

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，为用户提供精确贴心的服务，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，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，使其能够很好地展现销售商品的特征。

二、数据说明

1.数据来源

数据来源：某五台自动售货机

数据时间：2017 年全年

2.数据采集

由后台数据给出，有着全年的销售数据，包括价格、种类等。

3.数据量

由附件 1、附件 2 给出，主要有两大类：饮料和非饮料。

三、任务分析

任务一：数据预处理。

首先观察所给附件的数据，发现有一行错误数据，其时间显示有问题，是 C 售货机上的销售数据，这里选择把它删除。

任务 1.1 从附件 1 中所有售货机的数据中，提取每台售货机的销售数据，分别保存到“task1-1A”、“task1-1B”、“task1-1C”、“task1-1D”、“task1-1E”中。

利用 python 的 pandas 模块中 loc 函数提取每台售货机的销售数据，将结果保存即可。
(这里保存的时候考虑到有中文的关系，编码方式 encoding 选择了 utf_8_sig)

任务 1.2 计算每台售货机 2017 年 5 月的交易额、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。

主要步骤为先读取各个售货机的销售数据，提取五月份的数据，再进行计算即可。（订单量即为提取出的数据的长度（行数），交易额利用 numpy 中的 sum 函数计算。）

源程序将会以附件的形式上传，这里给出表格。

	A	B	C	D	E	总和
五月份交易额	3385.1	3681.2	3729.4	2392.1	5699.0	18886.8
五月份订单量	756	869	789	564	1292	4270

任务 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，以表格形式体现在报告中。

主要步骤与任务 1.2 相似，重复操作，画成表格即可。

日均订单量

	A	B	C	D	E
一月	10.2	10.7	11.3	7.7	10.9
二月	4.0	6.5	7.3	5.0	9.1
三月	7.0	7.3	7.1	4.9	9.7
四月	14.2	19.3	23.5	14.0	28.1
五月	24.4	28.0	25.5	18.1	41.7
六月	51.6	59.3	58.2	33.3	82.8
七月	14.0	9.5	23.0	9.4	22.4
八月	21.3	30.6	40.3	22.7	56.7
九月	33.0	54.6	53.0	30.8	130.6
十月	49.3	64.0	70.8	37.7	87.7
十一月	38.5	67.3	64.5	40.2	166.6
十二月	64.1	70.4	75.8	52.6	103.5

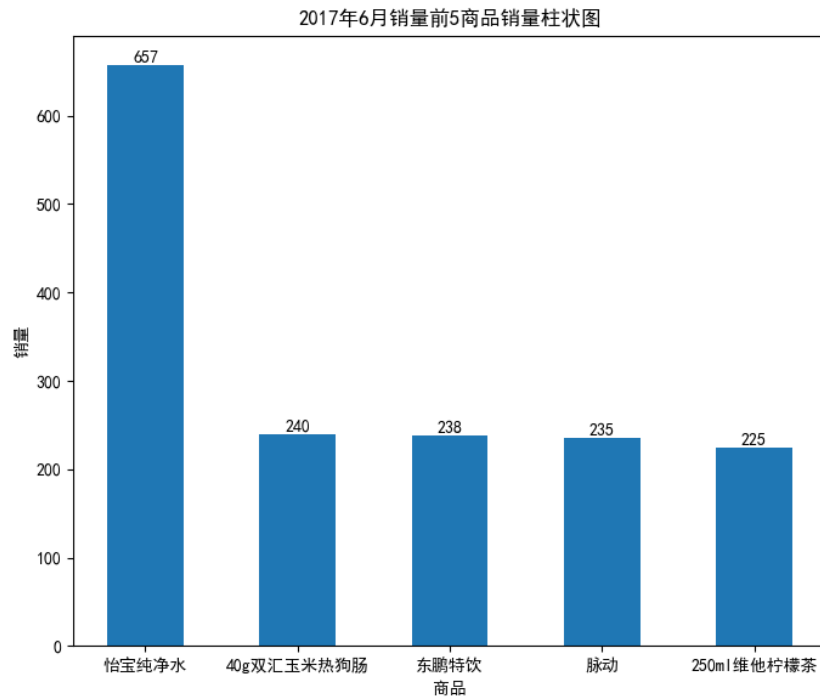
每单平均交易额

	A	B	C	D	E
一月	4.6	3.8	4.4	3.8	4.7
二月	3.9	3.3	3.8	3.1	3.6
三月	3.4	3.6	3.8	3.4	4.3
四月	4.0	4.1	4.4	3.8	4.2
五月	4.5	4.2	4.7	4.2	4.4
六月	4.1	4.1	4.5	4.0	3.8
七月	4.1	4.4	4.0	4.2	3.9
八月	3.4	3.6	3.9	3.3	3.8
九月	4.3	4.2	4.4	3.8	4.1
十月	4.0	4.1	4.3	3.9	3.6
十一月	4.5	4.3	4.3	3.9	4.3
十二月	3.8	3.7	4.0	3.6	4.2

