

新零售无人机智能售货机 商务数据分析项目分析报 告

目录

一，项目背景.....	3
二，项目目的.....	3
三，项目目标.....	3
四，项目分析过程.....	4
任务一：数据预处理和分析.....	4
任务二：数据分析与可视化.....	5
任务三：生成各自动售货机画像.....	10
任务四：销售额的预测.....	12

一，项目背景

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经营需要重点关注的问题。因此，科学的商务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，是掌握经营反方向的重要手段，对自动售货机这一营销模式的发展有着非常重要的意义。

二，项目目的

通过新零售无人智能售货机商务数据分析项目，掌握 python 中 NumPy 数值计算、Matplotlib 的基础语法，即各种图表的绘制，以及 Pandas 的统计分析，使用 Pandas 进行数据预处理的基础知识。

三，项目目标

- 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 月商品销量的预测问题。
- 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。

四，项目分析过程

任务一：数据预处理和分析

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”。

执行提取每台售货机对应的销售数据源代码时，发现出现“day is out of range for month”，处理发现异常数据：

订单号	商品	时间
DD201708167493241554692026752	55g 奥利奥原味芝士饼干	2017/2/29 3:44:00 PM

直接删除该数据，得到文件“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	5 月交易额	5 月订单量
A	3385.1	756
B	3681.2	869
C	3729.4	789
D	2392.1	564
E	5699.0	1292
合计	18886.8	4270

1.3，计算每台售货机每月的每单交易额与日均订单量：

售货机	每月每单交易额											
	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
A	4.57	3.87	3.37	4.03	4.48	4.09	4.12	3.36	4.32	4.02	4.48	3.78
B	3.8	3.27	3.6	4.12	4.24	4.08	4.4	3.6	4.18	4.11	4.27	3.66
C	4.43	3.82	3.76	4.37	4.73	4.54	3.98	3.92	4.41	4.27	4.36	3.95
D	3.8	3.09	3.44	3.79	4.24	4.03	4.23	3.33	3.83	3.89	3.87	3.59
E	4.72	3.64	4.3	4.17	4.41	3.81	3.91	3.81	4.13	3.65	4.29	4.19
	日均订单量											
	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
A	10	4	7	14	24	52	14	21	33	49	38	64
B	11	6	7	19	28	59	9	31	55	64	67	70
C	11	7	7	24	25	58	23	40	53	71	65	76
D	8	5	5	14	18	33	9	23	31	38	40	53
E	11	9	10	28	42	83	22	57	131	88	167	103

