

一、任务重述

1 任务一：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”。
- 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。
- 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。

2 任务二：数据可视化

- 2.1 绘制 2017 年 6 月销量前五的商品销量柱状图。
- 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。
- 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%）
- 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。
- 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？

3 任务三：生成个自动售货机画像

- 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”
- 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”。

4 任务四：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。
- 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？

二、任务假设

1. 附件1，附件2无错误数据，且仅有附件1最后一行数据为脏数据并手动删除。
2. 订单量即为商品实际金额的个数。

三、任务的解决

3.1 任务一：数据预处理与分析

3.1.1 任务 1.1 的解决

读取附件 1 的内容，分别提取地点为 A, B, C, D, E 的数据，利用 ‘loc’ 提取指定行。保存为 ‘task1_1A’ , ..., ‘task1_1E’。

3.1.2 任务 1.2 的解决

提取每台售货机 5 月份的数据，‘sum’ 和 ‘count’ 计算实际金额的总量（即交易额）以及个数（即订单量），数据如下表所示：

2017 年 5 月份各售货机数据表						
售货机	A	B	C	D	E	总计
交易额	3385.1	3681.2	3729.4	2392.1	5699.0	18886.8
订单量	756	869	789	564	1292	4270

可以看出售货机 E 在 2017 年 5 月份相较于其他售货机交易额更大。

3.1.3 任务 1.3 的解决

利用 for 循环将售货机每个月的交易额和订单量统计出来，再计算每单平均交易额以及日均订单量。计算结果如下表所示：

A 售货机(结果保留两位小数)						
月份	1 月	2 月	3 月	4 月	5 月	6 月
平均交易额	4.51	3.86	3.59	4.04	4.48	4.05
日均订单量	10.81	4.07	8.23	14.90	24.39	55.63
月份	7 月	8 月	9 月	10 月	11 月	12 月
平均交易额	4.50	3.36	4.31	4.02	4.47	3.79
日均订单量	15.35	21.48	34.67	50.48	38.67	64.61

B 售货机(结果保留两位小数)						
月份	1 月	2 月	3 月	4 月	5 月	6 月
平均交易额	3.75	3.26	3.61	4.08	4.24	4.07
日均订单量	11.81	6.61	8.55	20.1	28.03	61.87
月份	7 月	8 月	9 月	10 月	11 月	12 月
平均交易额	4.40	3.58	4.13	4.11	4.27	3.67
日均订单量	11.13	31.65	58.17	65.35	67.70	71.29

C 售货机(结果保留两位小数)						
月份	1 月	2 月	3 月	4 月	5 月	6 月
平均交易额	4.33	3.83	3.77	4.40	4.73	4.50
日均订单量	12.23	7.39	8.48	24.47	25.45	62.73
月份	7 月	8 月	9 月	10 月	11 月	12 月
平均交易额	3.99	3.91	4.43	4.27	4.35	3.94
日均订单量	24.65	40.61	55.93	71.48	64.77	76.74

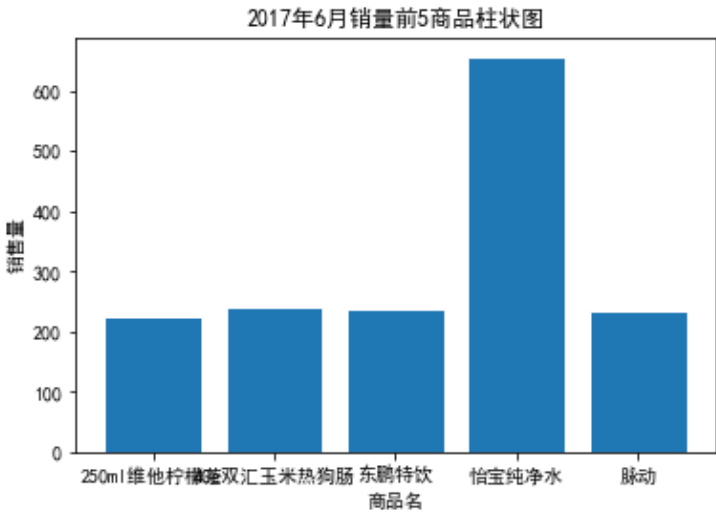
D 售货机 (结果保留两位小数)						
月份	1 月	2 月	3 月	4 月	5 月	6 月
平均交易额	3.69	3.09	4.31	3.79	4.24	4.03
日均订单量	8.35	5.04	6.19	14.77	18.19	34.67
月份	7 月	8 月	9 月	10 月	11 月	12 月
平均交易额	4.23	3.32	3.90	3.88	3.86	3.57
日均订单量	10.23	23.06	32.77	38.26	40.33	53.65

