

BDA 数据分析师实践考核

报告名称： 烟草销售评估分析及规律预测

考核级别： BDA 初级数据分析师

作 者： 江昱峰

单 位： 无

联系方式： 13914250795

提交时间： 2023.1.20

目录

目录.....	2
执行性摘要	3
引言	4
正文.....	5
一、问题分析	5
1.1 卷烟四个价位销量占比的特点	5
1.2 对于真龙（起源）、硬盒芙蓉王、84 软盒玉溪的讨论	6
1.3 其他价位段的三种主销商品	10
1.4 灵敏度检验.....	12
二、局限性	13
三、结论与建议.....	13
3.1 结论	13
3.2 建议	14
附录：	15
一、参考文献	15
二、Python 绘图柱形图代码.....	16

执行性摘要

本文针对我国现有的烟草行业仍存在的产品差异性、品牌规模与竞争等问题，通过对 2019-2021 年间一家烟草公司的客户相关数据以及多指标销售数据的筛选、分析与处理，通过数据分析和时间序列的回归预测方法求解，得到了最终的评估方案以及预测结果，给出了结论。

在已知卷烟的四个价位段之后，对原始数据进行筛选处理，去除了其中 2019 年的部分存疑数据，完成数据预处理。通过 Excel 和 Python 处理数据后得到相应图像，并结合同比进行对稳定性的评估，得到下列最终结果：

1、低端卷烟销量占比 57%，中端占比 34%，中高端占比 8%，高端占比 1%。

2、中端三种主销商品的销量特征、数量关系、三者间存在竞争关系而不存在替代关系和对它们未来一年销量变化趋势的预测。

3、其他三个价位的主销商品名称，销量间的关系以及预测未来一年销量占比变化。

在评估中，通过灵敏度分析，得到的误差在 3.3%，证明计算过程具有较强的鲁棒性。

关键词：销售分析，时间序列，评估预测

引言

烟草行业是我国国民经济体系里重要的一环，对于满足我国国民的消费需求以及增加国家财政收入都有重要的意义。但是尽管近年来政府以及各烟草公司通过各种手段力图提高烟草产品的质量、销量等，烟草行业目前仍然存在许多问题，如产品差异性不显著，产品特色不突出，产品结构有待优化，品牌规模小，优势品牌竞争力与影响力不够等。因此，我将通过对一家烟草公司 2019-2021 年的相关销售数据进行分析与预测，解决如下问题：

分析卷烟价位的四个段位，说明其销售占比特点；并具体分析真龙（起源）、硬盒芙蓉王、84 软盒玉溪这三种中端销售商品，说明他们销量的特征、数量的关系、是否存在竞争和替代关系以及未来一年的趋势是否会发生变化；分析其他三个价位的主销商品，并回答它们在销量上的关系以及未来 1 年的占比变化。

正文

卷烟的价位一共被分成了四档：100-150 元/条的低端卷烟，200-250 元/条的中端卷烟，250-500 元/条的中高端卷烟以及 500-1000 元/条的高端卷烟。之后的讨论都在此基础上进行。

由于原始数据是公司相关人员处理后的数据，经检查并没有出现数据缺失、重复情况，但 2019 年的部分数据和公司实际销售情况仍有出入，考虑到数据量本身较大，本文直接剔除了这些存疑数据，之后将针对不同的问题利用相应的数据进行处理分析。

一、问题分析

1.1 卷烟四个价位销量占比的特点

对于这个问题，我将 2019 年-2021 年的四档位的卷烟销量分别相加，得到四个档位卷烟在这 3 年中各自的销量总和，进而计算得到它们的销量占比，处理后的数据如下：

表 1：四价位销量占比

价位段（元/条）	3 年总销量（箱）	销量占比
100-150	332701	57%
200-250	199553	34%
250-500	47478	8%
500-1000	7565	1%

