

目录

一、 背景介绍.....	2
1、 任务背景.....	2
2、 任务目标.....	2
二、 项目任务分析.....	2
1、 任务 1 数据预处理与分析.....	2
2、 任务 2 数据分析与可视化.....	2
3、 任务 3 生成自动售货机画像.....	2
4、 任务 4 业务预测.....	2

一、背景介绍

1、任务背景

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。

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。

2、任务目标

- 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，并预测每台售货机的销售额。
- 2、为每台售货机所销售的商品贴上标签，使其能够很好的展现商品的特征。

二、项目任务分析

1、任务 1 数据预处理与分析

1.1 根据附件 1 的数据，提取每台售货机对应的销售数据，保存到 csv 文件中，文件名分别为“task1_1A.csv”，“task1_1B.csv”，…，“task1_1E.csv”。解决方案是：先用 pandas.read_table 函数读取附件 1 数据，然后通过 loc 切片的方法提取出每台售货机对应的销售数据。

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，以表格的形式体现在报告中。解决方案是：通过任务 1.1 结果保存的文件“task1_1A.csv”，“task1_1B.csv”，…，“task1_1E.csv”，分别用 pandas.read_table 函数读取表中数据，再用 pandas.to_datetime 函数将表中的‘支付时间’列的时间类型数据转换为标准时间数据，对每台售货机的销售数据分别提取 2017 年 5 月的销售数据，并利用 sum 函数和 len 函数分别计算出每台售货机 2017 年 5 月份的交易额和订单量，相加得到所有售货机交易总额和订单总量。根据 task1-2 得到的结果如下表：

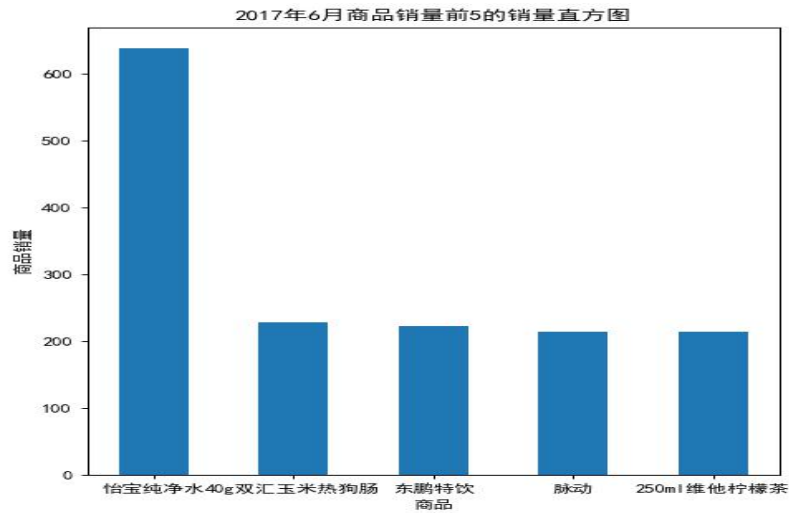
2017 年 5 月各售货机交易额与订单量						
售货机	A	B	C	D	E	总计
交易额	3385.1	3681.2	3729.4	2392.1	5699.0	286979.7
订单量	756	869	789	564	1292	70680

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，以表格的形式体现在 报告中。 解决方案与 1.2 大致相同，不同的是需要用 for 循环将每台售货机每月的数据分别提取出来，用 mean 函数求出平均交易额以及用 len 函数求出每月订单量，再除以每月天数得到日均订单量。（均保留两位小数）

