

新零售-无人智能售货机商务数据分析报告

一、背景介绍

分析背景

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。

分析任务

- 1、根据商品销售数据，对自动售货机进行商务数据分析，探讨销量，营业额，利润等之间的关系
- 2、绘制自动售货机画像，分析售货机商品热销程度，给出合理的营销意见。
- 3、了解预测的含义，掌握预测方法，尝试对预测次月的销售数据

二、任务一、数据预处理与分析

任务 1.1

对于所给数据附件 1，在拆分附件 1 中的销售数据前先使用 python 进行预处理。

- A、检测与处理重复值（无）
- B、检测与处理缺失值（无）
- C、检测与处理异常值（最后一行数据的日期为 2.29）

从运行结果可以看出数据无重复也无缺失但在数据末行出现日期异常值

得到预处理后的数据后，即可对附件 1 的数据进行拆分，得到每台售货机对应的销售数据，即 task1-1A, task1-1B, task1-1C, task1-1D, task1-1E,

任务 1.2:

利用任务 1.1 中的得到的各个数据以及附件 1，在导入后对其字符串时间实现 `pd.to_datetime()` 操作，使其格式转为标准时间。使用循环筛选出各个数据的月份作为 'month' 列保存在相对应的 dataframe 里，再对其 'month' 进行 groupby 操作，将相同月份的合并在一起。

- 1、对 groupby 所得的 dataframegroupby 用 `.size()[5]` 取出五月份的订单量
- 2、对 groupby 所得的 dataframegroupby 用 `.agg({'实际金额': np.sum})` 操作并用 `loc[5,:]` 提取出五月份的交易额

详细数据见下表：

2017 年 5 月份的交易额和订单量

售货机	交易额	订单量
A	3385.1	756
B	3681.2	869

C	3729.4	789
D	2393.1	564
E	5699.0	1292
汇总	18886.8	4270

任务 1.3

使用任务 1.2 groupby 之后得到的 dataframegroupby，先创建一个由 2017 年每个月日期数组成的的数组

- 1、对 dataframegroupby 进行.size()/a 的操作即可得到每月日均订单量(总订单量/日期数)
- 2、对 dataframegroupby 的['实际金额']进行 mean()操作,即得到了每月每单平均交易额

详细数据如下：

2017 每台售货机每月的每单平均交易额

月份	A 售货机的每单平均交易额	B 售货机的每单平均交易额	C 售货机的每单平均交易额	D 售货机的每单平均交易额	E 售货机的每单平均交易额
1	10.806452	11.806452	12.225806	8.354839	11.419355
2	4.071429	6.607143	7.392857	5.035714	9.214286
3	8.225806	8.548387	8.483871	6.193548	11.290323
4	14.900000	20.100000	24.466667	14.766667	29.833333
5	24.387097	28.032258	25.451613	18.193548	41.677419
6	55.633333	61.866667	62.733333	34.666667	86.433333
7	15.354839	11.129032	24.645161	10.225806	26.225806
8	21.483871	31.645161	40.612903	23.064516	57.000000
9	34.666667	58.166667	55.933333	32.766667	137.800000
10	50.483871	65.354839	71.483871	38.258065	89.580645
11	38.666667	67.700000	64.766667	40.333333	167.333333
12	64.612903	71.290323	76.741935	53.645161	104.903226

每台售货机每月的每单平均交易额

月份	A 售货机的日均订 单量	B 售货机的日均订 单量	C 售货机的日均订 单量	D 售货机的日均订 单量	E 售货机的日均订 单量
1	4.506567	3.753005	4.328496	3.692664	4.680226
2	3.864035	3.255676	3.826087	3.088652	3.638372
3	3.585490	3.614717	3.769962	4.305729	4.305714
4	4.036913	4.075290	4.403678	3.790293	4.159888
5	4.477646	4.236133	4.726743	4.241312	4.410991
6	4.047394	4.068050	4.501700	4.025962	3.817856
7	4.097689	4.401739	3.988351	4.229653	3.919311
8	3.358709	3.584200	3.913582	3.316503	3.804471
9	4.307212	4.130258	4.427294	3.899390	4.125375
10	4.020703	4.112340	4.273330	3.884233	3.676125

11	4.471552	4.268784	4.352393	3.862314	4.283227
12	3.787868	3.667014	3.943043	3.572580	4.168973