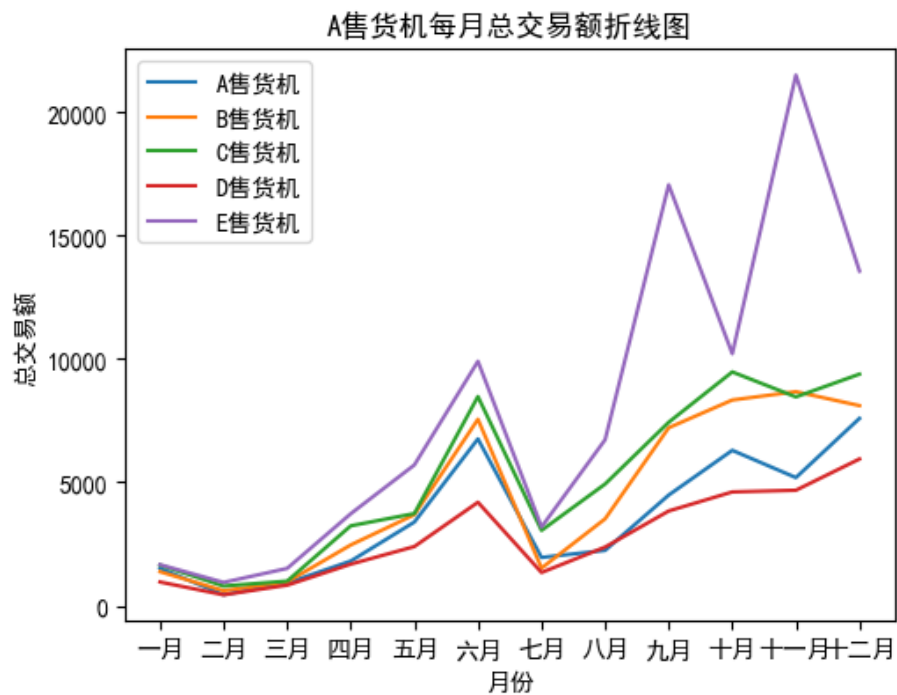
任务二：数据可视化

数据可视化是数据分析必不可少的一环，利用 python 中的 matplotlib 模块可以绘制图像，这里分别绘制了几个任务的图像。源程序将会以附件形式上传。

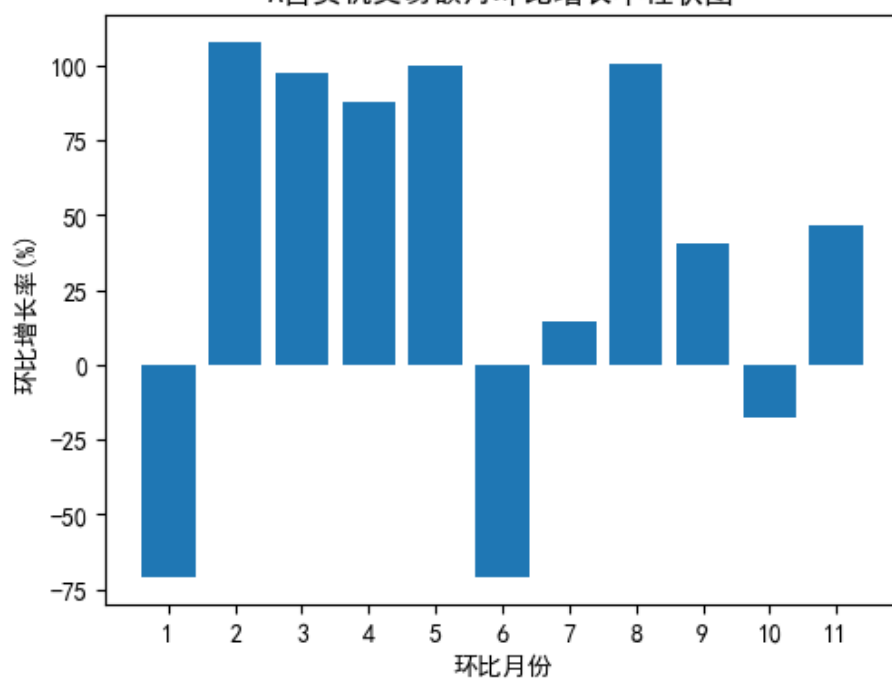
