目 录

[一、项目简介 2](#_Toc23016799)

[1.1项目背景 2](#_Toc23016800)

[1.2分析目标 2](#_Toc23016801)

[二、项目数据来源 2](#_Toc23016802)

[2.1数据的获取 2](#_Toc23016803)

[2.2数据的时间段 2](#_Toc23016804)

[2.3数据的大小 2](#_Toc23016805)

[三、项目数据分析 3](#_Toc23016806)

[3.1数据预处理与分析 3](#_Toc23016807)

[3.1.1数据清洗、提取及描述 3](#_Toc23016808)

[3.1.2统计交易额及订单量 3](#_Toc23016809)

[3.1.3统计各售货机每月每单平均交易额及日均订单量 5](#_Toc23016810)

[3.2数据分析与可视化 6](#_Toc23016811)

[3.2.1 6月份销量前五的商品销量柱状图 6](#_Toc23016812)

[3.2.2各售货机每月交易总额折线图及交易额月环比增长率柱状图 7](#_Toc23016813)

[3.2.3各售货机毛利润占比饼图 12](#_Toc23016814)

[3.2.4每月交易额均值气泡图 13](#_Toc23016815)

[3.2.5 6、7、8月份热力图 13](#_Toc23016816)

[3.3生成自动售货机画像 14](#_Toc23016817)

[3.3.1对饮料类商品生成标签 14](#_Toc23016818)

[3.3.2售货机画像 16](#_Toc23016819)

[3.4项目预测 18](#_Toc23016820)

[3.4.1指数平滑法 18](#_Toc23016821)

[3.4.2 预测交易额 18](#_Toc23016822)

一、项目简介

## 1.1项目背景

目前，随着移动商务的发展，自动售货机作为一种方便、直观的移动商务工具，开始在全球各地迅速普及和发展。在中国，自动售货机将成为一个潜在的巨大产业，继百货商店、超市之后掀起第三次零售业革命，其前景非常广阔。从自助售卖机进入中国市场以来，其市场的规模一直在扩大，需求量也一直在加大。自动售卖机的发展顺应了社会发展的需求。2018年，我国自动售货机行业保有量达到 50.5 万台，主要分布在东部经济发达地区，机场、火车站(地铁)、广场及小区、学校、医院、商业办公区域、商业街等场所人流量比较大，且潜在需求比较高的产所。

自动售货机的各种商品销售情况，以及不同地点不同时间各台售货机销售的商品类型和销售额的对比是商家进行货品补给和配置的依据，同时为商家制定提高销售额获取更大利润的方案提供参考。因此，对售货机各方面的数据进行分析对商家来说是必要的，这样才能为消费者提供更优质更合适的商品，从而在市场获取更大比例的利润，商家才能更好更长远经营下去。

## 1.2项目目标

项目在不同地点选取5台具有代表性的自动售货A、B、C、D、E对其各类商品进行数据分析，以此预测总体的销售情况。

（1）计算每台商品交易额、订单量、交易额总额、订单总量，平均交易额和订单量；

（2）绘制关于商品销量，交易额，环比增长率、毛利润、订单量的相应图标；

（3）分析售货机商品数据，总结规律，对每台售货机进行标签化，使其能够好的展现销售情况的特征。

（4）依据分析结果预测未来销售额，分析数据不足，向经营者提出数据方面的要求

二、项目数据来源

## 2.1数据的获取

选取不同地点的5台售货机A、B、C、D、E商品销售数据

## 2.2数据的时间段

从2017年1月1日到2017年12月31日

## 2.3数据的大小

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 售货机 | A | B | C | D | E | 合计 |
| 数据大小（条） | 10486 | 13481 | 14493 | 8713 | 23505 | 70678 |

# 三、项目数据分析

## 3.1数据预处理与分析

### 3.1.1数据清洗、提取和描述

（1）数据导入及更改。将附件1导入python查看数据，将“支付时间”时间类型转化为datetime型数据，并通过对时间的遍历将“支付时间”的月份和日期提取出来作为新的一列，以便后面数据分析。

