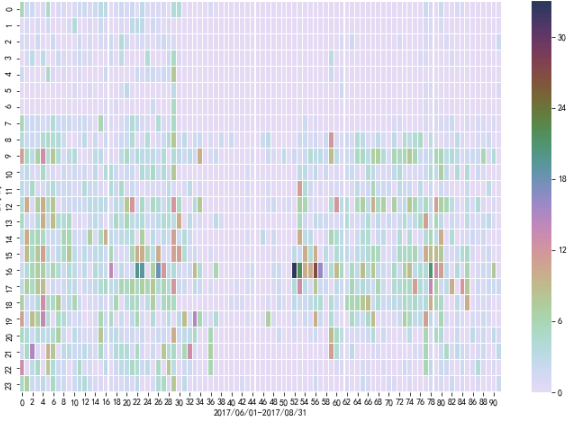


图9 C售货机6、7、8三个月订单量的热力图

C售货机主要售出时间集中在6月1号-7月2号的7:00-23:00、7月23号-8月31号的8:00-21:00这两个时间段。在这两个时间段中，15：00和16：00销量较多，7月23号-7月27号销量大致在20以上，且在7月23号的16：00销量最大，超过30。

C售货机在7月3号-7月22号的日销量几乎都小于6，因此，虽然7月23号是单日销量的最大值，但整体上看，7月的订单总量比6月、8月少；相应地，7月的日均订单量比6月、8月少；7月的总销售额也比6月、8月少，这和上面的分析是一致的。

# 四、商品标签设置与画像分析

## 1.标签设置

将5台售货机的商品分为非饮料类和饮料类，对每一类商品，根据销量设置标签1（热销/正常/滞销），根据销售总额设置标签2（大额/适中/小额），结果保存在表格中。由于各售货机整体销售情况不在同一水平，且非饮料类商品和饮料类商品销售情况差距较大，因此对各售货机非饮料类商品和饮料类商品标签1和标签2的设置是根据售货机各自的实际情况设定的。具体如下：

记销量为x，销售额为y：

* A自动售货机

（1）饮料类商品

标签1：热销：x>=200；正常：60<=x<200；滞销：x<60

标签2 ：大额：y>=1000；适中：400<=y<1000；小额：y<400

（2）非饮料类商品

标签1：热销：x>=148；正常：50<=x<148；滞销：x<50

标签2 ：大额：y>=450；适中：200<=y<450；小额：y<200

* B自动售货机

（1）饮料类商品

标签1：热销：x>=300；正常：100<=x<300；滞销：x<100

标签2 ：大额：y>=1000；适中：400<=y<1000；小额：y<400

（2）非饮料类商品

标签1：热销：x>=145；正常：50<=x<145；滞销：x<50

标签2 ：大额：y>=600；适中：300<=y<600；小额：y<300

* C自动售货机

（1）饮料类商品

标签1：热销：x>=300；正常：100<=x<300；滞销：x<100

标签2 ：大额：y>=1000；适中：300<=y<1000；小额：y<300

（2）非饮料类商品

标签1：热销：x>=120；正常：40<=x<120；滞销：x<40

标签2 ：大额：y>=600；适中：200<=y<600；小额：y<200

* D自动售货机

（1）饮料类商品

标签1：热销：x>=180；正常：80<=x<180；滞销：x<80

标签2 ：大额：y>=700；适中：250<=y<700；小额：y<250

（2）非饮料类商品

标签1：热销：x>=100；正常：30<=x<100；滞销：x<30

标签2 ：大额：y>=350；适中：150<=y<350；小额：y<150

* E自动售货机

（1）饮料类商品

标签1：热销：x>=450；正常：150<=x<450；滞销：x<150

标签2 ：大额：y>=1500；适中：400<=y<1500；小额：y<400

（2）非饮料类商品

标签1：热销：x>=200；正常：80<=x<200；滞销：x<80

标签2 ：大额：y>=800；适中：300<=y<800；小额：y<300

## 2.画像分析

根据标签生成自动售货机图像，并结合生成的表格分析如下：

* A自动售货机：



图10 A自动售货机画像

A自动售货机饮料类商品中怡宝纯净水、东鹏特饮、阿萨姆奶茶、脉动，雪碧、营养快线、维他柠檬茶等销量较高，销量最高的是怡宝纯净水，达到531；非饮料类商品中无穷农场盐鸡蛋、无穷烤鸡小腿、卫龙亲嘴烧、咪咪虾条、双汇玉米热狗肠销量较高，销量最高的是无穷农场盐鸡蛋，为264。