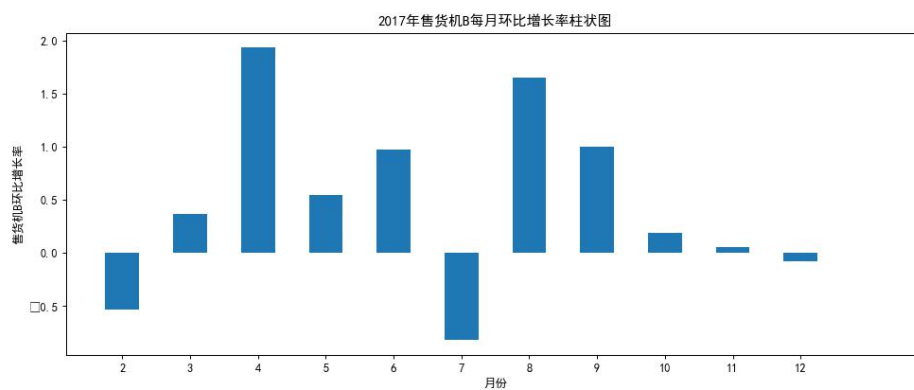
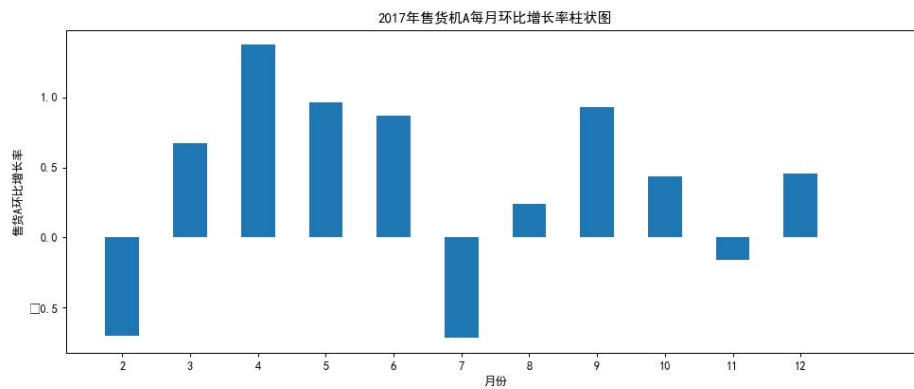
时间	售货机 A 每月每单平均交易额	日均订单量 A	售货机 B 每月每单平均交易额	日均订单量 B	售货机 C 每月每单平均交易额	日均订单量 C	售货机 D 每月每单平均交易额	日均订单量 D	售货机 E 每月每单平均交易额	日均订单量 E
201701	4.57	10.19	3.8	10.74	4.43	11.26	3.8	7.68	4.72	10.9
201702	3.87	4	3.27	6.5	3.82	7.29	3.09	5	3.64	9.07
201703	3.37	6.94	3.6	7.29	3.76	7.1	3.44	4.87	4.3	9.74
201704	4.03	14.23	4.12	19.3	4.37	23.53	3.79	14.03	4.17	28.13
201705	4.48	24.39	4.24	28.03	4.73	25.45	4.24	18.19	4.41	41.68
201706	4.09	51.6	4.08	59.3	4.54	58.23	4.03	33.33	3.81	82.83
201707	4.12	14.03	4.4	9.45	3.98	22.97	4.23	9.39	3.91	22.39
201708	3.36	21.29	3.6	30.65	3.92	40.26	3.33	22.74	3.81	56.71
201709	4.32	33.07	4.18	54.63	4.41	53.07	3.83	30.77	4.13	130.63
201710	4.02	49.32	4.11	64.03	4.27	70.81	3.89	37.71	3.65	87.77
201711	4.48	38.5	4.27	67.33	4.36	64.53	3.87	40.2	4.29	166.63
201712	3.78	64.1	3.66	70.35	3.95	75.84	3.59	52.61	4.19	103.48