（2）数据清洗。对附件1的数据去除重复值、缺失值、异常值。

（3）数据描述。查看数据大概的描述性统计，输出如图1。因为应付金额和实付金额的数据结构相同，所以项目后续都以实际金额为代表进行数据分析。



图1 描述性统计

（4）数据提取。通过索引提取每台售货机对应的销售数据，将五个数据表分别导出并存为相应的csv文件。

### 3.1.2统计交易额及订单量

（1）分别统计每台售货机各月份交易总额及订单总量。通过以月份作为索引，对售货机A、B、C、D、E创建数据透视表。对每月实际金额求和得到每月交易总额，对订单号计数得到每月订单总量。其中售货机A输出结果如下：



图2 A点售货机各月份交易额及订单量

（2）汇总每台售货机2017年5月份的交易额、订单量。结果如表1：

表1 每台售货机2017年5月份的交易额、订单量

|  |  |  |
| --- | --- | --- |
| 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  合计 | 5699.0  18886.8 | 1292  4279 |

（3）统计所有售货机的交易总额及订单总量。以月份为索引，创建含有所有售货机数据的数据透视表，输出结果如下：

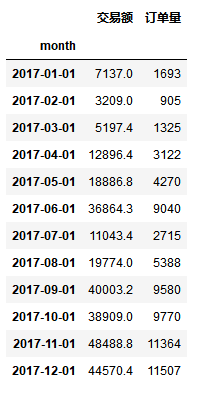


图3 所有售货机各月份的交易额及订单量

对交易额和订单量求和得到所有售货机的交易总额为286979.7元，订单总量为70679单。

### 3.1.3统计各售货机每月每单平均交易额及日均订单量

（1）计算每单平均交易额。分别对各台售货机的数据以月份为索引，创建数据透视表，对实际金额统计平均值，从而得到每单平均交易额。其中售货机A输出结果如下：

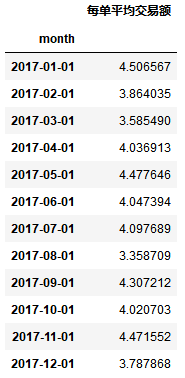


图4 A点售货机每月每单平均交易额

（2）分别统计每月的日均订单量。分别对售货机A、B、C、D、E的数据以月份及日作为索引创建数据透视表，对订单号进行计数处理，得到每日订单总量，进而再对每日订单总量按照月份作为索引，统计平均值，从而得到各售货机每月的日均单量。其中售货机A的结果如图：



图5 A点售货机每月日均订单量

（3）汇总售货机A、B、C、D、E的每月每单平均交易额（保留2位小数）及日均订单量（取整数），汇总结果如下表：



图6 A、B、C售货机每月每单平均交易额及日均订单量



图7 D、E售货机每月每单平均交易额及日均订单量

## 3.2数据分析与可视化

### 3.2.1 6月份销量前五的商品销量柱状图

以月份为作为索引，提取6月份的所有商品销售数据，创建数据透视表，再以商品作为索引，通过对订单号的计数得到相应的销售量。对销售量进行降序排序，并提取前5的商品，对其绘制销量柱状图，输出结果如下：

