

自动售货机商务数据分析

目录

自动售货机商务数据分析	1
一、背景介绍	1
二、数据说明	1
三、自动售货机的销售情况的分析	1
四、业务预测	9

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

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。

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

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。附件 1 提供了从 2017 年 1 月 1 日至 2017 年 12 月 31 日每台自动售货机的商品销售数据，附件 2 提供了商品的分类。

三、自动售货机的销售情况的分析

1、数据预处理与分析

1.1 根据附件 1 中的数据，提取每台售货机对应的销售数据，保存在 CSV 文件中，文件名分别为“task1-1A.csv”、“task1-1B.csv”、...、“task1-1E.csv”。

1.2 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交易总额和订单总

量。

售货机	A	B	C	D	E	Sum
交易额	3385.1	3681.2	3729.4	2392.1	5699.0	18886.8
订单量	756	869	789	564	1292	4270

从上表可以看出在 2017 年 5 月中售货机 E 的商品的销售情况最好，无论是交易额还是订单量都是最多的，是同时间段销售情况最差的售货机 D 的两倍多，而售货机 A、B、C 的销售情况差不多，处于中等水平。

1.3 计算每台售货机每月的每单平均交易额与日均订单量

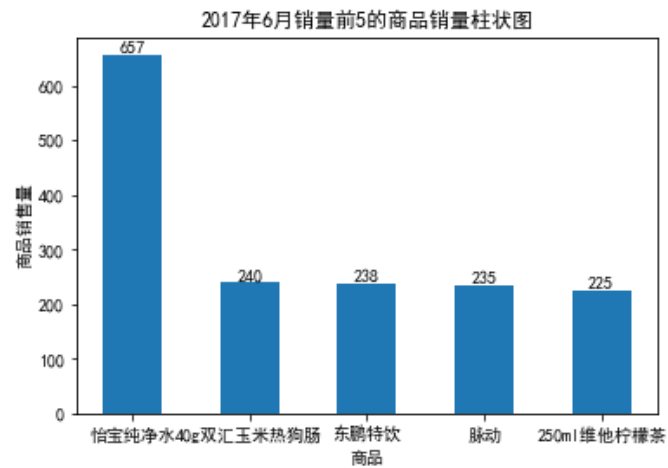
售货机每月的每单平均交易额	A	B	C	D	E
1 月	4.506567	3.753005	4.328496	3.692664	4.680226
2 月	3.864035	3.255676	3.826087	3.088652	3.638372
3 月	3.585490	3.614717	3.769962	4.305729	4.305714
4 月	4.036913	4.075290	4.403678	3.790293	4.159888
5 月	4.477646	4.236133	4.726743	4.241312	4.410991
6 月	4.047394	4.068050	4.501700	4.025962	3.817856
7 月	4.097689	4.401739	3.988351	4.229653	3.919311
8 月	3.358709	3.584200	3.913582	3.316503	3.804471
9 月	4.307212	4.130258	4.427294	3.899390	4.125375
10 月	4.020703	4.112340	4.273330	3.884233	3.676125
11 月	4.471552	4.268784	4.352393	3.862314	4.283227
12 月	3.787868	3.667014	3.943043	3.572580	4.168973

售货机每月的日均订单量	A	B	C	D	E
1 月	10.80645161	11.80645161	12.22580645	8.35483871	11.41935484
2 月	4.07142857	6.60714286	7.39285714	5.03571429	9.21428571
3 月	8.22580645	8.5483871	8.48387097	6.19354839	11.29032258
4 月	14.9	20.1	24.46666667	14.76666667	29.83333333
5 月	24.38709677	28.03225806	25.4516129	18.19354839	41.67741935
6 月	55.63333333	61.86666667	62.73333333	34.66666667	86.43333333
7 月	15.35483871	11.12903226	24.64516129	10.22580645	26.22580645
8 月	21.48387097	31.64516129	40.61290323	23.06451613	57.
9 月	34.66666667	58.16666667	55.93333333	32.76666667	137.8
10 月	50.48387097	65.35483871	71.48387097	38.25806452	89.58064516
11 月	38.66666667	67.7	64.76666667	40.33333333	167.33333333
12 月	64.61290323	71.29032258	76.74193548	53.64516129	104.90322581

从上表可以看出每台售货机每月的每单平均交易额都差不多，但从每台售货机每月的日均订单量可知售货机 E 除了 1 月份外其他月份的日均订单量都是最高的，说明售货机 E 的销售量真的很好。