E 售货机 (结果保留两位小数)						
月份	1 月	2 月	3 月	4 月	5 月	6 月
平均交易额	4.68	3.64	4.31	4.16	4.41	3.82
日均订单量	11.42	9.21	11.29	29.83	41.68	86.43
月份	7 月	8 月	9 月	10 月	11 月	12 月
平均交易额	3.92	3.80	4.13	3.68	4.28	4.17
日均订单量	26.23	57.00	137.80	89.58	167.33	104.90

3.2 任务二：数据分析与可视化

3.2.1 任务 2.1 的解决

读取 2017 年 6 月份的数据，将销量前 5 即订单量前 5 表示出来，绘制散点图。散点图绘制如下：



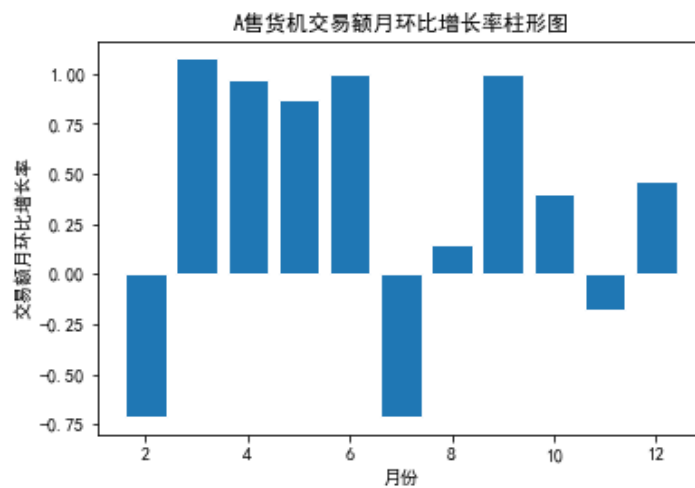
可以看出怡宝纯净水销售量最高，其余四个商品销售量基本相同。

3.2.2 任务 2.2 的解决

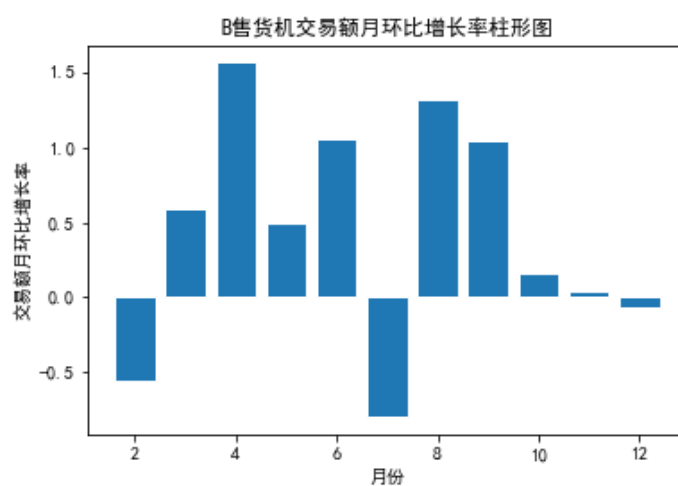
方法类似于任务1.3，利用for循环将售货机每个月的交易额统计出来，交易额月环比则是利用如下公式计算：