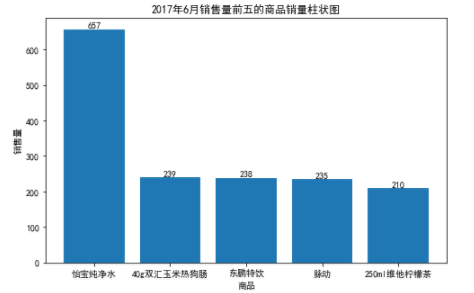


图8 2017年6月销量前5的商品的销量

### 3.2.2各售货机每月交易总额折线图及交易额月环比增长率柱状图

（1）：每月交易总额折线图如下：



图9 A点售货机每月交易总额折线图



图10 B点售货机每月交易总额折线图

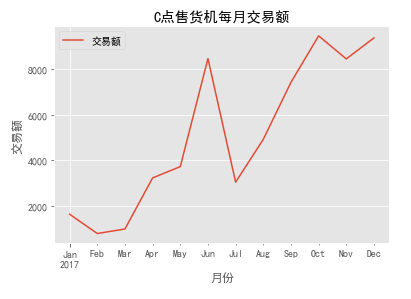


图11 C点售货机每月交易总额折线图

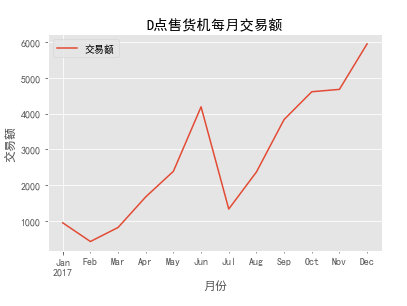


图12 D点售货机每月交易总额折线图

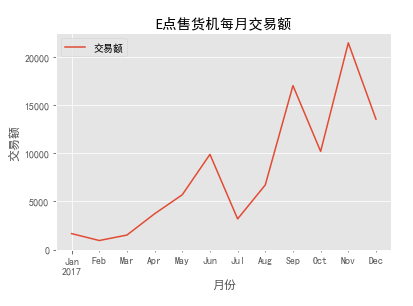


图13 E点售货机每月交易总额折线图

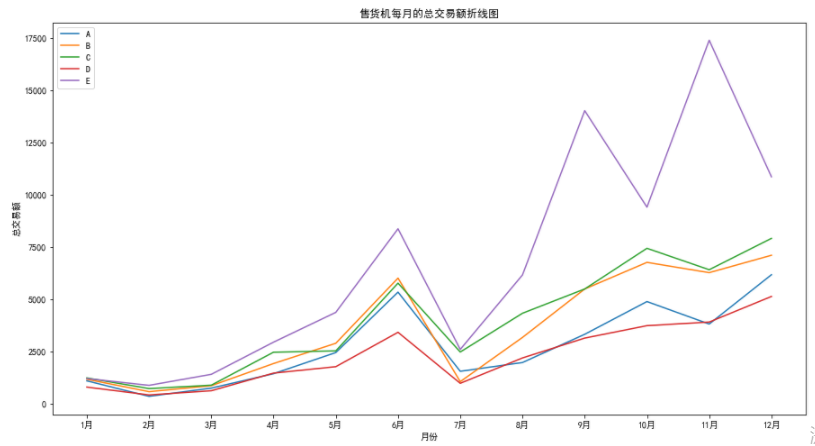


图14 各售货机每月交易总额折线图

（2）计算各每月交易额月环比增长率。环比增长率，一般是指和上期相比较的增长率。即环比增长率=(本期的某个指标的值-上一期这个指标的值)/上一期这个指标的值\*100%。那么

首先创建数据透视表，统计各售货机每月的交易额，然后通过差值函数计算（本月交易额-上月交易额），最后用错位函数提取上月交易额，从而得到月环比增长率。其中A点售货机输出结果如下：