任务 2.1 绘制 2017 年 6 月销量前 5 的商品销量柱状图



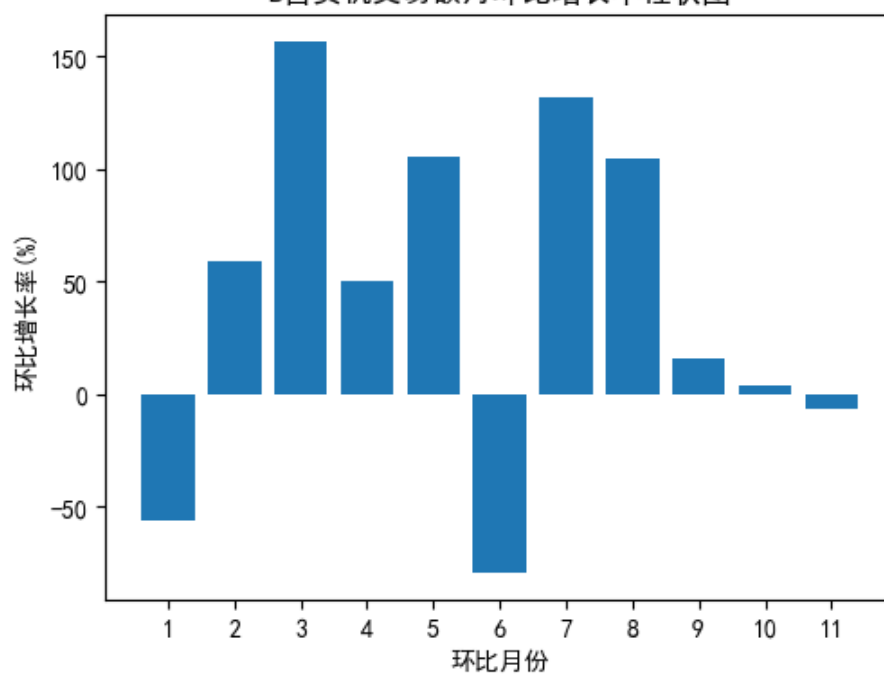
任务 2.2 绘制每台售货机每月总交易额折线图及交易额月环比增长率柱状图



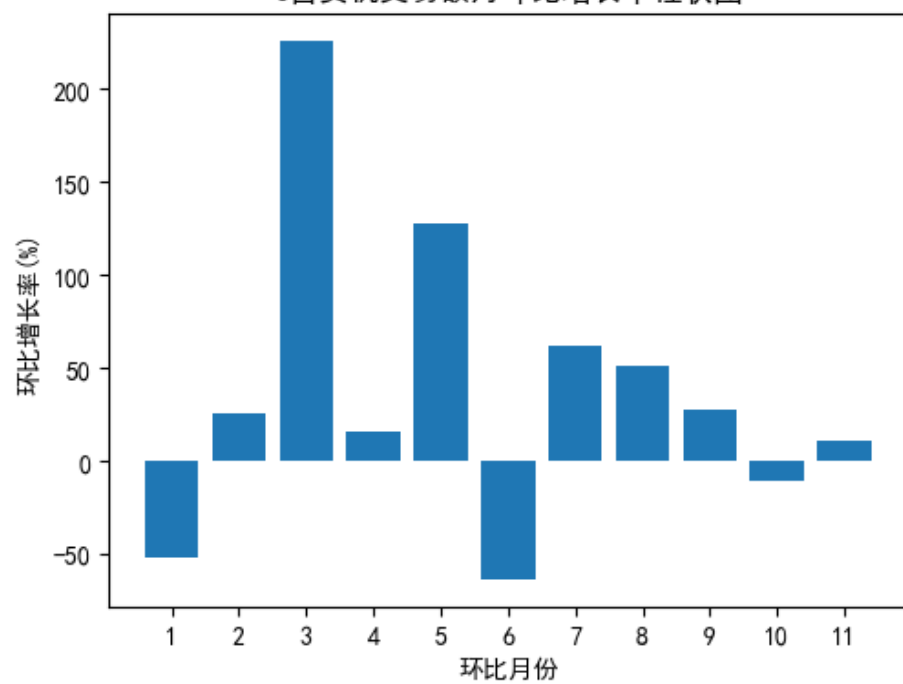