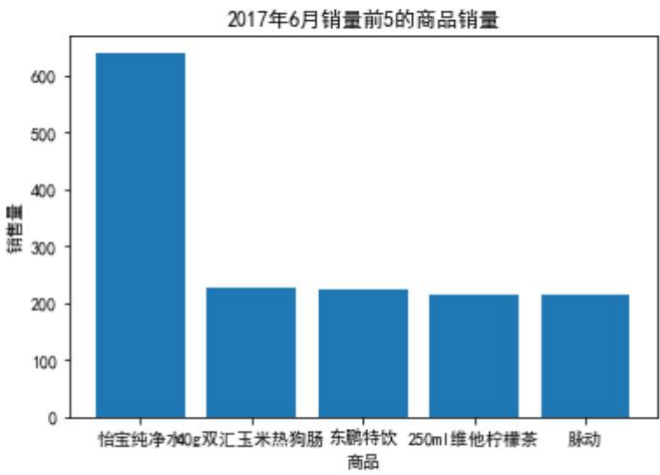
任务二：数据分析与可视化

2.1，绘制 2017 年 6 月销售量前 5 的商品销售柱状图

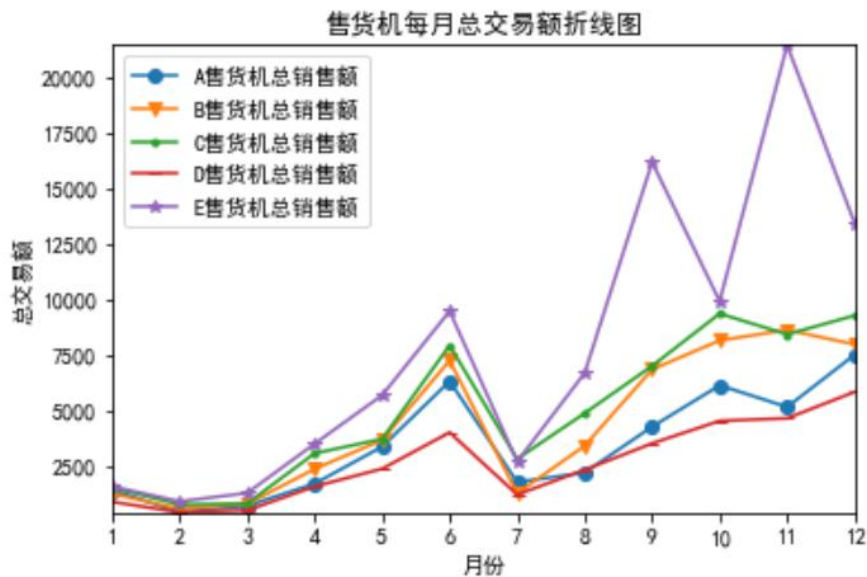
取前五的商品销售数据：

怡宝纯净水	40g 双汇玉米热狗肠	东鹏特饮	250ml 维他柠檬茶	脉动
638	228	223	215	215

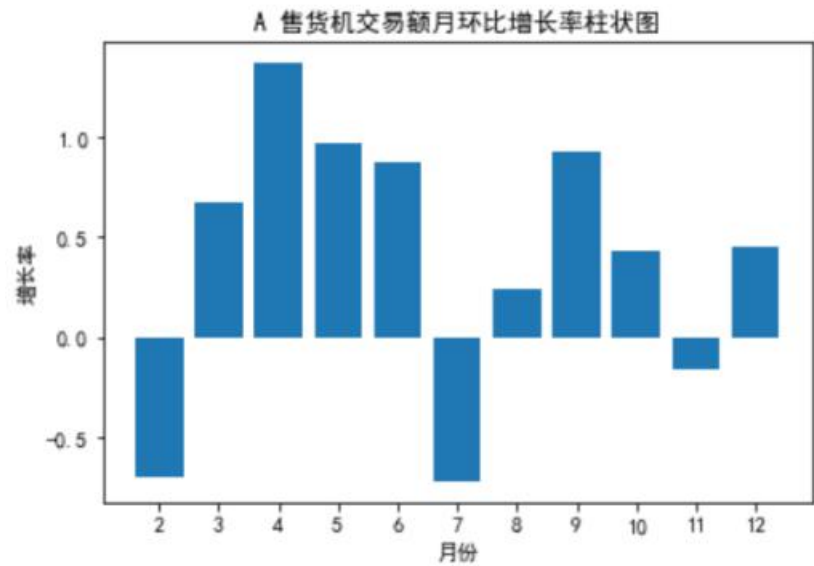
制得柱状图：



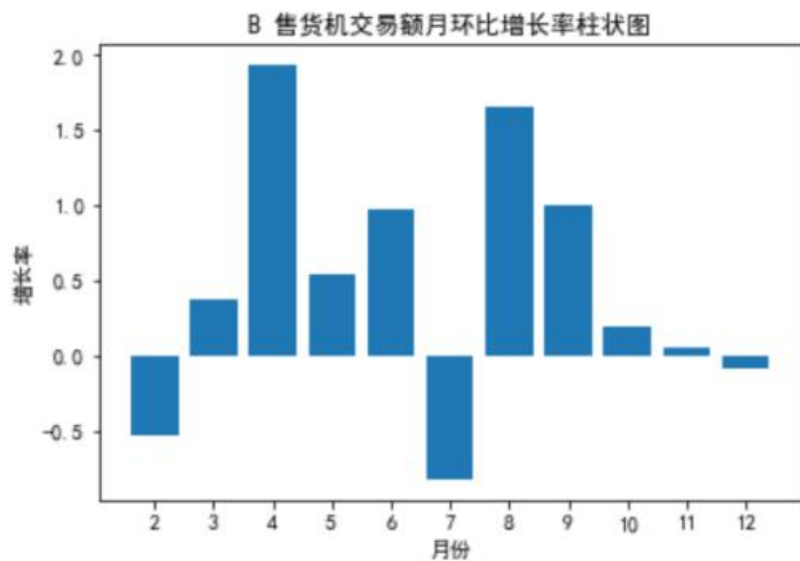
2.2，绘制每台售货机每月交易额折线图及交易额月环比增长率柱状图：



A 售货机 2017 年 1-12 月交易总额：