数据可视化后结果如下：

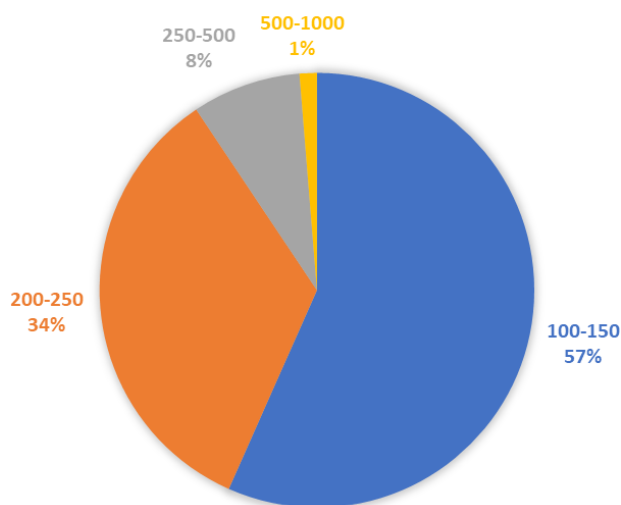


图 1：四价位销量占比图像

由上图可知，4 个价位中，低端卷烟（100-150 元/条）的销量占比最大且超过 50%，这符合我国百姓经济状况的整体情况；中端卷烟（200-250 元/条）的销量占比第二大，超过 30%；剩下的中高端卷烟（250-500 元/条）和高端卷烟（500-1000 元/条）占比依次为 8%与 1%，两者之和小于 10%。说明整体上看来我国大部分烟民日常买的卷烟都属于中低端范围，节假日时可能会购买中高端卷烟作为送礼。

1.2 对于真龙（起源）、硬盒芙蓉王、84 软盒玉溪的讨论

真龙（起源），硬盒芙蓉王、84 软盒玉溪是中端价位（200-250 元/条）的三种主要销售商品，下面我将对他们就销量特征，数量关系，竞争替代关系以及未来趋势进行研究。

(1) 三者的销量特征：对于销量特征，我小队经过对原始数据的处理，将每种卷烟在 3 年中每月的销量进行加权平均，得到了它们各自每月销量的样本值。对于权重，由于 2021 年处于更靠后的位置，因此我将其权重设置的较大。最终，三者的权重分别为：0.25，0.35，0.4。计算并使用 python 处理后得到最后的结果图如下：