$$\text{交易额月环比} = (\text{本期交易额} - \text{上期交易额}) \div \text{上期交易额}$$

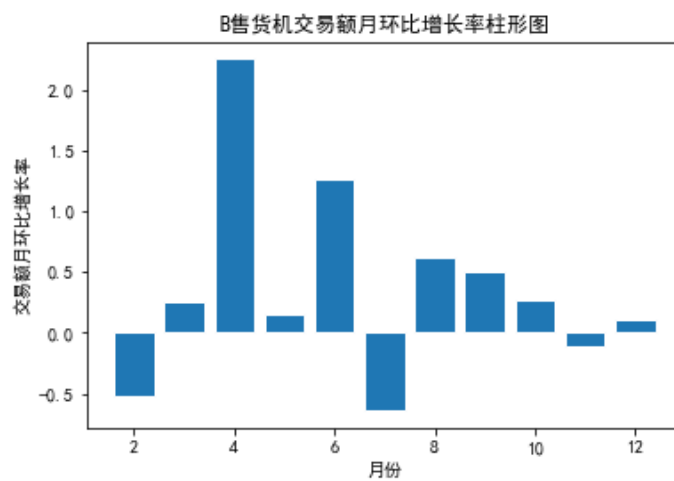
绘制A售货机交易额月环比柱形图如下：



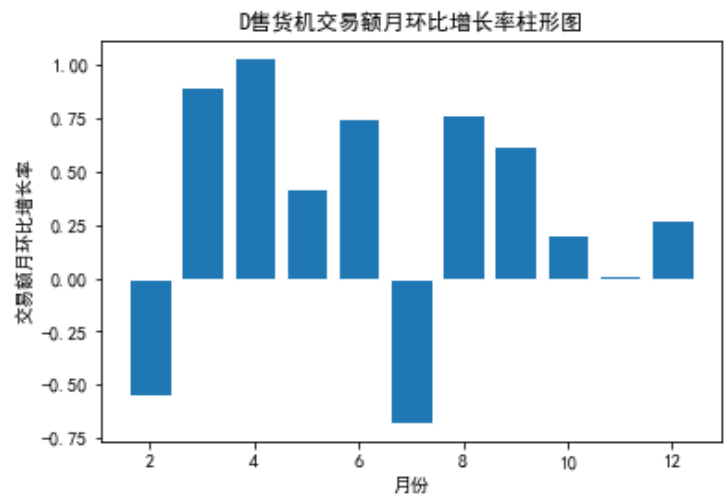
绘制B售货机交易额月环比柱形图如下：



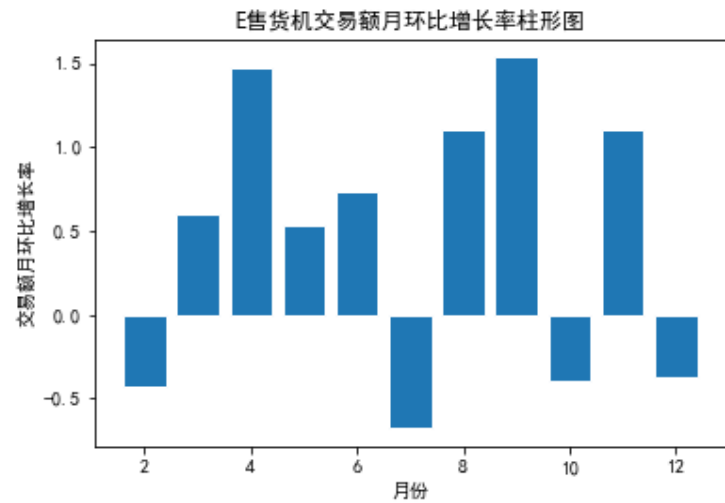
绘制C售货机交易额月环比柱形图如下：



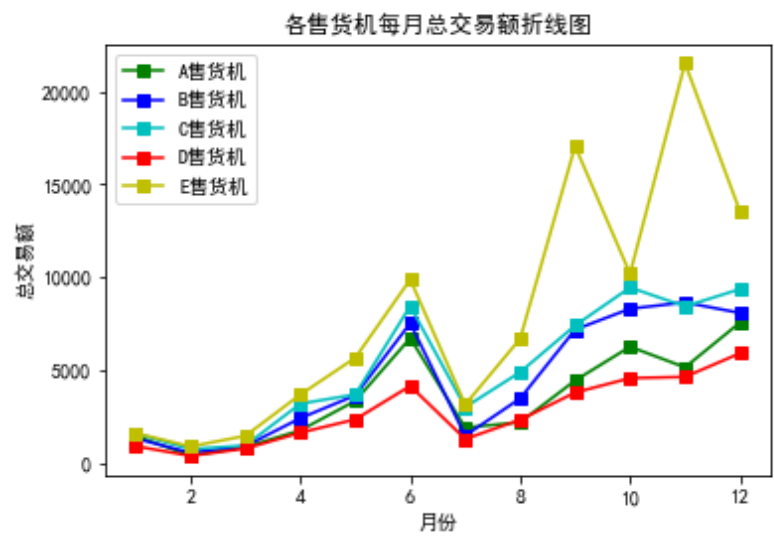
绘制D售货机交易额月环比柱形图如下：



绘制E售货机交易额月环比柱形图如下：



每个售货机每月交易额折线图绘制如下：



可以看出E售货机的销售额比其他售货机高，且各个售货机的大致趋势进本相同。

3. 2. 3 任务 2. 3 的解决

先将附件1和附件2以商品为中心合并为一张表上，提取大类为饮料类的商品，将其实际金额乘以25%得出饮料类毛利润，再提取大类为非饮料类的商品，将其实际金额乘以20%得出非饮料类毛利润。然后将每台售货机的饮料类和非饮料类的毛利润相加，得出每台售货机的毛利润，最后绘制每台售货机毛利润占总毛利润的饼图。