[1443.7, 433.5, 725.6, 1722.4, 3385.1, 6327.7, 1791.6, 2217.4, 4284.7, 6141.5, 5169.4, 7519.6]

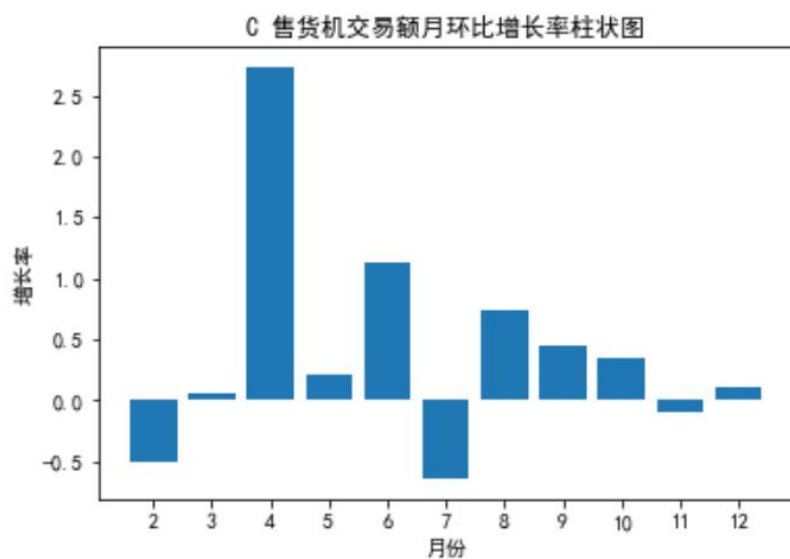


B 售货机 2017 年 1-12 月交易总额：
[1264.6, 594.8, 812.5, 2383.9, 3681.2, 7256.4, 1289.8, 3423.6, 6856.8, 8160.4, 8632.4, 7983.8]



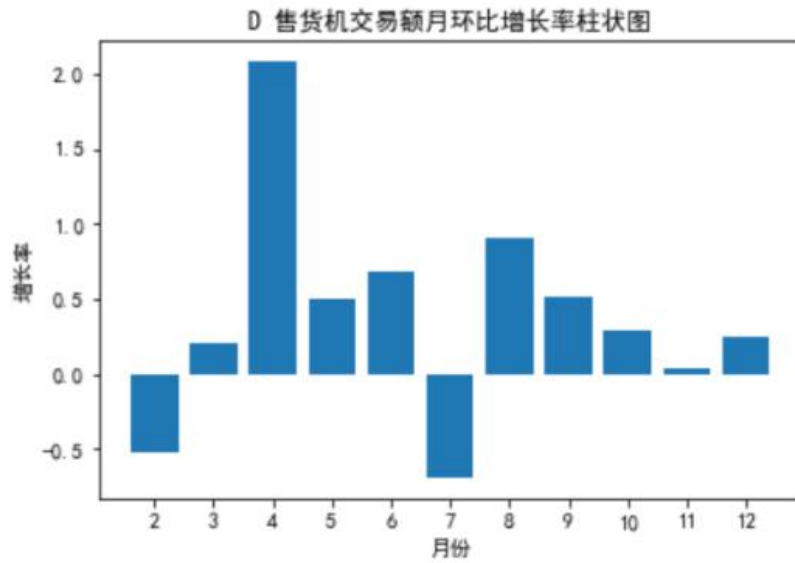
C 售货机 2017 年 1-12 月交易总额:

[1547.5, 780.0, 826.4, 3086.2, 3729.4, 7937.5, 2833.4, 4887.7, 7014.9, 9367.0, 8432.7, 9283.0]