* B自动售货机：

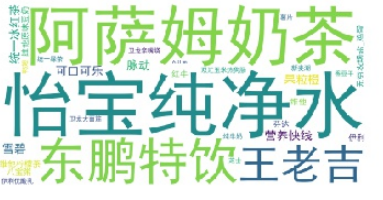


图11 B自动售货机画像

B自动售货机饮料类商品中怡宝纯净水、东鹏特饮、阿萨姆奶茶、王老吉、可口可乐等销量较高，其中怡宝纯净水销量最高，为1342，是销量第二的东鹏特饮的2倍多；非饮料类商品中无穷农场盐鸡蛋、卫龙大面筋、咪咪虾条、双汇玉米热狗肠、鸭翅销量较高，销量最高的是无穷农场盐鸡蛋，为242。

* C自动售货机：



图12 C自动售货机画像

C自动售货机饮料类商品中怡宝纯净水、脉动、东鹏特饮、阿萨姆奶茶、营养快线、王老吉、果粒橙、雪碧等销量较高，其中怡宝纯净水销量最高，接近1000；非饮料类商品中无穷农场盐鸡蛋、卫龙大面筋、咪咪虾条、双汇玉米热狗肠、鸭翅销量较高，销量最高的是无穷农场盐鸡蛋，为274 。

* D自动售货机：

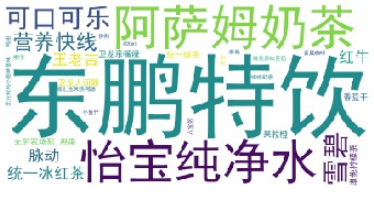


图13 D自动售货机画像

D自动售货机饮料类商品中热销的有东鹏特饮、怡宝纯净水、阿萨姆奶茶、营养快线、统一冰红茶、脉动、雪碧、可口可乐等，其中东鹏特饮销量最高，为411；非饮料类商品中卫龙大面筋、无穷农场盐鸡蛋、香豆干、咪咪虾条、双汇玉米热狗肠较为热销，销量最高的是卫龙大面筋，为205。

* E自动售货机：

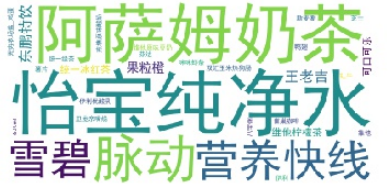


图14 E自动售货机画像

E自动售货机饮料类商品中热销的有怡宝纯净水、脉动、营养快线、阿萨姆奶茶、东鹏特饮、雪碧等，其中怡宝纯净水销量最高，达到1705；非饮料类商品中双汇玉米热狗肠、无穷农场盐鸡蛋、鸭翅、咪咪虾条较为热销，销量最高的是双汇玉米热狗肠，为415。

由以上分析可知，5台自动售货机主要以销售饮料类商品为主，饮料类商品中热销商品主要集中在怡宝纯净水、东鹏特饮、阿萨姆奶茶、脉动、营养快线、雪碧等几类；值得注意的是，D自动售货机中东鹏特饮销量最高，其他的均是怡宝纯净水销量最高；此外，A自动售货机的维他柠檬茶、B自动售货机的可口可乐、王老吉，C自动售货机的王老吉、果粒橙，D自动售货机的统一冰红茶、可口可乐，E自动售货机的果粒橙销量均不错。

非饮料类商品中，热销的集中为无穷农场盐鸡蛋、卫龙大面筋、咪咪虾条、双汇玉米热狗肠、鸭翅等几类。其中，A、B、C自动售货机销量最高的为无穷农场盐鸡蛋，D自动售货机销量最高为卫龙大面筋，E自动售货机销量最高的为双汇玉米热狗肠。此外A自动售货机的无穷烤鸡小腿、卫龙亲嘴烧，D自动售货机的香豆干销量都较为可观。

综上，对各自动售货机投放的商品主要为饮料类商品中的怡宝纯净水、东鹏特饮、阿萨姆奶茶、脉动、营养快线、雪碧，对于各自的最高销量商品怡宝纯净水和东鹏特饮可加大投放；A自动售货机增加投放维他柠檬茶，B自动售货机增加投放可口可乐、王老吉，C自动售货机增加投放王老吉、果粒橙，D自动售货机增加投放统一冰红茶、可口可乐，E自动售货机增加投放果粒橙。

此外，对各自动售货机可投放适当比例的非饮料类商品，主要为：无穷农场盐鸡蛋、卫龙大面筋、咪咪虾条、双汇玉米热狗肠、鸭翅等，对售货机各自的高销量商品无穷农场盐鸡蛋、卫龙大面筋、双汇玉米热狗肠适当加大投放；并且注意增加A自动售货机无穷烤鸡小腿、卫龙亲嘴烧，D自动售货机的香豆干的投放。