饼图绘制如下：

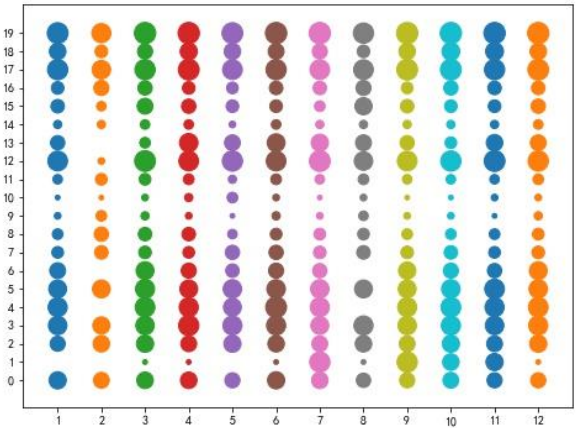


可以看出售货机E占比最大，占33.6%。

3. 2. 5 任务 2. 4 的解决

先以商品为中心合并附件1和附件2, 在利用for循环得出每个二级类目商品的每个月交易额均值，气泡大小与每个月各二级类目商品的交易额均值大小有关。

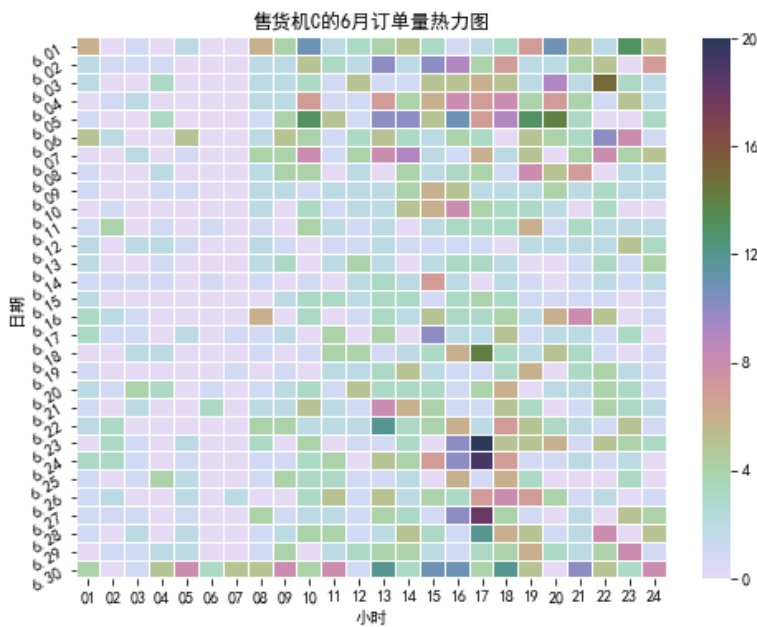
绘制气泡图如下：



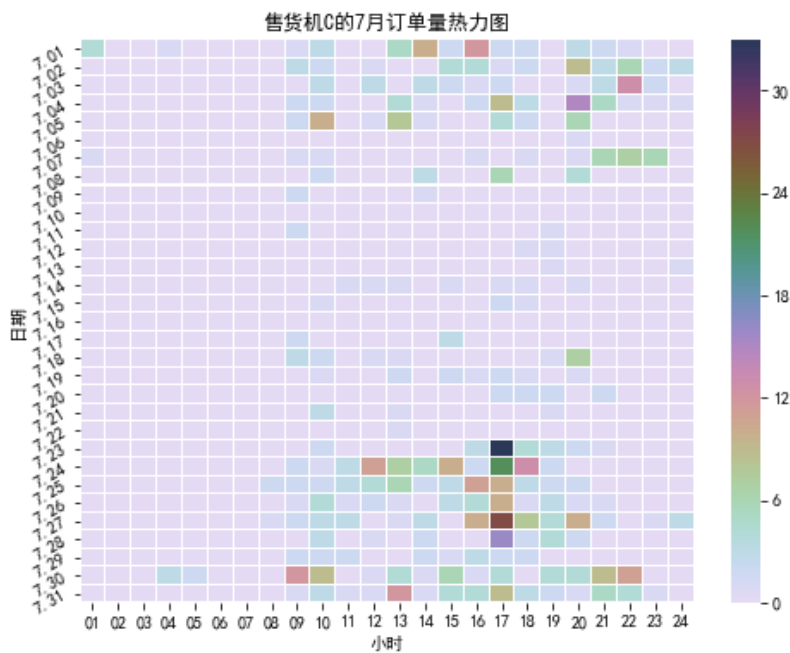