A售货机交易额月环比增长率柱状图



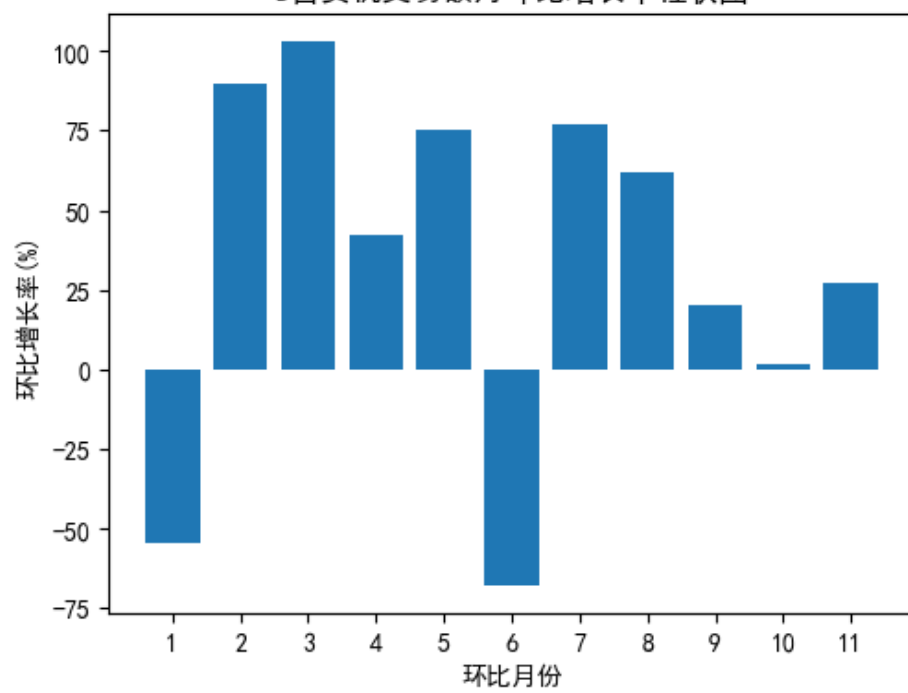
B售货机交易额月环比增长率柱状图

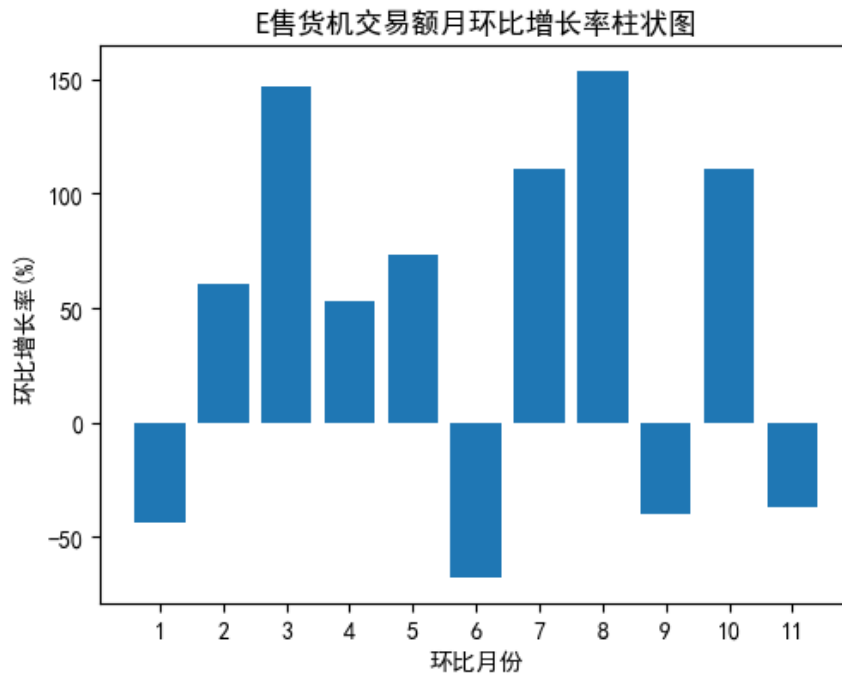


C售货机交易额月环比增长率柱状图



D售货机交易额月环比增长率柱状图