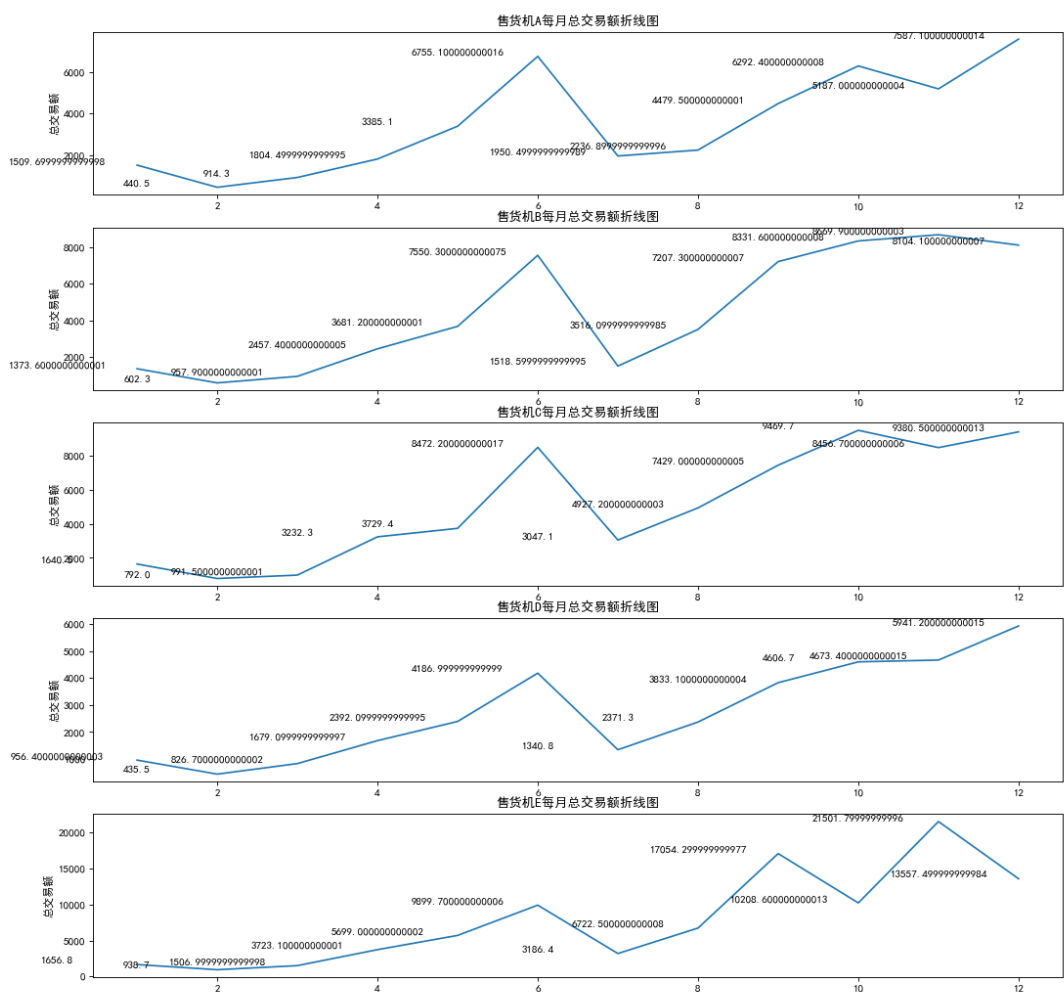
2、数据分析与可视化

2.1 绘制 2017 年 6 月销量前 5 的商品销量柱状图

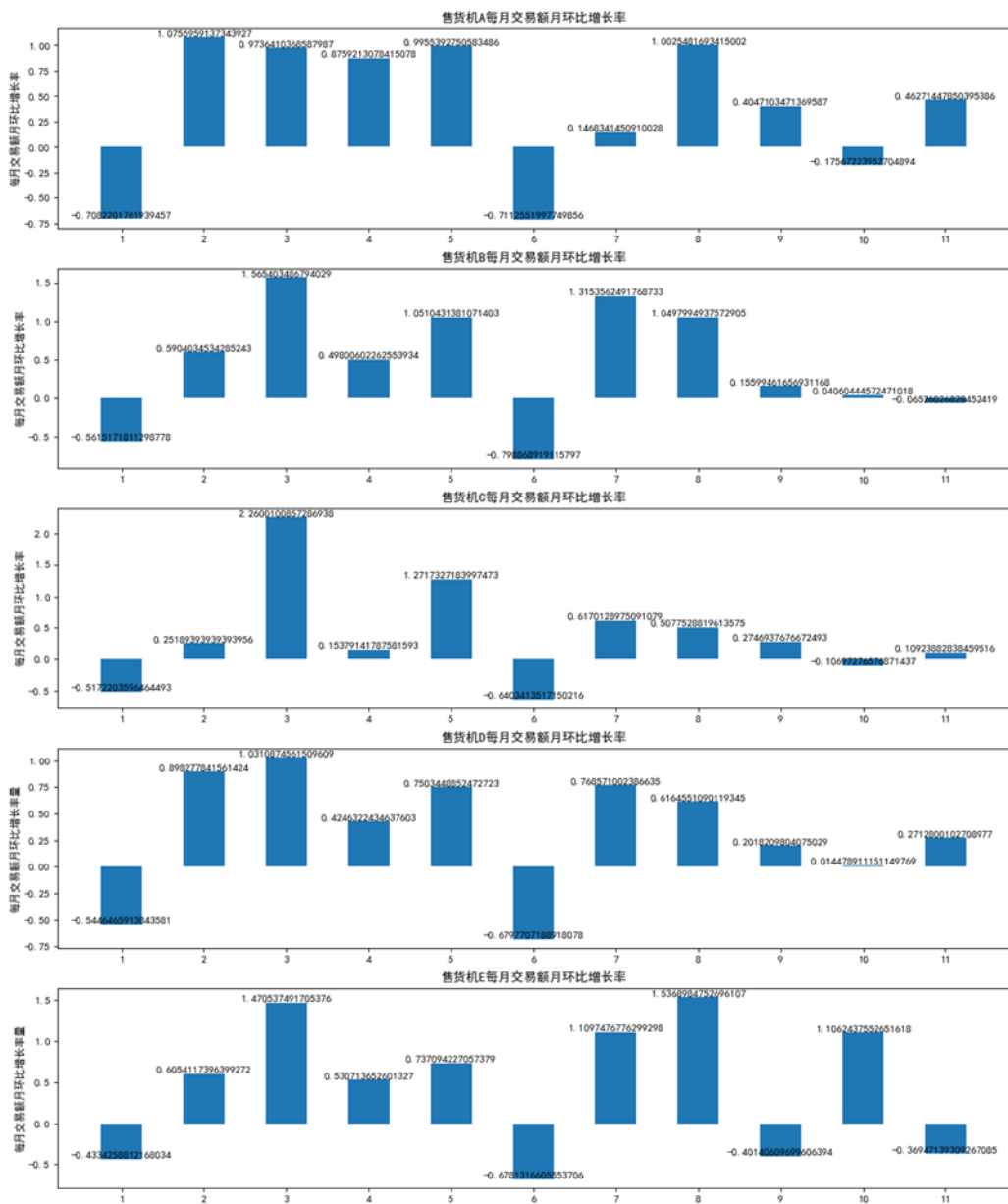


从上图 2017 年 6 月销量前 5 的商品销量柱状图可知除了 40g 双汇玉米热狗肠是非饮料，其他的是饮料，其中怡宝纯净水的销量是所有商品之中最高的，是前五其他商品的销量的 2~3 倍，所以整体可知因为时间气温的原因，在 2017 年 6 月中饮料的需求很大，在销售过程中占很大的优势。

2.2 绘制每台售货机每月总交易额折线图及交易额月环比增长率柱状图



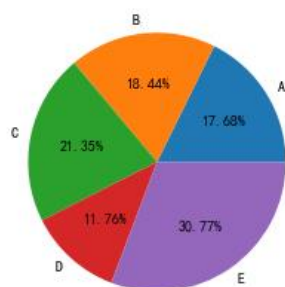
每台售货机交易额月环比增长率柱状图



从上图每台售货机每月总交易额折线图和交易额月环比增长率柱状图可知，售货机的交易额整体都是上升的趋势，在6月份到达最高峰，但紧接的7月份却下降的很厉害，8月份趋势才回暖，可见7月份必定有很重要的因素发生影响到7月份的销售额，使其急剧下降。