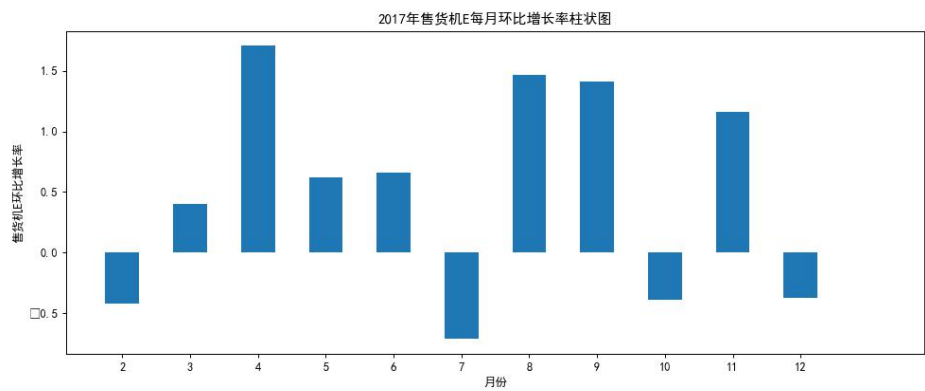
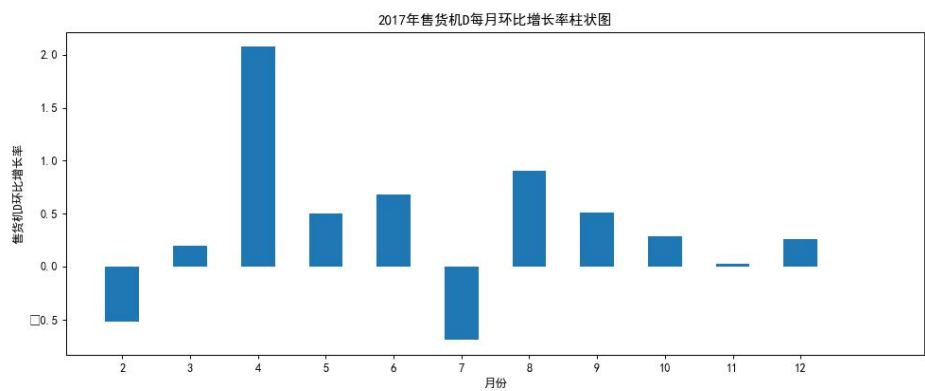
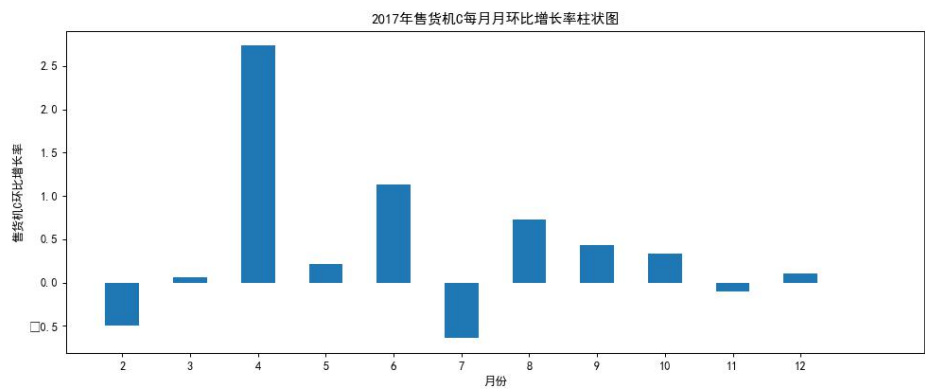
2、任务 2 数据分析与可视化