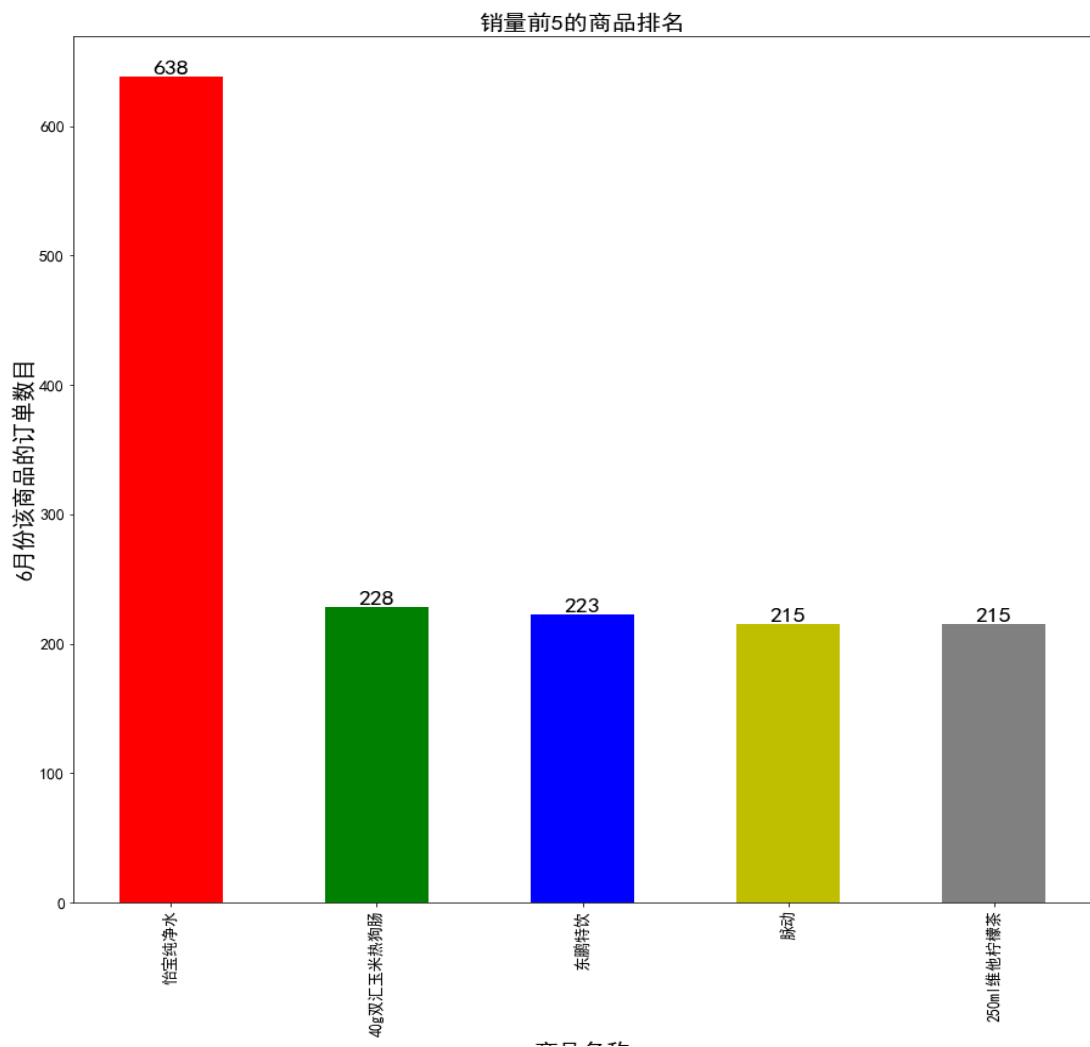
三、任务二

利用任务 1.1 中的得到的各个数据以及附件 1，附件 2，导入后对其进行 pd.to_datetime()和 pd.DatetimeIndex() 转化时间格式，方便接下来画图。

任务 2.1

使用 if 条件语句，将其在 6 月份的数据筛选出来，然后对这个 dataframe 数据进行 value_counts()操作，再使用 loc 筛选出销量前五的商品