由对各类商品的销售额分析所得数据表可知：A、B、C、D自动售货机中红牛、口味王槟榔，C自动售货机中维他柠檬茶，E自动售货机中红牛、维他柠檬茶、王老吉、马来西亚进口茱蒂丝口口香饼干等虽销量正常，但销售额较大，因此也可按适当比例投放。

## 3.小结

表5 自动售货机商品投放建议表

|  |  |
| --- | --- |
| 自动售货机 | 投放建议 |
| A | * 水：怡宝纯净水 * 功能饮料：脉动、东鹏特饮、红牛 * 茶饮料：阿萨姆奶茶、维他柠檬茶 * 碳酸饮料：雪碧 * 乳制品：营养快线 * 肉干/豆制品/蛋：无穷农场盐鸡蛋、卫龙亲嘴烧、双汇玉米热狗肠、无穷烤鸡小腿 * 膨化食品：咪咪虾条 * 蜜饯/果干：口味王槟榔 |
| B | * 水：怡宝纯净水 * 功能饮料：脉动、东鹏特饮、红牛 * 茶饮料：阿萨姆奶茶、王老吉 * 碳酸饮料：可口可乐 * 乳制品：营养快线 * 肉干/豆制品/蛋：无穷农场盐鸡蛋、卫龙大面筋、双汇玉米热狗肠、鸭翅 * 膨化食品：咪咪虾条 * 蜜饯/果干：口味王槟榔 |
| C | * 水：怡宝纯净水 * 功能饮料：脉动、东鹏特饮、红牛 * 茶饮料：阿萨姆奶茶、王老吉、维他柠檬茶 * 碳酸饮料：雪碧 * 乳制品：营养快线 * 果蔬饮料：果粒橙 * 肉干/豆制品/蛋：无穷农场盐鸡蛋、卫龙大面筋、双汇玉米热狗肠、鸭翅 * 膨化食品：咪咪虾条 * 蜜饯/果干：口味王槟榔 |
| D | * 水：怡宝纯净水 * 功能饮料：脉动、东鹏特饮、红牛 * 茶饮料：阿萨姆奶茶、统一冰红茶 * 碳酸饮料：雪碧、可口可乐 * 乳制品：营养快线 * 肉干/豆制品/蛋：无穷农场盐鸡蛋、双汇玉米热狗肠、卫龙大面筋、香豆干 * 膨化食品：咪咪虾条 * 蜜饯/果干：口味王槟榔 |
| E | * 水：怡宝纯净水 * 功能饮料：脉动、东鹏特饮、红牛 * 茶饮料：阿萨姆奶茶、维他柠檬茶、王老吉 * 碳酸饮料：雪碧 * 果蔬饮料：果粒橙 * 肉干/豆制品/蛋：无穷农场盐鸡蛋、双汇玉米热狗肠、鸭翅 * 膨化食品：咪咪虾条 * 饼干糕点：马来西亚进口茱蒂丝口口香饼干 |

# 五、建模分析

## 1.预测未来销售额的原理

1. 根据过去的已知的销售额数据，对未来的情况进行建模并预测。可以用趋势外推预测方法，即进行时间序列分析，根据事物的历史和现实数据，寻求事物随时间推移而发生变化的规律，从而推测其未来状况的一种常用预测方法。
2. 根据已知的销售额数据和影响销售额变化的各种自变量数据进行分析，建立反映销售额和自变量之间关系的模型，再根据自变量数据预测因变量销售额的数据。故可以选用回归模型，进行函数拟合，选择一条函数曲线使其能较好地拟合数据并预测未指数据。

由项目给出的2017年5台自动售货机的数据，可得出每台售货机每月的销售额和订单量。

若用趋势外推预测方法进行预测，则需要数据较多，至少需要各售货机5年以上各月的销售额。

若通过建模根据自变量预测销售额，由给出的数据，选择订单量为自变量，销售额为因变量，建立合适的模型，但要预测2018年1月的销售额，仍需要售货机2018年1月的订单量。