图15 A点售货机月环比增长率

对各售货机的环比增长率绘制柱形图，输出如下：

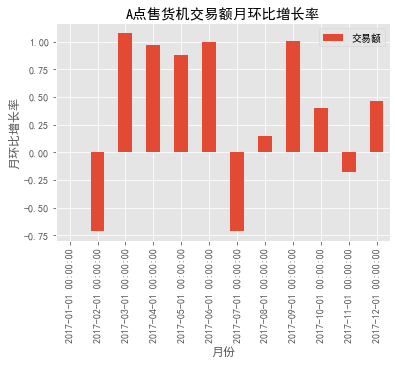


图16 A点售货机交易额月环比增长率柱状图

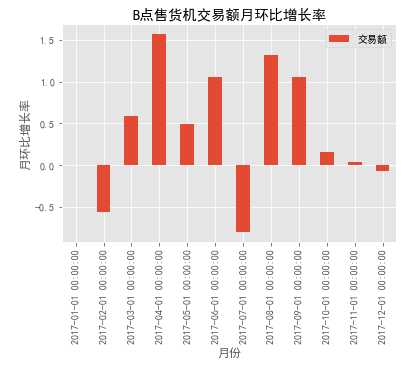


图17 B点售货机交易额月环比增长率柱状图

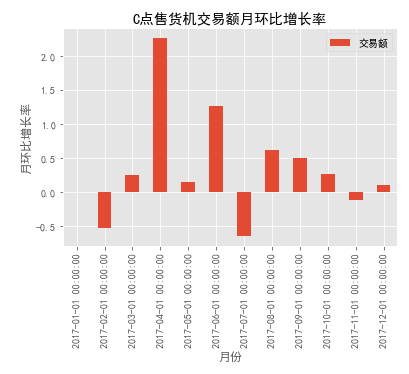


图18 C点售货机交易额月环比增长率柱状图

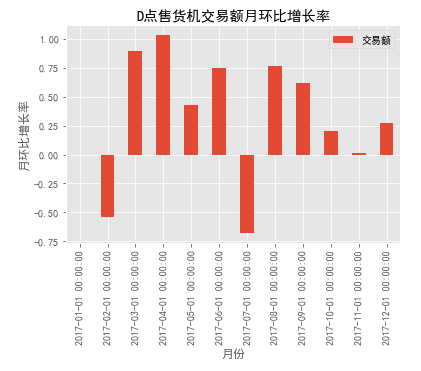


图19 D点售货机交易额月环比增长率柱状图

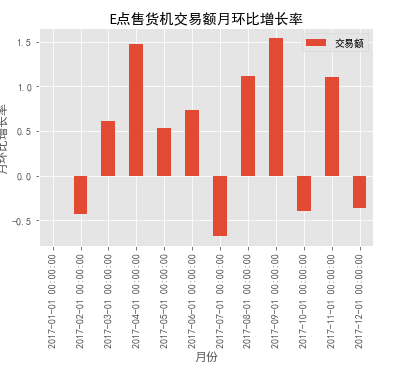


图20 E点售货机月环比增长率柱状图

### 3.2.3各售货机毛利润占比饼图

Step1：将数据与附件2中的数据进行合并，进行内连接。对总数据及5台售货机数据都与附件2内连接。因为饮料类（0.25）和非饮料类（0.20）的毛利率不同，因此将饮料和非饮料数据分别进行索引，分别对他们的销售金额求和，然后再跟利率相乘从而得到毛利润。售货机A、B、C、D、E的毛利润占比汇总如下表：

表2 各售货机毛利润占比

|  |  |
| --- | --- |
| 售货机 | 毛利润占比（元） |
| A | 14.58% |
| B | 18.84% |
| C | 21.46% |
| D | 11.57% |
| E | 33.55% |

对售货机A、B、C、D、E毛利润占比绘制饼图，如下：

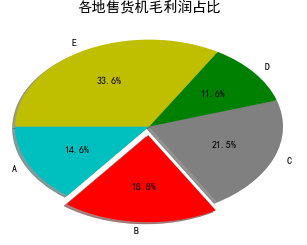


图21各地售货机毛利润占比饼图

### 3.2.4每月交易额均值气泡图

以月份和二级类目作为索引创建数据透视表，统计实际金额的平均值作为每月交易额均值。以月份作为x轴，二级类目作为y轴，每月交易额均值作为气泡大小，绘制气泡图如下：

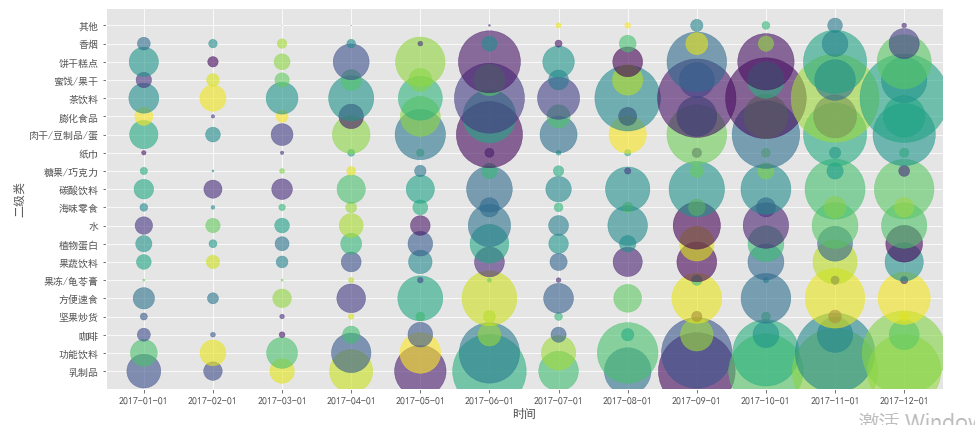


图22每月交易额均值气泡图

### 3.2.5 6、7、8月份热力图

在售货机C的数据中，从支付时间中提取小时及天作为新的一列添加到数据中。以小时为行，以天为列创建数据透视表，对订单量进行计数统计。通过条件函数对其绘制热力图。绘制C售货机6、7、8月份订单量热力图如下，纵轴以小时为单位，横轴以天为单位：