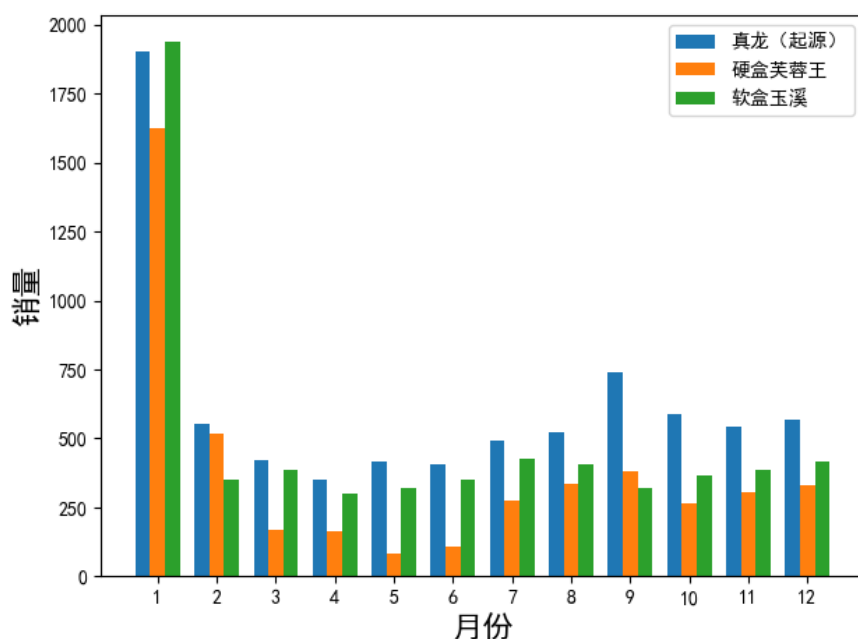


图 2：真龙（起源）、硬盒芙蓉王、84 软盒玉溪的销量特征（单位：箱）

由上图可知，三者销量的共同点为：1 月份销量最高，之后 11 个月份销量较低。不同点：总体上看硬盒芙蓉王的各月的销量基本都是三者中的最低；真龙（起源）虽然 1 月份没有软盒玉溪多，但是之后的所有月份都是三者中最高；84 软盒玉溪后 11 个月的销量最为稳定。

(2) 三者的数量关系：

本问题的结果仍可从图 1 中看出。三者在 1 月份的销量都超过 1500 箱，在 2-12 月份的销量都低于 1000 箱。除 1 月、2 月、9 月外，

三者的大小关系都是真龙（起源）＞ 84 软盒玉溪 ＞ 硬盒芙蓉王。1 月份中，84 软盒玉溪＞ 真龙（起源）＞ 硬盒芙蓉王；2 月份中，真龙（起源）＞ 硬盒芙蓉王 ＞ 84 软盒玉溪；9 月份中，真龙（起源）＞ 硬盒芙蓉王 ＞ 84 软盒玉溪。

（3）三者的竞争与替代关系：

对于本问题，我将三者加权处理后的每月销量数据制作成了如下折线图：

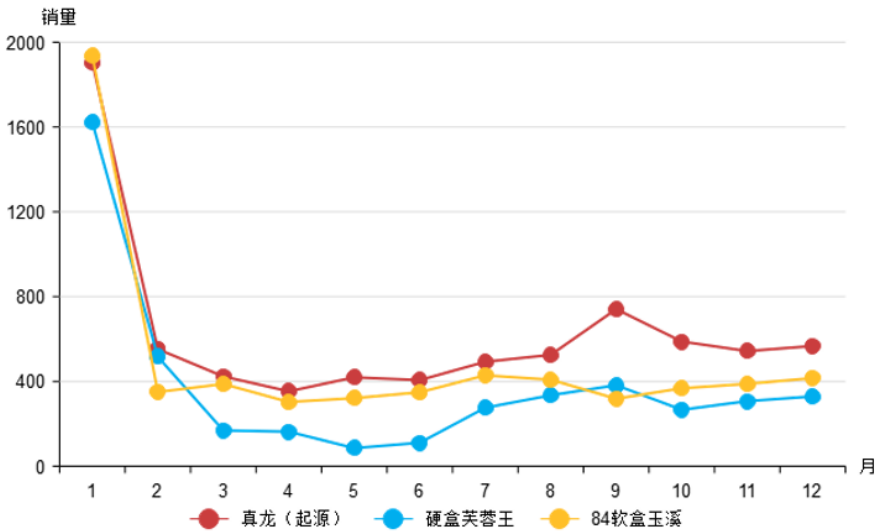


图 3：真龙（起源）、硬盒芙蓉王、84 软盒玉溪的竞争关系

分析上图可知：这三者的销量随时间的变化趋势类似，说明这三者在一年中的相同月份里营业状况类似，生意兴旺程度相似，也即三者有着较强的竞争关系；而三者的数据虽然有大小之分，但是总体变化趋势相同，不存在一种卷烟的销量逐渐变少直至趋于 0 的情况，说明三者之中并不存在替代关系。结合实际情况，烟民一般会持续购买同一种品牌的卷烟，不会随意更换，因此结论符合常理。