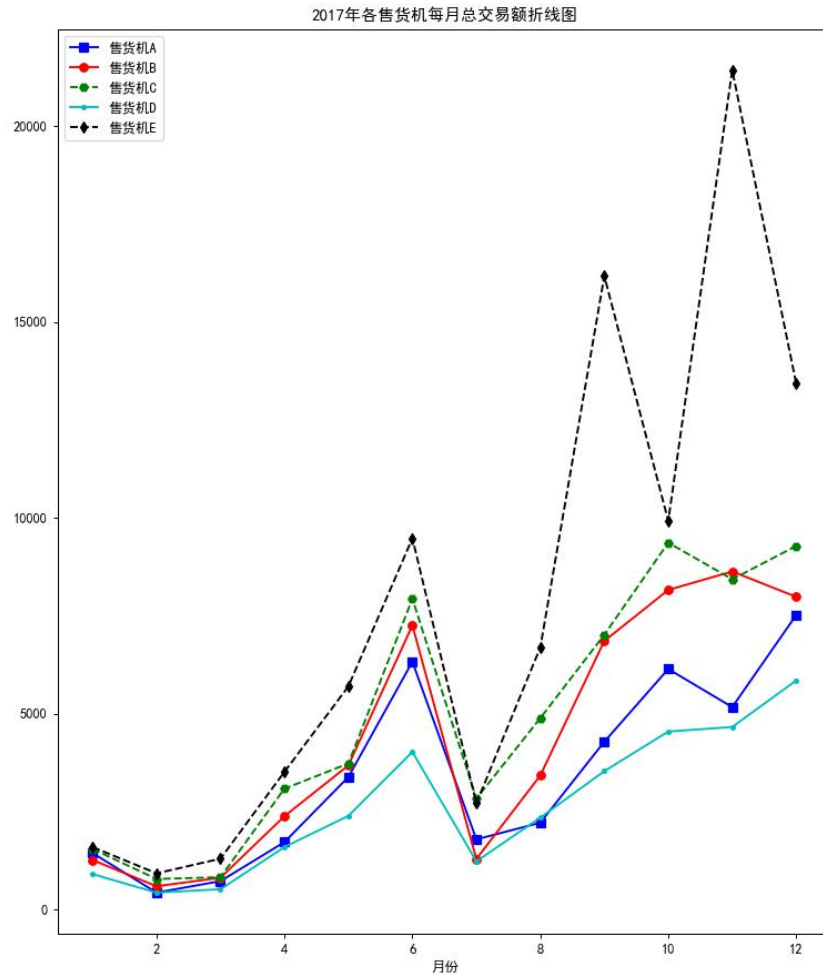
2.1 绘制 2017 年 6 月销量前五的商品销量柱状图。解决方案是：读取附件 1 中的数据，提取 2017 年 6 月的商品销售数据，用 count 函数提取出排名前 5 的商品，从而画出 2017 年 6 月销量前五的商品销量柱状图。结果如下：



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。解决方案：交易额月环比= (本期交易额-上期交易额)÷上期交易额，利用 for 循环将附件 1 的数据提取计算出每台售货机每月总交易额分别是多少，得到每台售货机每月总交易额折线图。通过交易额月环比计算公式得到每台售货机每月交易额月环比增长率柱状图。结果如下：