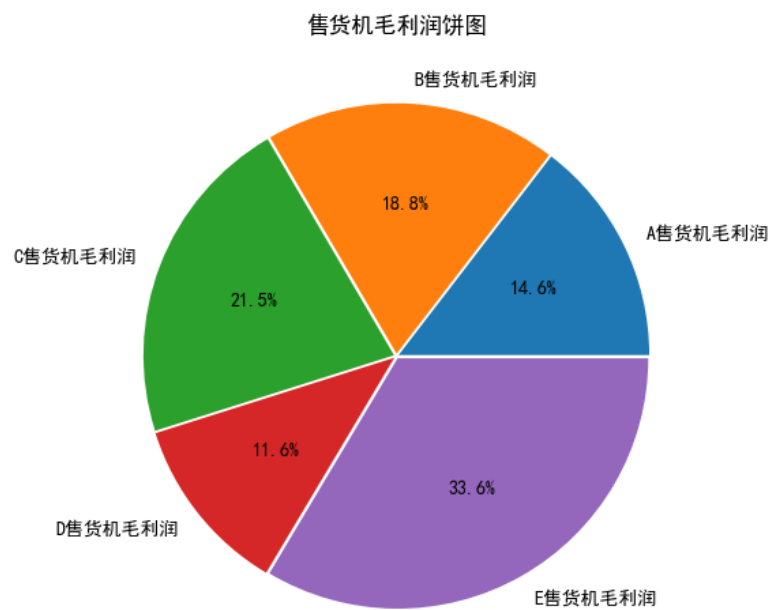




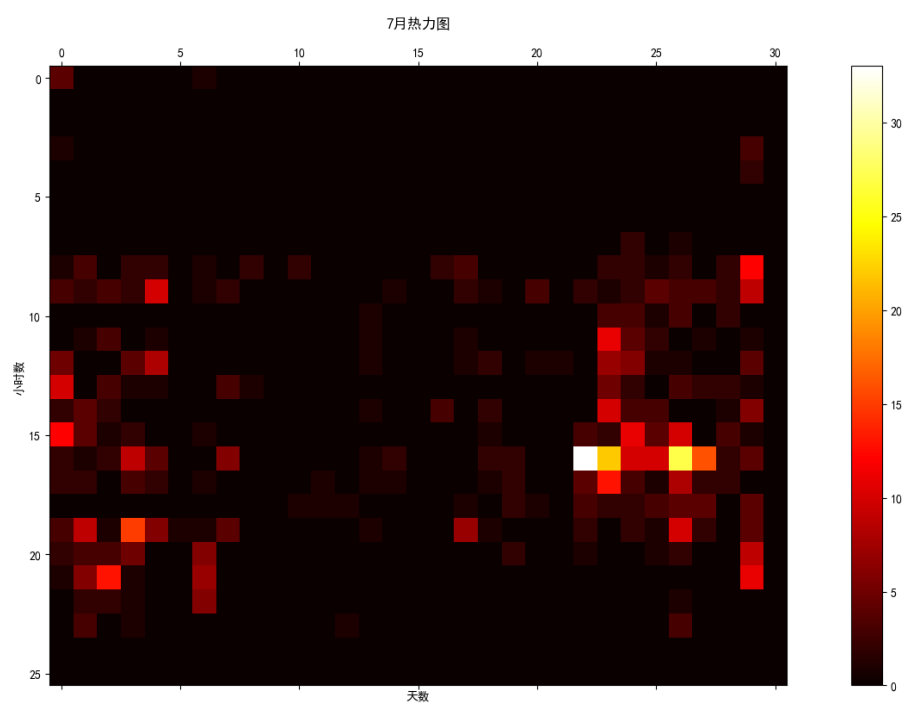
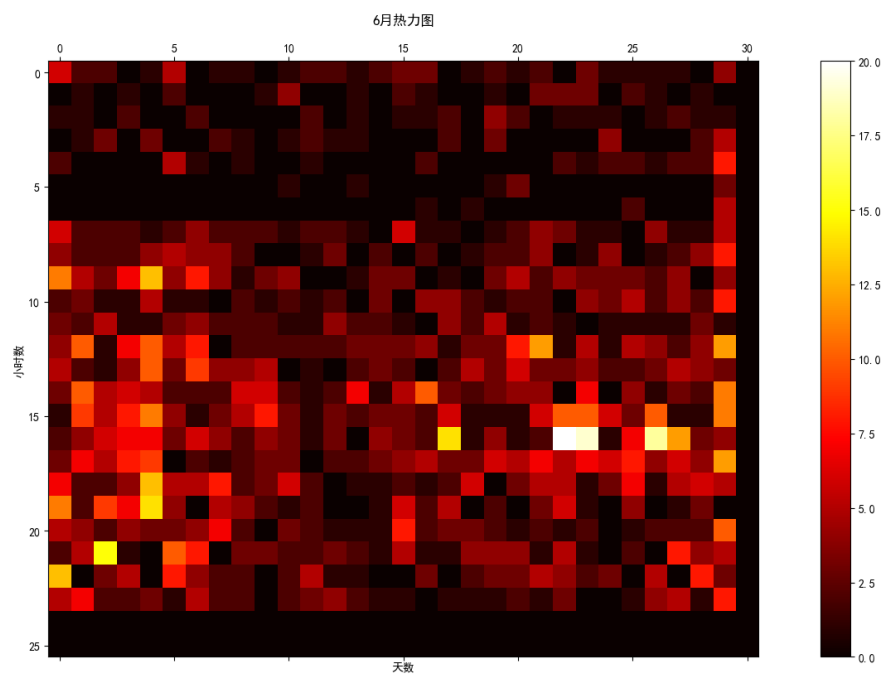
任务 2.3 绘制每台售货机毛利润占总毛利润比例的饼图（假设饮料类毛利率为 25%，非饮料类为 20%）。