- 1、进行.index 操作得到 x 轴刻度
 - 2、进行.values 操作得到各个商品的销量用于绘制柱状图
- 得到下图

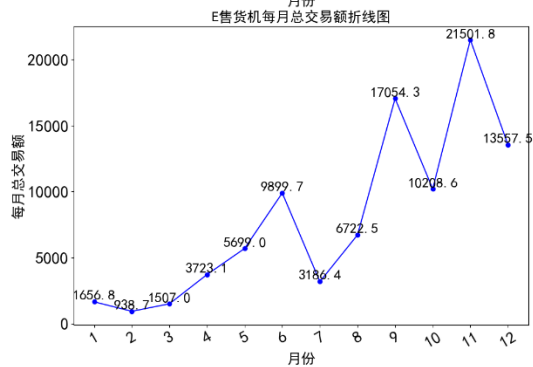
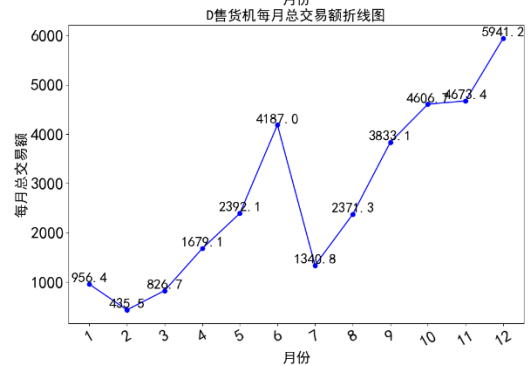
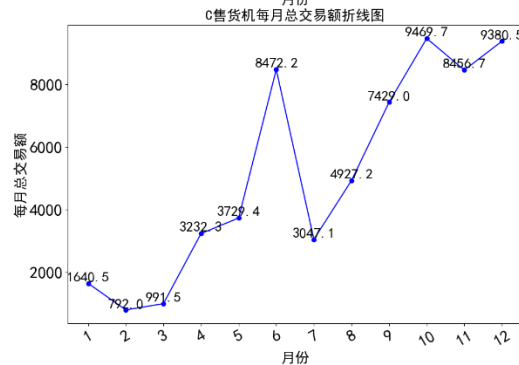
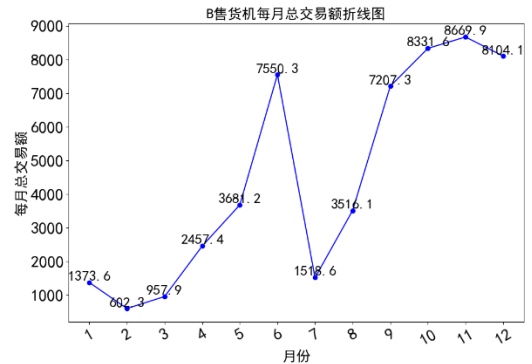
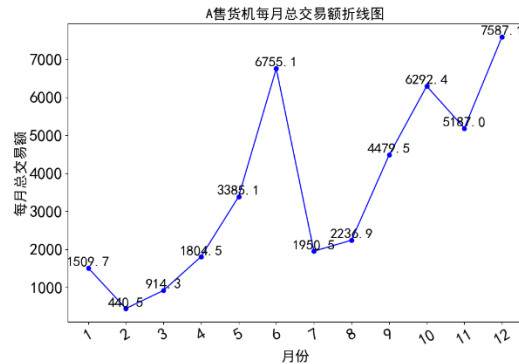