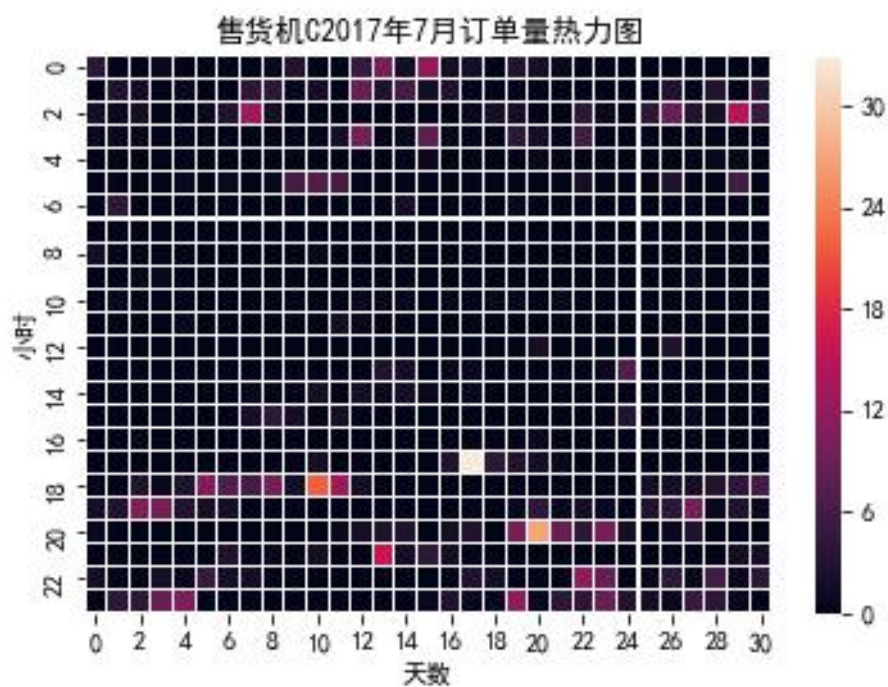
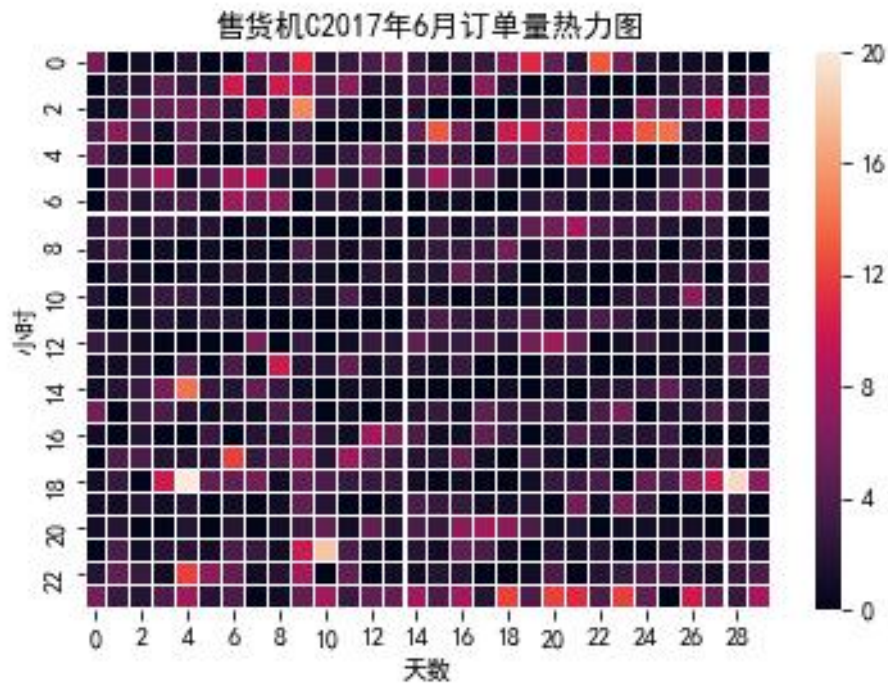
可以看出各售货机的月环比增长率柱状图大同小异，各售货机每月总交易额总体都呈现上升趋势，其中各售货机每月总交易额在 2017 年中基本呈现售货机 E>售货机 C>售货机 B>售货机 A>售货机 D 的规律。

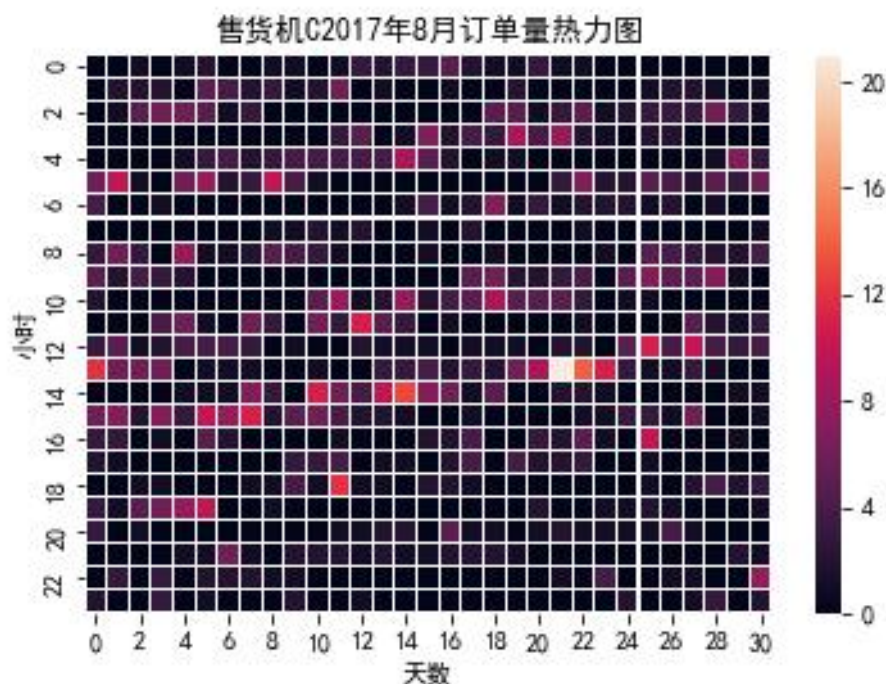
2.3 绘制每台售货机毛利润占总毛利润比例的饼图。（假设饮料毛利润为 25%，非饮料毛利润为 20%）解决方案是：首先读取附件 1 和附件 2 中的数据，可以看出附件 2 中大类分为