所以，由给出的项目数据无法完成对每台售货机的每个大类商品在2018年1月的交易额的预测。

要完成对每台售货机的每个大类商品在2018年1月的交易额的预测，需要的数据：各售货机5年以上的月销售额数据或各售货机在2018年1月的订单量数据。

## 2.模型建立与分析

本文以A自动售货机为例，根据已有数据对非饮料类和饮料类商品在2017年各月的销售额和订单量进行建模，探讨此预测方法的合理性及有效性。

首先对任务一提取出的A自动售货机的数据进行分类，分别提取出其中非饮料类商品和饮料类商品的数据，计算各自在2017年1-12月的销售额和订单量，数据如下。

表6 非饮料类商品2017年各月销售额和订单量

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 月份 | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 | 7月 | 8月 | 9月 | 10月 | 11月 | 12月 |
| 销售额 | 792.4 | 127 | 356.7 | 864.3 | 2031.9 | 3006.3 | 843.7 | 650.4 | 1696.8 | 2947.4 | 2082.3 | 3162.5 |
| 订单量 | 177 | 19 | 98 | 205 | 441 | 752 | 198 | 191 | 366 | 705 | 416 | 779 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 月份 | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 | 7月 | 8月 | 9月 | 10月 | 11月 | 12月 |
| 销售额 | 717.3 | 313.5 | 557.6 | 940.2 | 1353.2 | 3748.8 | 1106.8 | 1586.5 | 2782.7 | 3345 | 3104.7 | 4424.6 |
| 订单量 | 160 | 95 | 157 | 239 | 309 | 909 | 277 | 474 | 673 | 856 | 743 | 1218 |

表7 饮料类商品2017年各月销售额和订单量

做出散点图如下：

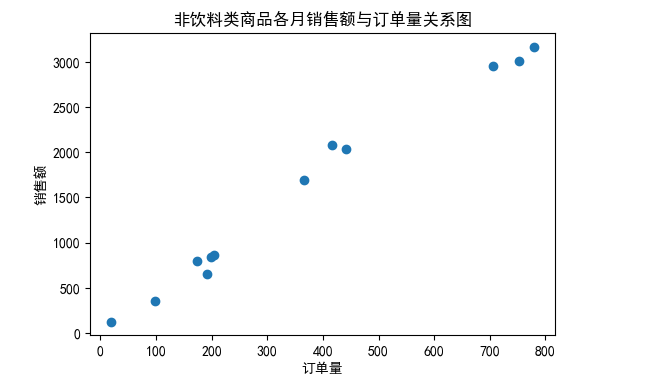


图15 非饮料类商品各月销售额与订单量关系图

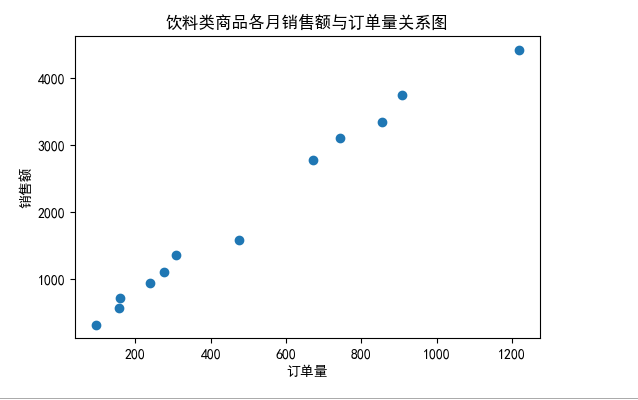


图16 饮料类商品各月销售额与订单量关系图

由图可知，非饮料类和饮料类商品的销售额和订单量大致呈线性关系，并且数据分布集中在某几处，故选用线性回归LinearRegression和最近邻回归KNeighborsRegressor进行建模。由于数据较少，故先进行10折交叉验证，再分别选用LinearRegression和KNeighborsRegressor方法建模，以均方误差MSE的均值作为评分评价模型的好坏。

* 非饮料类商品

对于非饮料类商品，由LinearRegression和KNeighborsRegressor方法建模得出均方误差均值分别为-29301.26，-330561.10，由均方误差是一种损失函数，优化的目标是使其最小化，所以越小越好，可知选择最近邻回归KNeighborsRegressor建模比较合适。

选定KNeighborsRegressor建模后，对模型进行优化，选择最优参数，仍以均方误差作为评分，得出参数alpha与均方误差的关系图如下：

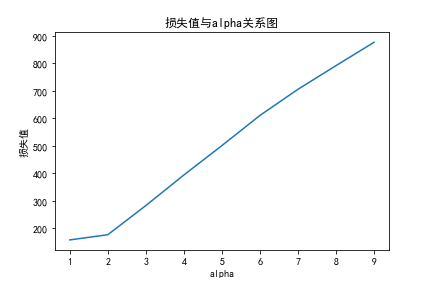


图17 损失值与alpha关系图

由图可知，alpha=1时，损失值最小，所以最优参数值为1。

* 饮料类商品

对于饮料类商品，由LinearRegression和KNeighborsRegressor方法建模得出均方误差均值分别为-54945.83，-395840.02，可知选择最近邻回归KNeighborsRegressor建模比较合适。对模型调参后，得出参数alpha与均方误差的关系图如下：