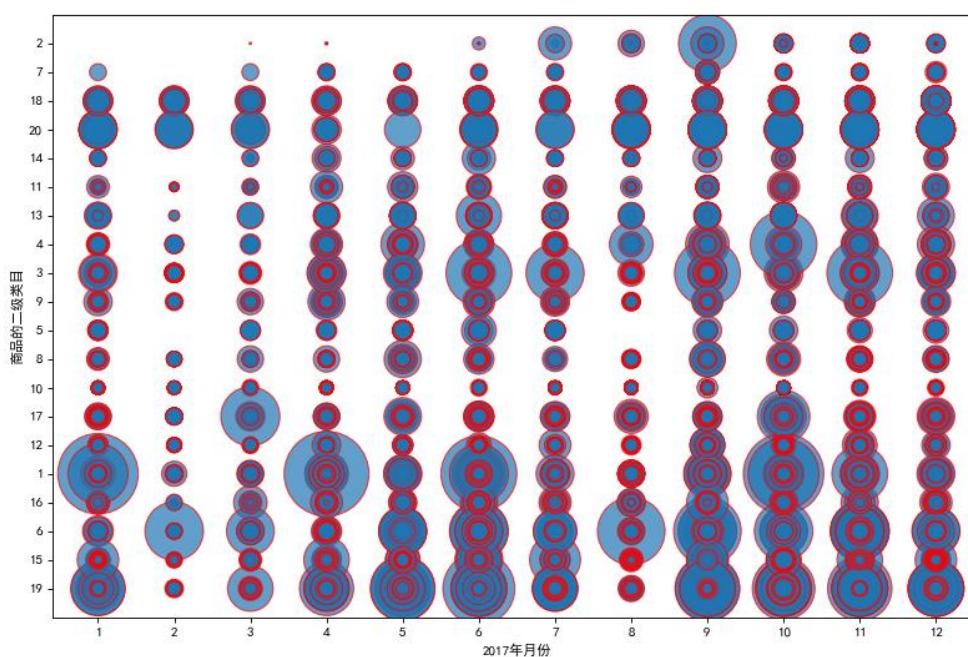
2.3 绘制每台售货机毛利润占总毛利润比例的饼图（假设饮料类毛利率为25%，非饮料类为20%）

每台售货机毛利润占总毛利润比例的饼图（假设饮料类毛利率为25%，非饮料类为20%



从饼图可以看中售货机 E 的毛利润占总的毛利润最大达到 30.77%，是毛利润最小的售货机 D 的 3 倍左右。

2.4 绘制每月交易额均值气泡图，横轴为时间，纵轴为商品的二级类目



从每月交易额均值气泡图可知

3、生成自动售货机画像

3.1 分析各售货机商品销售数据，总结规律，给出每台售货机饮料类商品的标签，结果保存在 CSV 文件中，文件名分别为 “task3-1A.csv”、“task3-1B.csv”、…、“task3-1E.csv”，格式如下表所示

中茶饮料，乳制品和碳酸饮料种类比较火，单一商品比较火热的是小茗同学、维他奶、可口可乐 330ml、王老吉等。

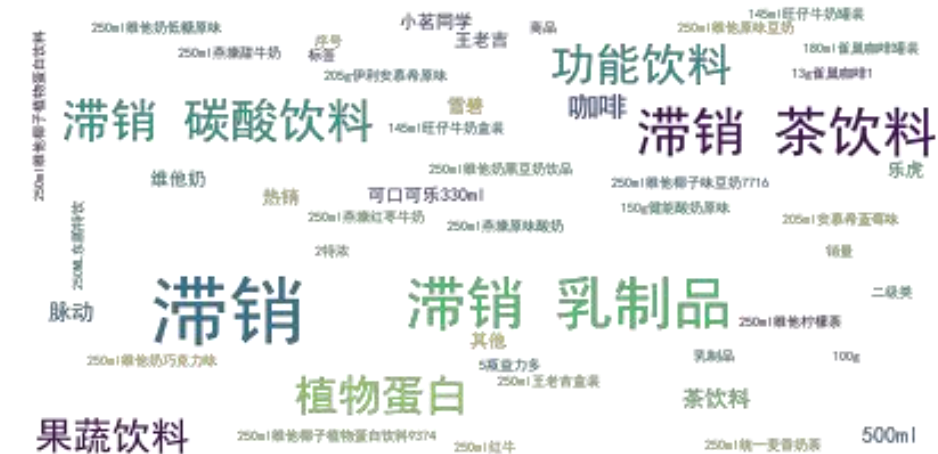


Figure 3 售货机 C

从售货机 C 的画像可知，售货机 C 的商品大多都是滞销品，正常的和热销的几乎没有，其中茶饮料和乳制品比较火，单一商品比较火热的是脉动、小茗同学、可口可乐 330ml 等。