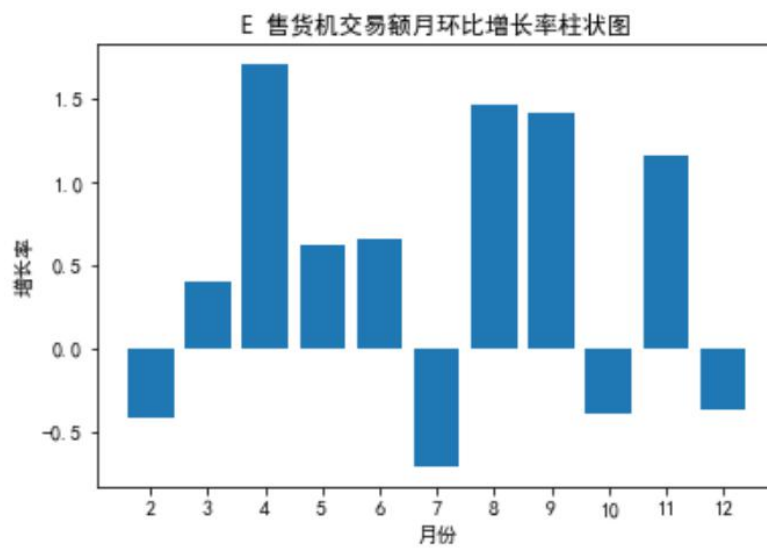


D 售货机 2017 年 1-12 月交易总额:

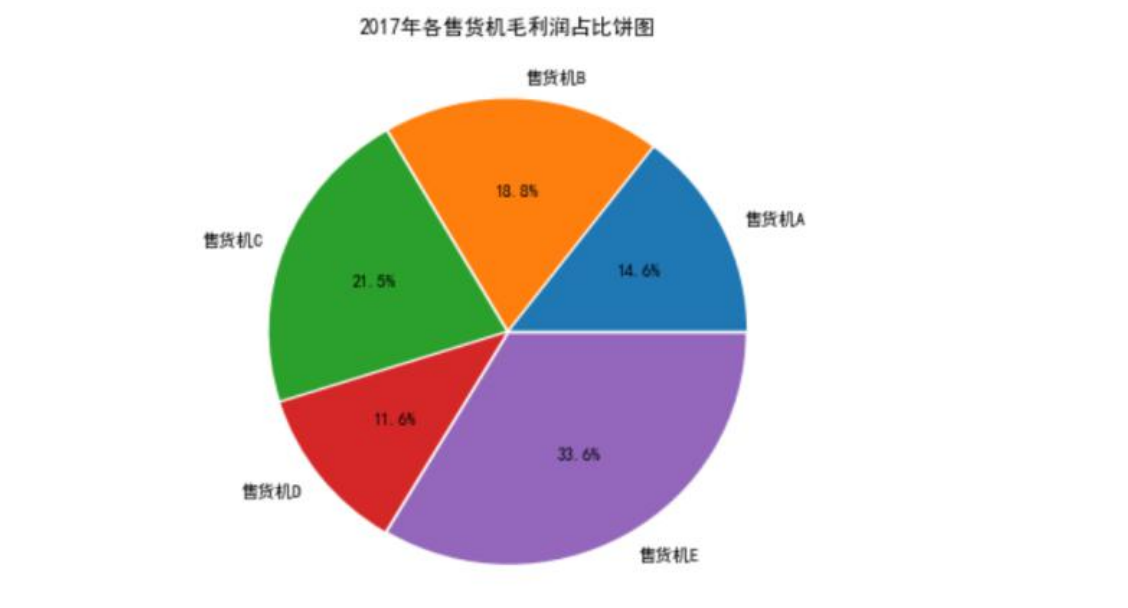
[905.4, 432.5, 519.0, 1596.1, 2392.1, 4025.0, 1230.7, 2346.3, 3533.5, 4544.9, 4661.4, 5848.7]



E 售货机 2017 年 1-12 月交易总额：
 [1596.3, 924.7, 1298.7, 3515.4, 5699.0, 9478.2, 2711.5, 6695.0, 16174.4, 9923.1,
 21433.8,
 13429.2]