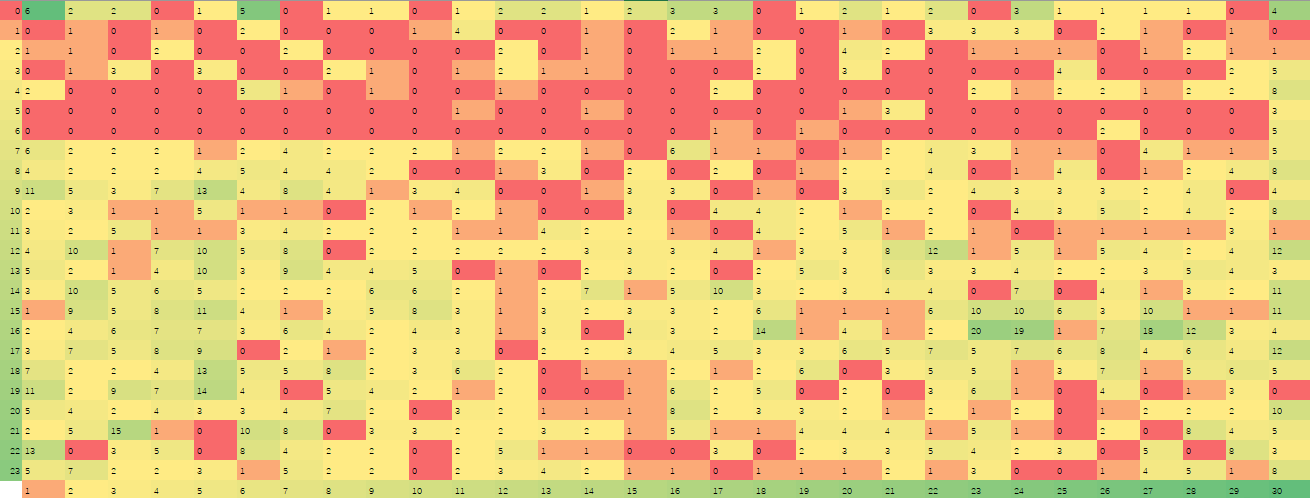


图23 C售货机6月份订单量热力图

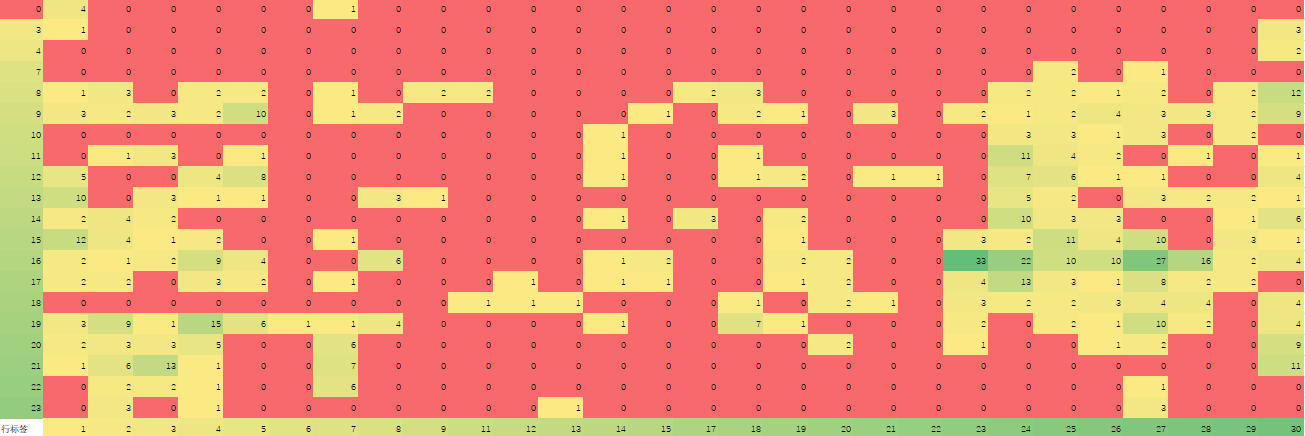


图24 C售货机7月份订单量热力图

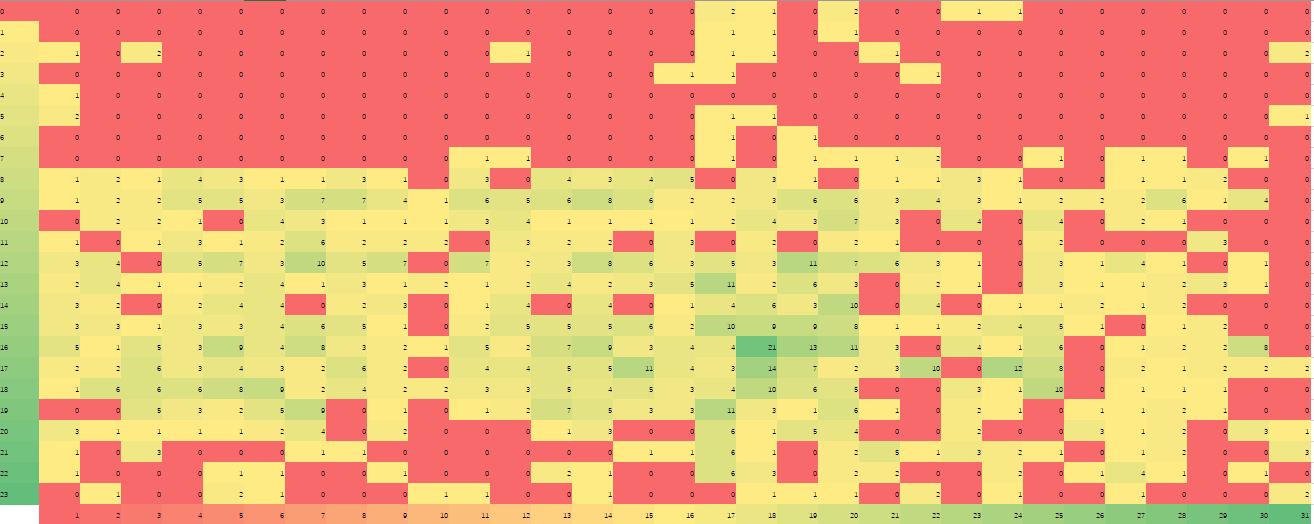


图25 C售货机8月份订单量热力图

## 3.3生成自动售货机画像

### 3.3.1对饮料类商品生成标签

（1）制定标签判断的数据标准。将所有售货机的数据进行合并，提取出饮料类的商品销售数据，按商品类型进行分类后对订单号进行计数得到各商品订单量，并对其进行描述统计分析，结果如下：

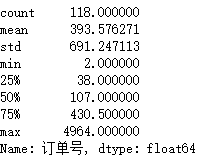


图26 各类商品订单量描述统计分析

通过输出结果可知：饮料类共有118种商品，订单量的平均数为394，最小值订单量为2，最大订单量为4964，订单量中位数为107，订单量下分位数为38，订单量上分位数为430。由此可以规定把订单量在[0,38]区间内认为是滞销，在[38,430]区间内认为是正常，在[430，4964]区间内认为是畅销。

（2）同样对售货机A、B、C、D、E的销售数据与类型数据合并，提取出饮料类产品的数据，统计各类商品的订单量，根据（1）得到的分类进行进行贴标签得到各类商品的销售标签。其中售货机A商品标签结果如下：