“饮料类”和“非饮料类”，通过 for 循环将附件 1 中的数据提取出特征为“饮料类/非饮料类”和“地点”分别为 A、B、C、D、E 的数据，将对应的附件 1 中的数据中的“实际金额”列中的数据乘以 0.25（饮料类）或 0.20（非饮料类），对应加总得到各售货机的毛利润，进而求出总毛利润，各售货机的毛利润除以总毛利润得到各售货机的毛利润占总毛利润比例，绘制出每台售货机毛利润占总毛利润比例的饼图。结果如下：

售货机	比例 (%)
售货机A	14.6%
售货机B	18.8%
售货机C	21.5%
售货机D	11.6%
售货机E	33.6%

2.5 绘制售货机 C 的 6、7、8 三个月的热力图，横轴以天为单位，纵轴以小时为 单位，从热力图可以分析得到什么结论？解决方案是：读取任务 1.1 中 task1-1C.csv 文件，通过 pandas.to_datetime 和 pandas.date_range 函数提取计算出以每小时为单位的售货机 C2017 年 6、7、8 月的订单量数据。再通过 sns.heatmap 函数画出货机 C 的 6、7、8 三个月的热力图。结果如下：

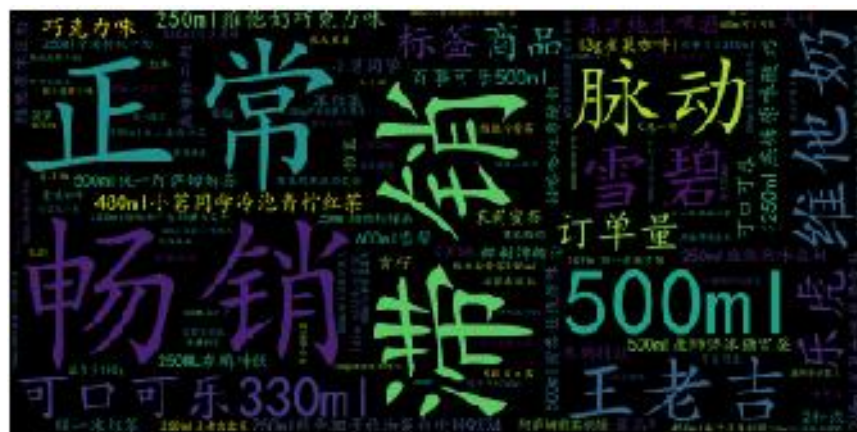


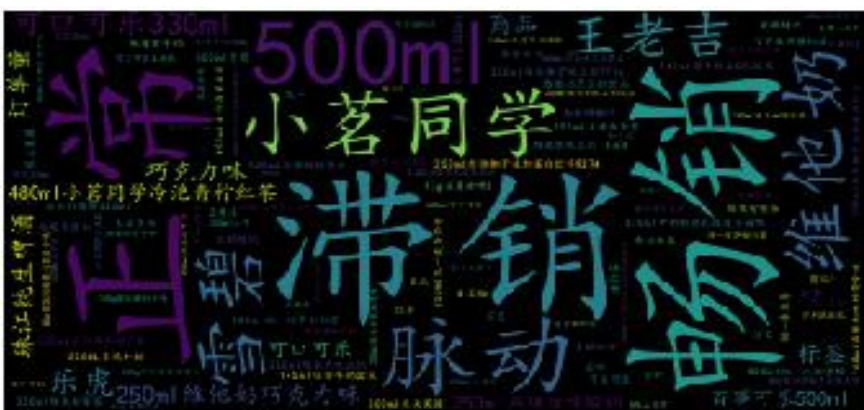
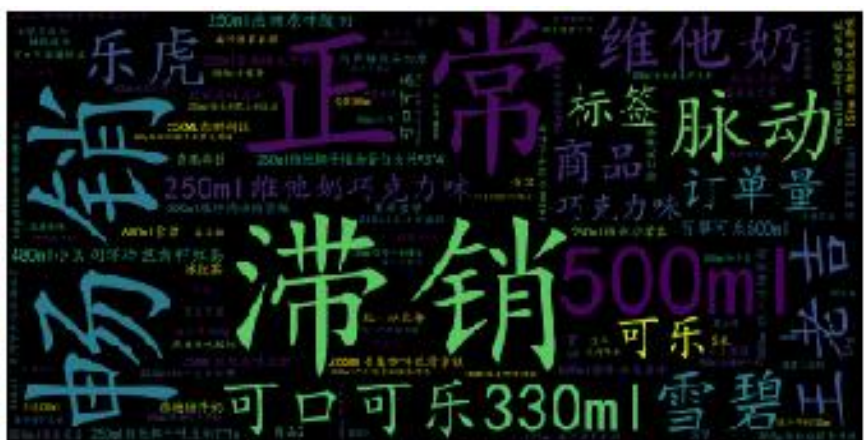


3、任务3 生成自动售货机画像

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， 结果保存到 csv 文件中，文件名分别为“task3_1A.csv”，...，“task3_1E.csv”。解决方案是：读取附件 1 和附件 2 中数据，根据附件 2 中饮料类商品名和附件 1 中“地名”来筛选出附件 1 中各售货机饮料类商品的销售数据，根据订单量的多少，分别取各售货机饮料类商品订单量按从少到多排序的三分位数为“滞销”、“正常”、“畅销”的临界值。再将各售货机饮料类商品贴上相应的标签，分别保存为“task3_1A.csv”，⋯，“task3_1E.csv”。