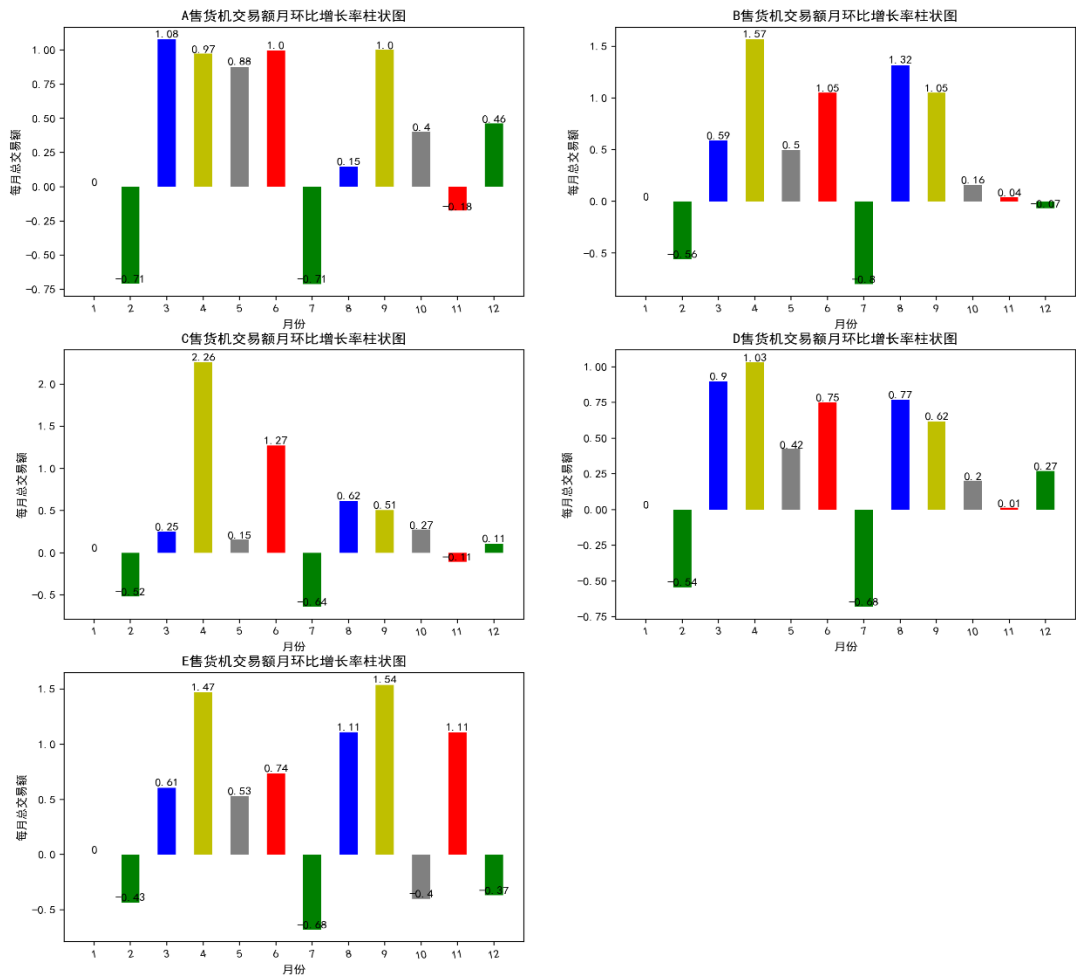
从图中可以看出 6 月份的销量前五均为饮料，且销量冠军怡宝矿泉水的销量要远远高于其他的饮料销量，可能是由于夏季原因，客户饮水需要，所以经营者可