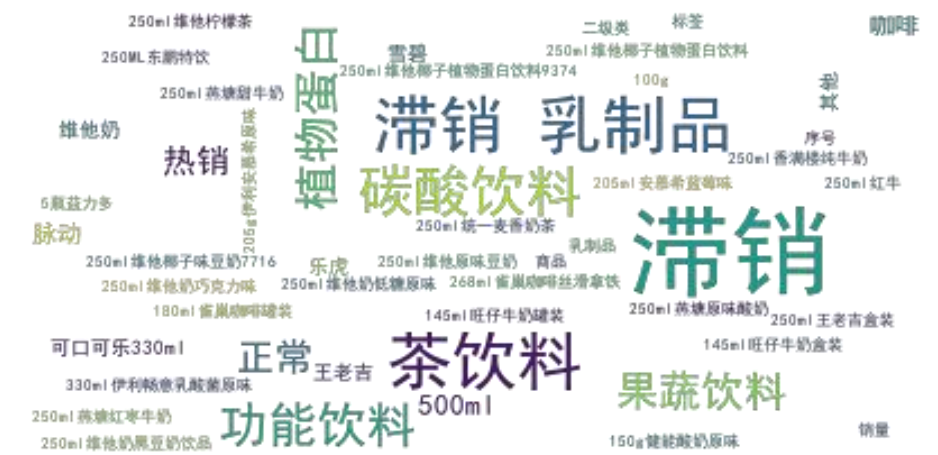


Figure 4 售货机 D

从售货机 D 的画像可知，售货机 D 的商品大多都是滞销品，正常的和热销的商品还是有的，其中茶饮料，乳制品和碳酸饮料种类比较火，单一商品比较火热的是脉动、可口可乐 330ml、雪碧、王老吉等。

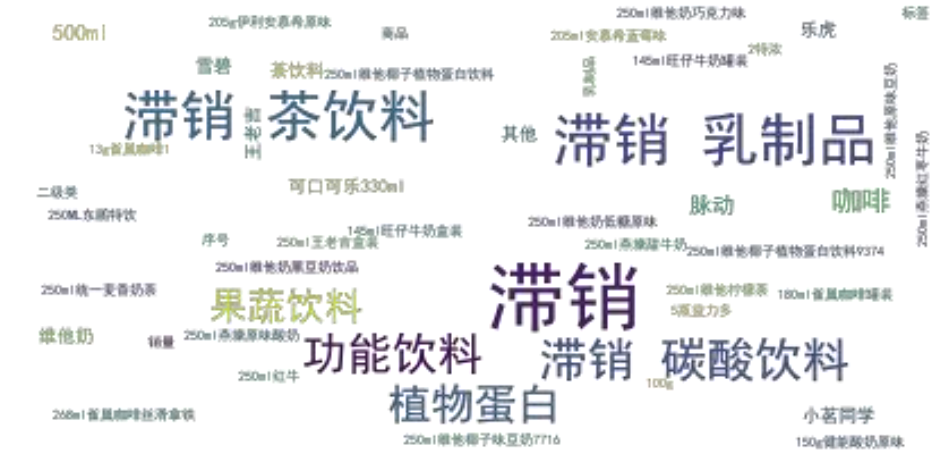


Figure 5 售货机 E

从售货机 E 的画像可知, 售货机 E 的商品大多都是滞销品, 正常的和热销的商品几乎没有的, 其中茶饮料, 乳制品种类比较火, 单一商品比较火热的是脉动、可口可乐 330ml、雪碧、维他奶等。

四、业务预测

假设自动售货机的经营者向数据分析师提出依据附件中的 1 年的数据来预测每台售货机的每个大类商品在 2018 年 1 月的交易额的要求。

1、预测未来销售额的原理是一元线性回归, 假设时间为解释变量 x , 交易金额为因变量 y , μ 为随机干扰项, 则模型为: $y = \beta_0 + \beta_1 x + \mu$