3.2 在任务 3.1 的基础上进行标签拓展,依据标签生成完整的售货机画像,结果 保存到 csv 文件中,文件名分别为“task3_2A.csv”, ..., “task3_2E.csv”。解决方案是利用词云图将 3.1 中得到的商品标签生成完整的售货机画像。结果如下:





4、任务 4 业务预测

4.1 预测未来销售额的原理是什么？是否可以根据附件的数据对每台售货机的每个大类商品在 2018 年 1 月的交易额进行预测，请阐述理由。

答：预测未来销售额的原理是使用 `sklearn` 估计器构建非线性回归模型。首先将附件 1 中的数据通过附件 2 中的分类数据的特征分别提取出 2017 年每月各售货机饮料类商品和 2017 年每月各售货机非饮料类商品，从而计算出 2017 年每月各售货机饮料类商品销售额和 2017 年每月各售货机非饮料类商品销售额，画出相应的时序图，通过时序图可以看出 2017 年每月各售货机饮料类商品销售额和非饮料类商品销售额与月份的关系均可近似看做为一元二次函数关系。然后再使用 `sklearn` 估计器构建非线性回归模型绘制拟合预测的曲线图并与 2017 年各售货机饮料类（非饮料类）实际销售额与月份的折线图相比较，发现两者几乎吻合。因此，将该方法作为预测每台售货机的每个大类商品在 2018 年 1 月的交易额的理论依据。

4.2 如果你们认为可以预测，请给出过程及结果；如果认为附件所给数据不足以支持预测工作，应该向经营者提出哪些数据方面的要求？

通过 4.1 的论述，基本可以认为可以根据附件的数据对每台售货机的每个大类商品在 2018 年 1 月的交易额进行预测。在执行 `task4.txt` 中的代码后的结果如下：

2018 年 1 月的预测交易额

	售货机 A	售货机 B	售货机 C	售货机 D
非饮料	20289	21778	25027	13118
饮料	26706	39693	45159	26781