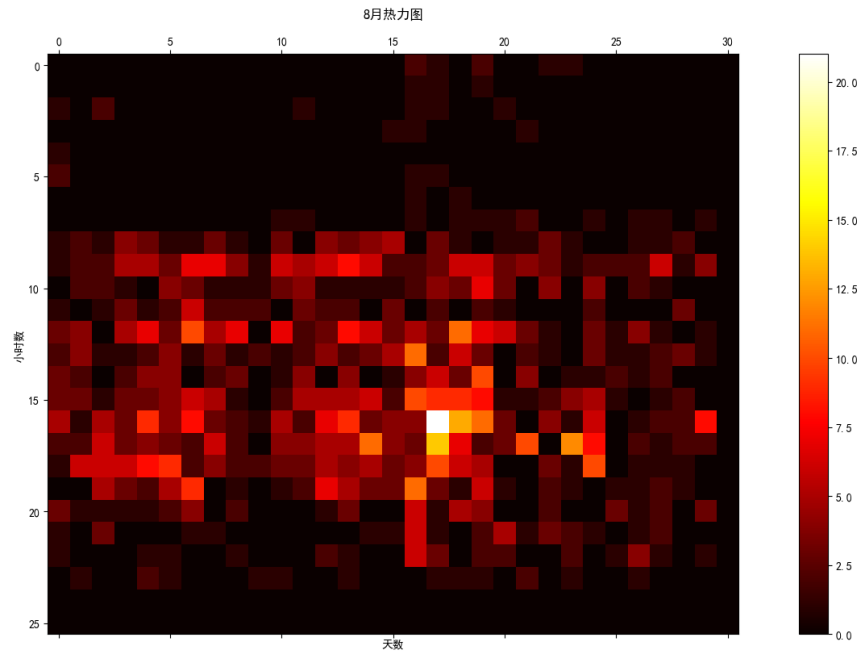
此任务要先把附件一二合并，再利用 matplotlib 模块画图。

合并后保存文件名为“合并数据”的 excel 文件。



任务 2.5 绘制售货机 C 6、7、8 三个月订单量的热力图，横轴以天为单位，纵轴以小时为单位。





分析：总体来看，9 时到 20 时是销售高峰期。

分别来看，6 月和 8 月是销售高峰期，7 月份订单量明显少于 6 与、8 月。

而 6、7 月份订单集中在月头和月尾，8 月份订单则较为集中在月头和月中旬。

任务三：为售货机贴标签

任务 3.1 分析各售货机商品的销售数据，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。

根据网上查阅的资料，商家判断商品畅销与否采用的一般为“二八原则”，即所有商品先根据销量排序，排队中前 20%的商品可以定义为“畅销”，在剩余的商品在再来定义“正常”与“滞销”。根据任务要求，我们可以做出如下改进：首先，销量为 0 的商品毫无疑问地直接将他定义为滞销。我们在剔除了销量为 0 的商品后，在有销售数据的商品中，前 20%的商品定义为“畅销”，前 20%-40%的商品定义为“正常”，剩余的商品定义为“滞销”。