1.1 根据附件提供的数据对售货机 A 的每个大类商品进行分析, 模型评价为

非饮料数据线性回归模型的平均绝对误差为: 44.167683884204

非饮料数据线性回归模型的均方误差为: 3868.660465852127

非饮料数据线性回归模型的中值绝对误差为: 28.388753220546647

非饮料数据线性回归模型的可解释方差值为: 0.20686693163872627

非饮料数据线性回归模型的 R 方值为: 0.18402861851318197

饮料数据线性回归模型的平均绝对误差为: 44.167683884204

饮料数据线性回归模型的均方误差为: 3868.660465852127

饮料数据线性回归模型的中值绝对误差为: 28.388753220546647

饮料数据线性回归模型的可解释方差值为: 0.20686693163872627

饮料数据线性回归模型的 R 方值为: 0.18402861851318197

1.2 根据附件提供的数据对售货机 B 的每个大类商品进行分析, 模型评价为

非饮料数据线性回归模型的平均绝对误差为: 51.35984148098553

非饮料数据线性回归模型的均方误差为: 5801.496498113489

非饮料数据线性回归模型的中值绝对误差为: 29.804702578669044

非饮料数据线性回归模型的可解释方差值为: 0.14216127020491398

非饮料数据线性回归模型的 R 方值为: 0.06381112990668736

饮料数据线性回归模型的平均绝对误差为: 78.9632007010525

饮料数据线性回归模型的均方误差为: 11195.484928578

饮料数据线性回归模型的中值绝对误差为: 53.4659102421604

饮料数据线性回归模型的可解释方差值为: 0.22901682413220192

饮料数据线性回归模型的 R 方值为: 0.19877152264030373

1.3 根据附件提供的数据对售货机 C 的每个大类商品进行分析, 模型评价为

非饮料数据线性回归模型的平均绝对误差为: 44.48818269143234

非饮料数据线性回归模型的均方误差为: 3555.7485811211736

非饮料数据线性回归模型的中值绝对误差为: 28.80849736681285

非饮料数据线性回归模型的可解释方差值为: 0.2270507670162989

非饮料数据线性回归模型的 R 方值为: 0.19536726038481123

饮料数据线性回归模型的平均绝对误差为: 57.99840904772982

饮料数据线性回归模型的均方误差为: 6278.450512852121

饮料数据线性回归模型的中值绝对误差为: 35.56020085103788

饮料数据线性回归模型的可解释方差值为: 0.4464969002237039

饮料数据线性回归模型的 R 方值为: 0.43305742247769885

1.4 根据附件提供的数据对售货机 D 的每个大类商品进行分析, 模型评价为

非饮料数据线性回归模型的平均绝对误差为： 27.585298773346103
非饮料数据线性回归模型的均方误差为： 1491.4606718741961
非饮料数据线性回归模型的中值绝对误差为： 20.986032113240675
非饮料数据线性回归模型的可解释方差值为： 0.12735074706693006
非饮料数据线性回归模型的 R 方值为： 0.08811960142685837
饮料数据线性回归模型的平均绝对误差为： 32.58021220883844
饮料数据线性回归模型的均方误差为： 2059.7493632924784
饮料数据线性回归模型的中值绝对误差为： 21.10317177750089
饮料数据线性回归模型的可解释方差值为： 0.438322378168425
饮料数据线性回归模型的 R 方值为： 0.42570535709309143

1.5 根据附件提供的数据对售货机 E 的每个大类商品进行分析，模型评价为

非饮料数据线性回归模型的平均绝对误差为： 81.93518947058564
非饮料数据线性回归模型的均方误差为： 13857.610944091442
非饮料数据线性回归模型的中值绝对误差为： 50.311953354735266
非饮料数据线性回归模型的可解释方差值为： 0.2343363636975473
非饮料数据线性回归模型的 R 方值为： 0.21319837830841715
饮料数据线性回归模型的平均绝对误差为： 127.42496752644216
饮料数据线性回归模型的均方误差为： 38986.89655811519
饮料数据线性回归模型的中值绝对误差为： 67.34751218737361
饮料数据线性回归模型的可解释方差值为： 0.32883191670584644
饮料数据线性回归模型的 R 方值为： 0.3282771878298876

从上述所有无人售货机的数据分析得出来的平均绝对误差、均方差和中值绝对误差的值远远偏离 0，可解析方差值和 R 方值也是远小于 1，可知该模型的拟合效果不理想和性能差，随机干扰项过大。。

2、 附件所给数据不足以支持预测工作。无人售货机的销售额主要与收入水平、人口结构、消费习惯、政府有关政策规定和支持因素有关。一般说来，收入水平越高，流动人口比例越大、容易接受新鲜事物和喜欢时尚消费的年轻人越多的城市 and 地段，自动售货机的销售额就越大。可在提供该地点的购买人的年龄，职业，是否节假日，重要节假日重点说明，该地点的收入水平等数据，增加可解释变量，减少随机误差。