表3 A点售货机部分商品标签

|  |  |  |
| --- | --- | --- |
| 序号 | 饮料类商品 | 标签 |
| 1 | 100g\*5瓶益力多 | 正常 |
| 2 | 13g雀巢咖啡1+2特浓 | 滞销 |
| 3 | 145ml旺仔牛奶盒装 | 正常 |
| 4 | 145ml旺仔牛奶罐装 | 滞销 |
| 5 | 150g健能酸奶原味 | 滞销 |
| 6 | 180ml雀巢咖啡罐装 | 滞销 |
| 7 | 205g伊利安慕希原味 | 正常 |
| 8 | 205ml安慕希蓝莓味 | 滞销 |
| 9 | 250ML东鹏特饮 | 滞销 |
| 10 | 250ml燕塘原味酸奶 | 正常 |
| 11 | 250ml燕塘甜牛奶 | 正常 |
| 12 | 250ml燕塘红枣牛奶 | 正常 |
| 13 | 250ml王老吉盒装 | 滞销 |
| 14 | 250ml红牛 | 正常 |
| 15 | 250ml统一麦香奶茶 | 滞销 |
| 16 | 250ml维他原味豆奶 | 正常 |
| 17 | 250ml维他奶低糖原味 | 滞销 |
| 18 | 250ml维他奶巧克力味 | 正常 |
| 19 | 250ml维他奶黑豆奶饮品 | 滞销 |
| 20 | 250ml维他柠檬茶 | 正常 |