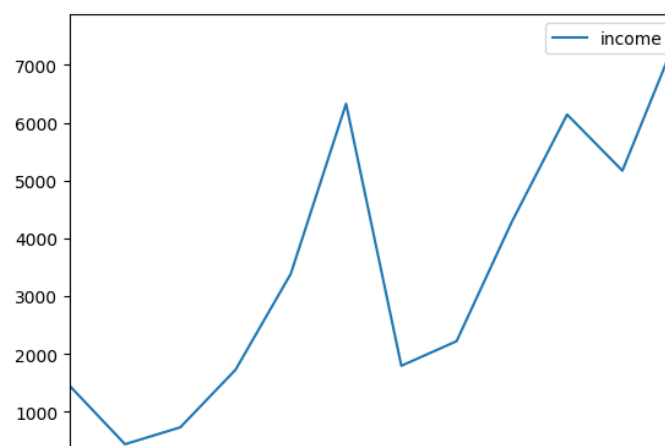
部分结果展示如图，完整结果在附件当中。

商品	标签
100g*5瓶益力多	滞销
145ml旺仔牛奶罐装	滞销
145ml旺仔牛奶盒装	畅销
150g健能酸奶原味	滞销
250ML东鹏特饮	滞销
250ml红牛	正常
250ml维他奶巧克力味	正常
250ml维他柠檬茶	畅销
250ml维他原味豆奶	畅销