（4）三者未来一年的趋势：

在之前的问题中，我主要针对的是销售中的销量、销售额、品牌宽度这 3 个属性。但是在对某一特定品牌的卷烟未来趋势的预测中，还可以使用到另外的几个属性。在考虑到这个问题的性质后，在这一问中，我引入了同比的概念：

$$X_{\text{同比}} = X_{\text{本期}} - X_{\text{同期}} \quad (1)$$

其中， $X_{\text{本期}}$ 指的是当月的数据， $X_{\text{同期}}$ 指的是去年当月的数据。因此同比的含义为本月数据减去去年本月数据。

在此基础上，我分别计算了 2018-2019 年与 2020-2021 年这三者的销量同比随月份的变化曲线，曲线如下：

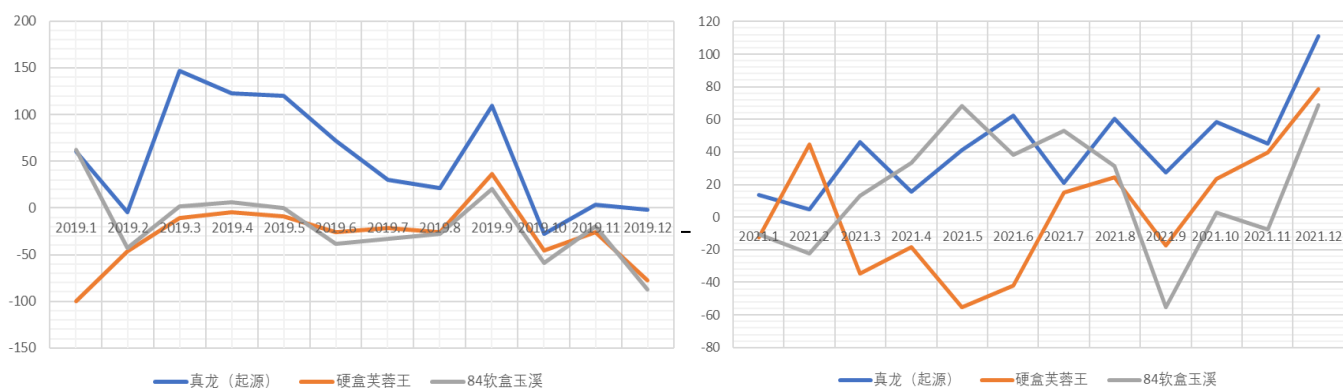


图 4：真龙（起源）、硬盒芙蓉王、84 软盒玉溪 2019 与 2021 年同比销量图（万元）

由上图左右对比后可知：真龙（起源）的同比保持在正值，且同比有持续增大趋势，说明真龙（起源）在未来一年的销量会保持上升趋势；硬盒芙蓉王的同比由一开始的稳定在 0 附近变为波动性大于 0，且最后有持续增大趋势，说明其在未来一年销量会波动上升，波动幅度可能较大；84 软盒玉溪与硬盒芙蓉王相似，但其在 2021 年同比大

于 0 的月数更多，说明其在未来一年销量会波动上升，波动幅度可能较小。

1.3 其他价位段的三种主销商品

根据上面对卷烟价位段的划分，在处理数据后，我得出以下结果（下列结果按照从销量大到小排列）：

- （1）100-150 元/条：真龙（软祥云），84 软盒红河（99），真龙（轩云）
- （2）250-500 元/条：真龙（海韵），硬盒中华，芙蓉王（硬蓝新版）
- （3）500-1000 元/条：软盒中华，真龙（海韵细支），真龙（软海韵）