### 3.3.2售货机画像

（1）通过power BI分别对售货机A、B、C、D、E饮料类商品的订单量绘制词云图，以此作为相应售货机的画像。输出结果如下：



图27 A点售货机画像



图28 B点售货机画像



图29点售货机画像



图30 D点售货机画像



图31 E点售货机画像

由该售货机画像可看出，售货机A、B、C、E中销量最好的是怡宝纯净水，在D中其销量也很好，所以整体上怡宝纯净水在各售货机是最热销的产品。售货机A销量前3的分别是怡宝纯净水、东鹏特饮、阿萨姆奶茶，售货机B销量前3的分别是怡宝纯净水、东鹏特饮、阿萨姆奶茶，售货机C销量前3的分别是怡宝纯净水、脉动、东鹏特饮，售货机D销量前3的分别是东鹏特饮、怡宝纯净水、阿萨姆奶茶，售货机E销量前3的分别是怡宝纯净水、脉动、营养快线。总体上，在各售货机中怡宝纯净水，东鹏特饮，阿萨姆奶茶比较热销的商品。对于像小茗同学，芦荟汁这类边缘产品，销售效果差，商家应减少对他们的购入。由此可以看出销量高的商品都大致相同，由地点摆放的不同有一定的区别。

（2）营销建议。自动售货机行业在我国有将近10年的累计，增长速度逐渐加快，已经进入了大发展的阶段。目前全球的售货机行业向中国等发展中国家进行成本的战略转移，这为中国的售货机企业提供良好的发展契机。

1. 增加商品品种，促使使用者更多的使用售货机，增加对自动售货机使用的兴趣。
2. 根据自动售货机市场的淡季和旺季调整价格，在旺季，可以以相对较高的价格进行销售，在淡季，可以适当降低价格。对于滞销类产品，在长期积压下，应考虑降价或折扣销售，减少损失。
3. 迎合消费者需求，对需求较大的商品（如：怡宝纯净水）可以大量购入，对于需求较小的产品可以减少购入，甚至不购入，二购新的商品，看看是否符合消费者需求，寻找新的能迎合消费者需求的热销产品。
4. 对人流量高，销售量高的地方可以多放置几台售货机，在销量低的地方减少自动售货机的数量

## 3.4项目预测

### 3.4.1指数平滑法

指数平滑预测法是中短期发展趋势预测的常用方法，是比较有效的销售额预测方法。指数平滑是用得最多的一种。简单的全期平均法是对时间数列的过去数据一个不漏地全部加以同等利用;移动平均法则不考虑较远期的数据，并在[加权移动平均法](https://baike.so.com/doc/361350-382792.html)中给予近期资料更大的权重;而指数平滑法则兼容了全期平均和移动平均所长，不舍弃过去的数据，但是仅给予逐渐减弱的影响程度，即随着数据的远离，赋予逐渐收敛为零的权数。因此项目使用通过指数平滑法的方法预测2018年1月的交易额。

## 3.4.2 预测交易额

合并售货机A、B、C、D、E销售数据与类型数据，以月份和大类作为分组依据，统计不同大类各个月份的交易额，并将统计结果导入csv文件中。通过FORECAST.ETS函数进行指数平滑法预测2018年1月交易额。预测结果如下表：

表3 2018年1月每个售货机每个大类的交易额预测结果

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 大类\地点 | A | B | C | D | E |
| 非饮料 | 3001.287 | 3393.074 | 3749.298 | 1977.633 | 5846.731 |
| 饮料 | 4199.335 | 6378.357 | 6360.203 | 3925.011 | 12444.32 |