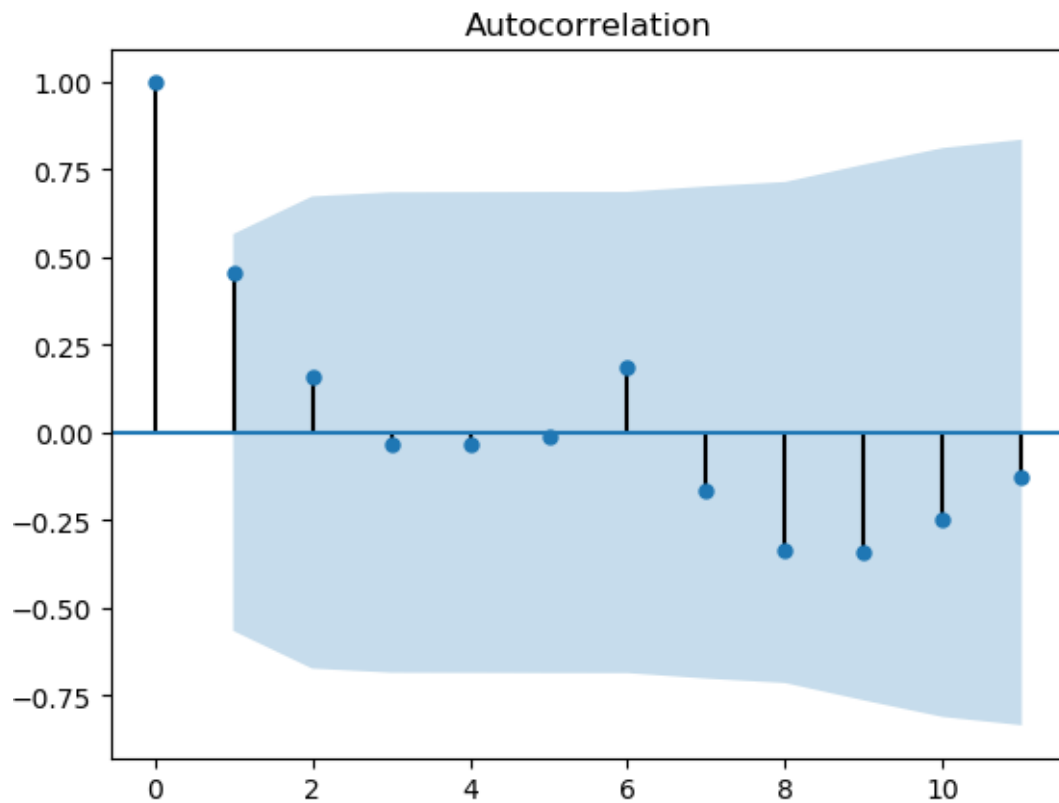
任务四：销量预测

本次销量预测利用的是时间序列预测方法，根据时间的移动和数据的变化来推测未来的数据大小，此报告以 A 售货机为例子预测下一年一月份的交易额。

首先得出 A 售货机 12 个月份的交易额，如图：

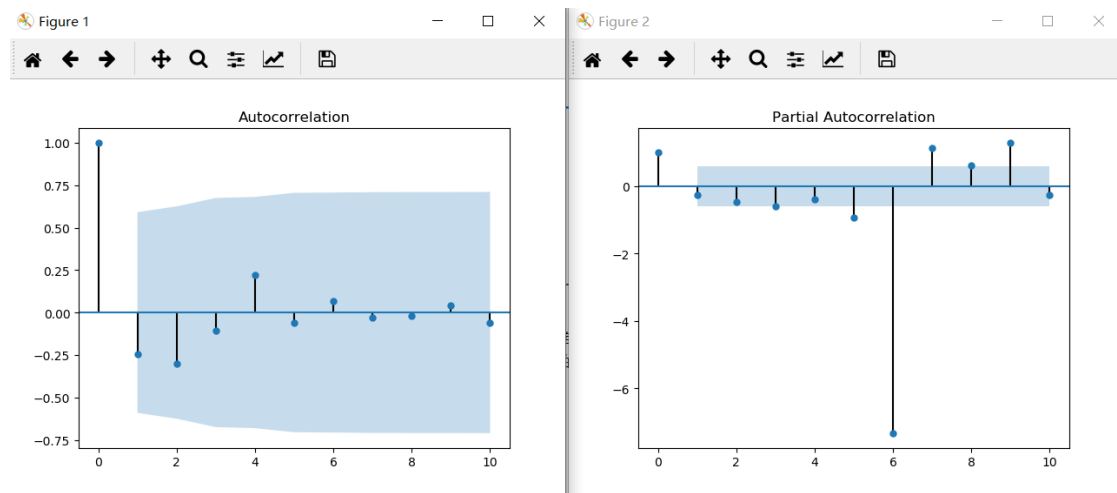


画出自相关系数图，



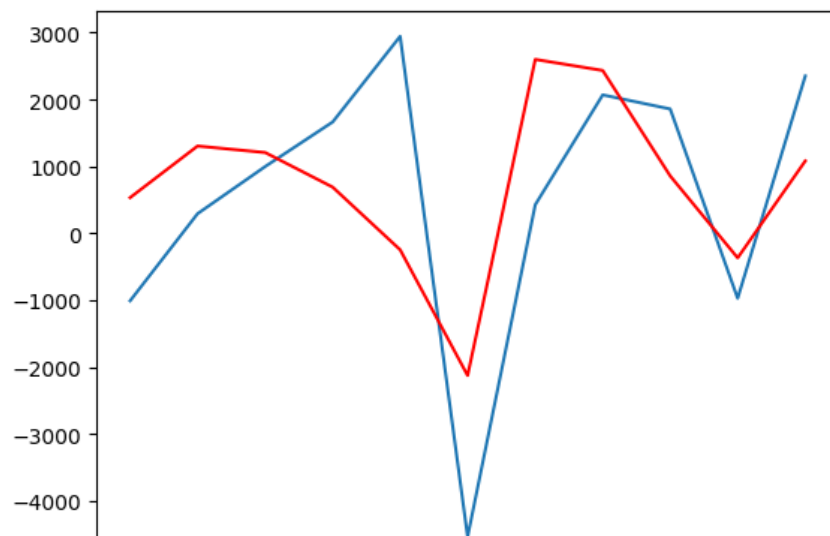
可以看出每月数据之间有较为明显的相关关系。

利用差分的方法使数据变得平稳，这里采用默认一阶差分，差分之后的自相关系数与偏相关系数如图，



模型的效果一般，这里采用 ARIMA (0, 1, 1) 模型进行时间序列分析

训练模型后结果如图：



拟合结果还可以，最后检验显著性水平

	AC	Q	Prob(>Q)
lag			
1.0	0.034926	0.017443	0.894927
2.0	-0.375298	2.255369	0.323782
3.0	-0.333079	4.238456	0.236840
4.0	-0.030476	4.257429	0.372286
5.0	0.050937	4.319266	0.504423
6.0	0.258490	6.230235	0.397900
7.0	0.033988	6.271533	0.508426
8.0	-0.046541	6.374783	0.605328
9.0	-0.022106	6.409724	0.698314
10.0	-0.070839	7.127326	0.713372

可以看出模型效果不错

接着预测下一月份的交易额，结果为 6782.18。

对每台售货机使用时间序列方法预测到的结果如下：

A: 6782.175128

B: 9253.290620

C: 10330.085169

D: 5498.155761

E: 21895.980443