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%）



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：将附件 1 和附件 2 主键合并成新附件“附件 1 合并附件 2”，通过该附件提取各个二级类的每月交易额均值得附件 3：

二级类	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
饼干糕点	27	4	6	40	79	124	28	28	110	101	133	94
茶饮料	27	24	26	65	63	154	52	137	194	207	256	249
方便速食	15	4	7	27	65	95	25	25	80	75	118	87
功能饮料	20	23	28	50	53	111	33	118	157	140	210	223
果冻/龟苓膏	0	0	0	1	1	0	1	0	4	1	2	2
果蔬饮料	7	6	4	13	18	28	9	27	51	41	65	46
海味零食	2	1	1	4	7	12	2	3	9	12	16	12
坚果炒货	2	0	0	1	3	5	2	0	3	3	5	3
咖啡	5	1	1	9	20	16	7	5	33	21	42	29
蜜饯/果干	7	6	5	13	17	31	12	29	39	43	55	66
膨化食品	0	0	0	0	0	0	0	0	0	0	0	0
其他	0	0	0	0	0	0	1	1	5	2	7	1
肉干/豆制品/蛋	24	8	11	46	82	140	40	45	113	143	133	124
乳制品	37	12	16	59	86	175	46	70	188	175	242	173
水	9	7	6	18	12	58	11	50	70	66	69	65
碳酸饮料	11	11	12	26	26	65	18	62	98	80	119	113
糖果/巧克力	2	0	1	3	4	8	3	1	5	9	9	4
香烟	5	3	0	2	1	8	1	9	15	8	22	29
植物蛋白	8	2	6	14	20	46	11	9	37	41	40	43
纸巾	1	0	0	1	2	3	1	1	3	3	4	3

由于保留的是整数部分，部分数据因为数值太小所以得到的数值为 0，再通过附件 3，运行源代码将数据转换成气泡图。（源代码编写错误，没有得到气泡图）

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，并分析热力图。

任务三：生成各自动售货机画像

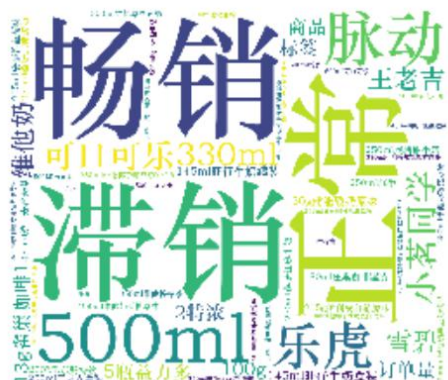
3.1，分析各售货机商品销售数据，总结规律，给出每台售货机饮料类商品的标签，并保存在 CSV 文件中，文件名分别为：

''task3-1A.csv''、''task3-1B.csv''、''task3-1C.csv''、''task3-1D.csv''、''task3-1E.csv''。

通过计算三分位点，判断 A、B、C、D、E 售货机的订单量达到“畅销”、“正常”、“滞销”中的哪个标准，并依据此标准进行商品标签张贴，保存在相应的文件中。

售货机	畅销	正常	滞销
A	大于 45.5	(7.5,45.5)	小于 7.5
B	大于 54	(7.5,54)	小于 7.5
C	大于 59.5	(11, 59.5)	小 11
D	大于 32.5	(5, 32.5)	小于
E	大于 132.5	(25.5, 132.5)	小于 25.5

售货机 E



由上五个图可以看出 A、C、D 售货机滞销的情况相对较小，而售货机 B、E 滞销的情况较大，可由此减少出现滞销的商品，增加画像中出现且标签较大的商品，从而减少商品的物质消耗，减小成本扩大收益。

任务四：销售额的预测

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。请你们完成如下的问题：

4.1，预测未来销售额的原理是什么？是否可以根据附件提供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，请阐述原因。

4.2，如果你们认为可以进行预测，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？

使用 Python 训练回归模型并进行预测。由上各个售货机销售额的折线图和月环比增长率柱状图可以大致看出，每台售货机的交易额并非遵循简单的一元线性回归模型。而现在需要对每个大类商品预测，则可以选 A、B 的售货情况为训练集，C、D、E 为测试集。