以在夏季考虑多摆放饮料类商品

任务 2.2

先对 A-E 的数据进行 for 循环，提取出每个月的总交易额用于绘制折线图，提取出每个售货机的月环比用于绘制月环比柱状图，得到下图

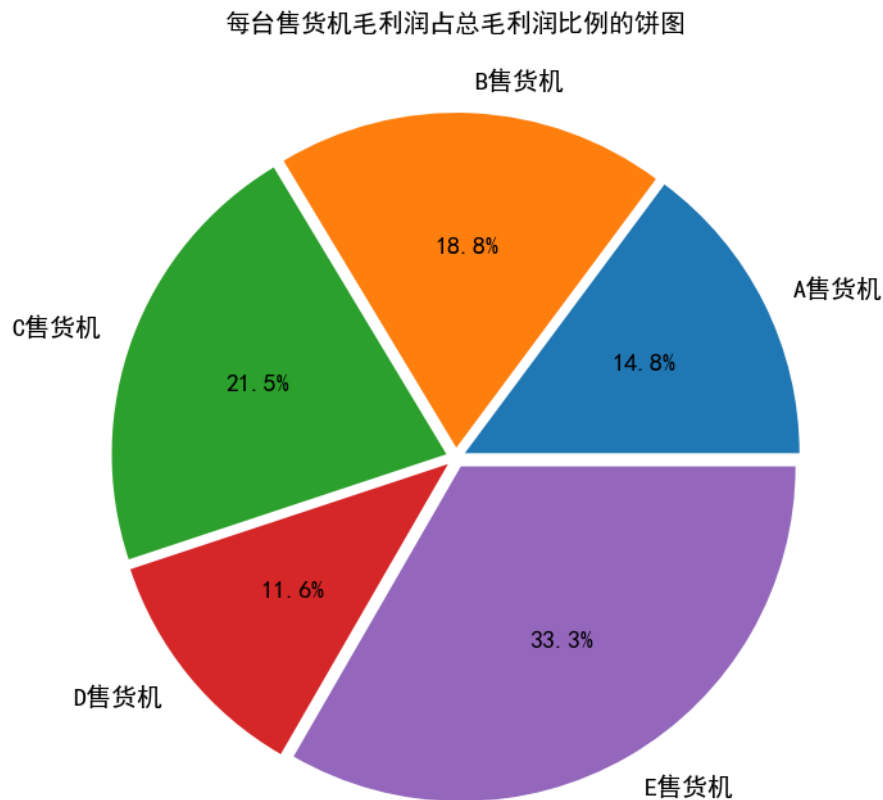




从图中看出每台售货机交易额的变化趋势都十分接近，大致都是上升趋势但在 6 月份过后会出现销量骤减。而如果分成 1 月-6 月，7 月-12 月两个部分，两个部分的交易额都是大致呈上升趋势，所以可以在夏冬两季增加各个类型商品供应。考虑到 1 月 2 月以及 7 月 8 月是寒暑假期间，故可能这五台售货机都是摆放在在学校附近的。故在学生的一些假期期间，运营者可以考虑减少供应，以免造成资金浪费。

任务 2.3

对各个数据进行 for 循环之后得到各个售货机的毛利润和总毛利润用于绘画饼图。得到下图



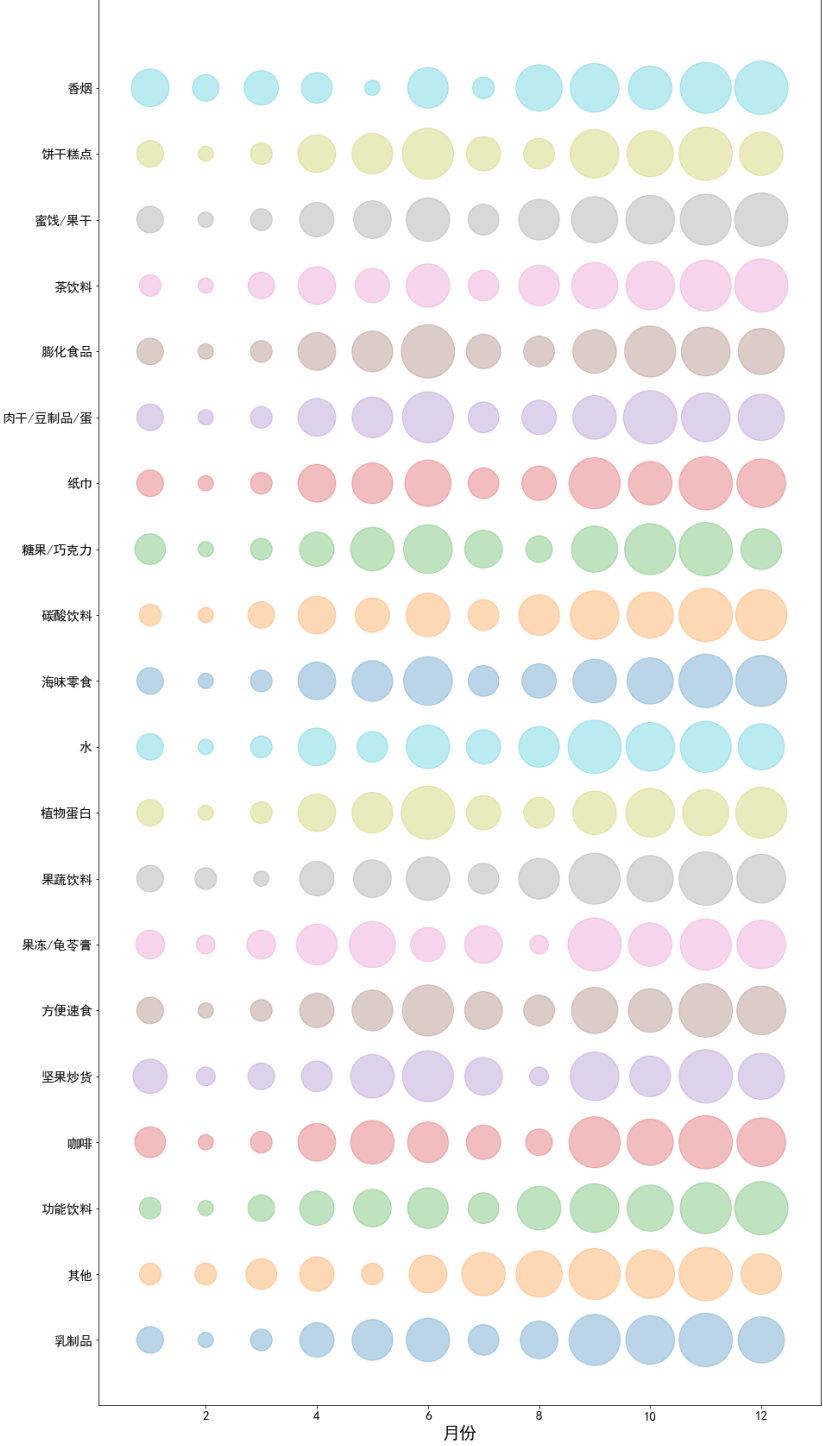
从图中可以看出各个售货机的毛利润高低排名为 E、C、B、A、D，E 的毛利润是总毛利润的 $\frac{1}{3}$ ，且 E 的毛利润接近于 D 的三倍。因此，可以考虑增加对 E 售货机的供应，减少对 D 的供应甚至移除。