横轴表示月份，纵轴表示商品，气泡越大代表该商品在本月交易额均值越高。

3.2.5 任务 2.5 的解决

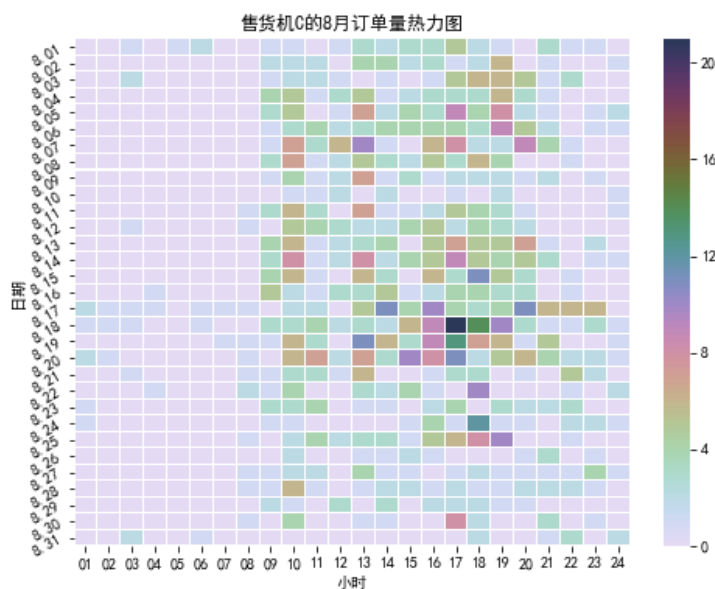
先提取售货机6, 7, 8月的数据，将6月份的数据一小时的单位分成($30 \times 24 = 720$)份，再计算每份的个数即为每小时的订单量，利用for循环得出 $[1 \times 720]$ 的列表，再将其转换为 $[30 \times 24]$ 的数组，最后利用其绘制售货机6月订单量的热力图。(7, 8月方法类似)
售货机6月订单量的热力图绘制如下：