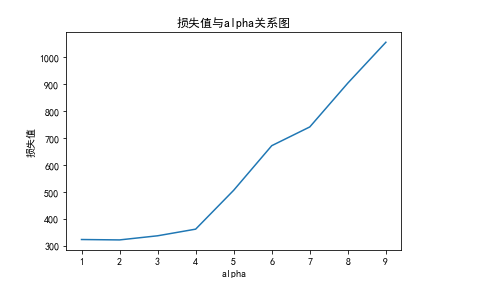


图18 损失值与alpha关系图

由图可知，alpha=1时，损失值最小，所以最优参数值为1。

* **结论**

通过以上建模过程，可知可选用最近邻回归KNeighborsRegressor对非饮料类和饮料类商品数据进行建模，并且此方法比较有效，当已知自变量时可用此算法预测销售额。

附表：二级类商品在2017年1-12月各月的销售额

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 一月 | 二月 | 三月 | 四月 | 五月 | 六月 | 七月 | 八月 | 九月 | 十月 | 十一月 | 十二月 |
| 饼干糕点 | 856 | 108 | 254 | 1288 | 2453 | 3846 | 995 | 888 | 3558 | 3174 | 4001 | 2928 |
| 茶饮料 | 898 | 684 | 1013 | 2040 | 1938 | 4905 | 1777 | 4303 | 6150 | 6512 | 7716 | 7803 |
| 方便素食 | 457 | 123 | 346 | 825 | 2022 | 3010 | 891 | 773 | 2482 | 2470 | 3547 | 2720 |
| 功能饮料 | 718 | 663 | 955 | 1558 | 1658 | 3654 | 1181 | 3672 | 4984 | 4403 | 6322 | 7025 |
| 果冻/龟苓膏 | 5 | 0 | 5 | 27 | 32 | 17 | 22 | 0 | 108 | 30 | 66 | 55 |
| 果蔬饮料 | 217 | 181 | 140 | 393 | 554 | 891 | 296 | 849 | 1583 | 1293 | 1953 | 1462 |
| 海味零食 | 61 | 15 | 42 | 119 | 222 | 372 | 78 | 108 | 278 | 370 | 475 | 380 |
| 坚果炒货 | 50 | 0 | 21 | 31 | 80 | 147 | 62 | 0 | 109 | 80 | 165 | 94 |
| 咖啡 | 170 | 23 | 34 | 297 | 611 | 515 | 230 | 154 | 1064 | 675 | 1253 | 898 |
| 蜜饯/果干 | 240 | 160 | 205 | 429 | 536 | 988 | 394 | 920 | 1223 | 1333 | 1654 | 2067 |
| 膨化食品 | 325 | 12 | 145 | 602 | 1591 | 2799 | 536 | 328 | 1594 | 1949 | 1851 | 1742 |
| 其他 | 0 | 0 | 0 | 0 | 0 | 3 | 26 | 32 | 150 | 62 | 216 | 22 |
| 肉干/豆制品/蛋 | 803 | 219 | 478 | 1417 | 2537 | 4379 | 1373 | 1397 | 3575 | 4569 | 3998 | 3887 |
| 乳制品 | 1148 | 343 | 595 | 1857 | 2652 | 5406 | 1574 | 2179 | 5893 | 5554 | 7279 | 5407 |
| 水 | 305 | 208 | 217 | 570 | 386 | 1795 | 396 | 1571 | 2229 | 2068 | 2107 | 2060 |
| 碳酸饮料 | 384 | 329 | 426 | 802 | 793 | 2088 | 648 | 1966 | 3079 | 2496 | 3606 | 3555 |
| 糖果/巧克力 | 53 | 4 | 24 | 78 | 132 | 241 | 109 | 40 | 172 | 266 | 267 | 118 |
| 香烟 | 165 | 73 | 90 | 75 | 23 | 228 | 50 | 281 | 485 | 236 | 665 | 910 |
| 植物蛋白 | 263 | 64 | 197 | 439 | 609 | 1497 | 381 | 271 | 1197 | 1288 | 1216 | 1349 |
| 纸巾 | 19 | 0 | 12 | 49 | 58 | 85 | 24 | 40 | 90 | 79 | 132 | 87 |