-

任务 2.4

使用 `pd.merge` 合并附件 1 和附件 2，并使用循环筛选出各个数据的月份作为 'month' 列保存在相对应的 dataframe 里，之后使用 `groupby` 合并得到 dataframegroupby 数据用于绘制气泡图。得到下图

每月交易总额均值气泡图

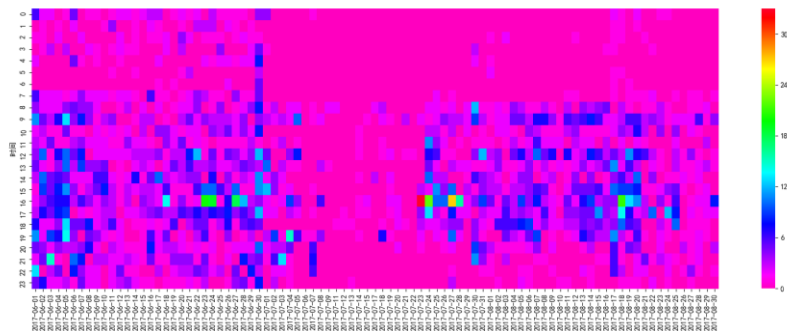


由图可知，对于每个月来说，每种类型的商品的交易额几乎都相差不多，虽然零食类，膨化食品类等这类商品的交易额会稍微高一点，几乎可以不用在类型上进行调整，需要的话就增加一些对零食类，膨化食品类的供应。

而对于每个类型来说，分成 1-6 月 和 7-12 月两部分，两个部分的交易额都成上升趋势，所以可以在夏冬两季增加各个类型商品的供应。

任务 2.5

使用 if 条件语句，将其在 678 月的数据筛选出来，然后对这个 dataframe 数据进行 groupby 操作，然后对其标准时间进行拆分成日期和小时，得到一个新的数据表，然后对其进行透视表操作用于绘制热力图。得到下图



从图中可以看出 6 月份的订单量明显大于 7 8 月的订单量，因此可以考虑在增加 6 月份的供应。并且每个月末的订单量都会大幅增高，所以在每个月末增加供应

任务三

对于热销正常滞销，由于每个售货机的情况不一样，无法给定确切的数字，所以我采取了以下定义

热销=销量排名前百分之十

滞销=销量排名后百分之五十

正常=滞销与热销之间

任务 3.1

对各个数据采取 groupby 操作后，使用.size()可以得到各个商品的订单量，得到一个新的数据表，对其进行 for 循环操作后得到了饮料类的标签。

任务 3.2

用 Excel 操作选取各个数据的商品名保存为 txt 格式，用于来绘制词云图，得到各个售货机的画像。字体越大代表订单量越高。

A 售货机画像



对于 D 售货机来说，明显怡宝纯净水订单量最高，所以对于 D 售货机可以增加矿泉水类的供应。

E 售货机画像

无法进行预测。

任务 4.2

经营者如果可以提供两年及两年以上的数据, 那么便可以通过时间序列预测方法, 得到 2018 年 1 月的交易额的预测, 数据越多则预测的会更加精准。