售货机7月订单量的热力图绘制如下：



售货机8月订单量的热力图绘制如下:



颜色越浅代表订单量越少，显然6月订单量较多。

3.3 任务三：生成自动售货机的画像

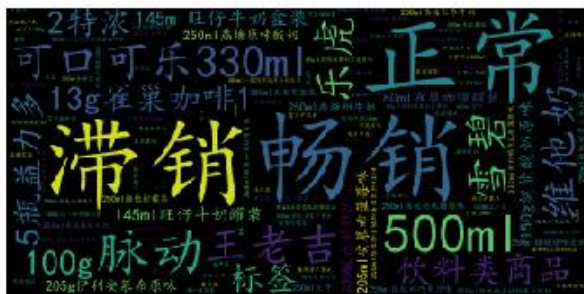
3.3.1 任务 3.1 的解决

与任务2.3类似，先将附件1和附件2以商品为中心合并为一张表上，提取大类为饮料类，地点为A的商品，再利用透视表将每类商品的订单量汇总计算，计算订单量三分位点，运用if语句，根据三分位点定义商品标签（滞销，正常，畅销），最后保存为csv文件。售货机B, C, D, E定义标签的方法类似。

3.3.2 任务 3.2 的解决

利用任务 3.1 保存的 csv 文件绘制画像，使用词云图将画像表示出来。

A 售货机的画像如下



B 售货机的画像如下



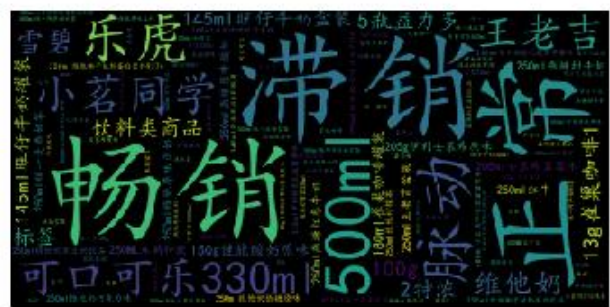
C 售货机的画像如下



D 售货机的画像如下



E 售货机的画像如下



从画像可以看出畅销，正常，滞销大小相差不大，是因为分类依据是三分位点。

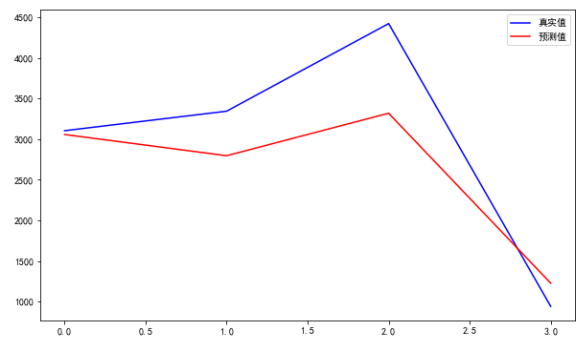
3.4 任务四：业务预测

3.4.1 任务 4.1 的解决

先合并附件1和附件2，将售货机A饮料类每月交易额数据提取出来进行预测，利用回归模型进行预测，其他售货机预测方法类似。

3.4.2 任务 4.2 的解决

售货机A饮料类每月交易额仅有12个数据，要预测2018年1月的交易额，数据太少。下图是售货机A饮料类每月交易额的预测值与真实值的比较图像



虽然数据可能趋势相近但是误差太大，下表为预测误差表

方法名称	最优值	预测误差值（保留两位小数）
平均绝对误差	0.0	496.53
均方误差	0.0	401425.73
中值误差	0.0	417.33

由上表可以看出仅由 12 个数据所进行的预测太过粗糙，误差值都太大，显然其他售货机的预测也将出现这种结果，故此不一一尝试。建议经营商给出 2017 年之前的几年数据，数据量越大，预测更精确，利用的方法也会变多。季节变动的影响也会减少，时间序列的预测方法也会变多，如一次移动平均等方法。