1. 对于各个价位段上三种主销商品销量上的关系，我分别分析了各个价位段三种主销商品加权平均后每个月的销量，绘制成了如下柱状图：

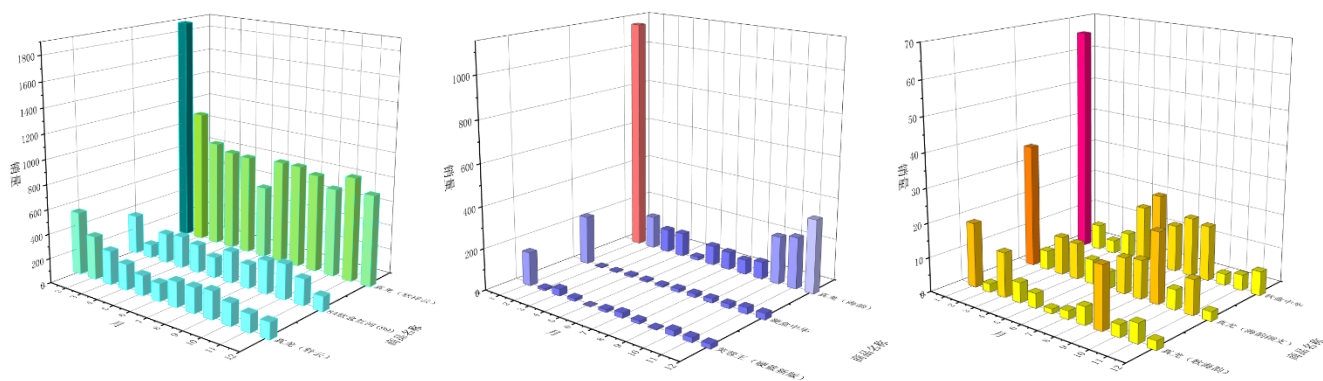


图 5：（从左到右）低端、中高端、高端卷烟前三种主销商品的各月销量关系（单位：箱）

在上面三张图中，商品名称处于坐标轴右端的年总销量大，处于坐标轴左端的年总销量小。从销量上看，无论是哪种价位上的主销商品，其销量都在 1 月份时达到巅峰，在 2-12 月份有起伏但相较 1 月份而言都较小。且整体上看来低端卷烟的销量远高于中高端卷烟的销量高于高端卷烟的销量。

2. 对于这些主销商品未来 1 年销售占比的变化，我计算了这些主销商品 3 年的占比情况，绘制成了如下的表格：

表 2：低端卷烟 3 种主销商品 2019-2021 销售占比情况

名称/月份	2019	2020	2021
真龙（软祥云）	26%	34%	31%
真龙（轩云）	7%	9%	7%
84 软盒红河（99）	4%	9%	9%

表 3：中高端卷烟 3 种主销商品 2019-2021 销售占比情况

名称/月份	2019	2020	2021
真龙（海韵）	54%	59%	53%
硬盒中华	14%	9%	8%
芙蓉王（硬蓝新版）	6%	8%	7%

表 4： 高端卷烟 3 种主销商品 2019-2021 销售占比情况

名称/月份	2019	2020	2021
软盒中华	31%	29%	22%
真龙（海韵细支）	12%	17%	17%
真龙（软海韵）	9%	6%	10%

由上表可以预测出，未来 1 年各个价位的前 3 种主销商品的占比变化为：

- （1）对于低端卷烟：真龙（软祥云）的占比波动上升，真龙（轩云）的占比在 8%附近波动，84 软盒红河（99）的占比上升；
- （2）对于中高端卷烟：真龙（海韵）的占比波动上升，硬盒中华的占比缓慢下降，芙蓉王（硬蓝新版）的占比在 7%附近波动；
- （3）对于高端卷烟：软盒中华的占比下降，真龙（海韵细支）的占比缓慢上升，真龙（软海韵）的占比波动上升。

1.4 灵敏度检验

在问题求解中，对于特定销售商品每个月销量的情况，我通过其在 2019-2021 年同一月份的销量进行加权求和，而权重的确定具有主观的成分，我根据年份越相近影响程度越大的规律而人为分配了 0.25、0.35、0.4 的权重。为了检验这样计算的结果的稳定性，我选取了对

200-250 价位段中三种主销商品真龙（起源）、硬盒芙蓉王、84 软盒玉溪的计算结果为灵敏度分析对象，对原结果进行了一定的干扰，调整了权重分配，把权重分别设置为 0.2、0.3、0.5 和 0.15、0.25、0.6，计算得到了两组干扰后的结果，并与原结果进行了可视化比对。

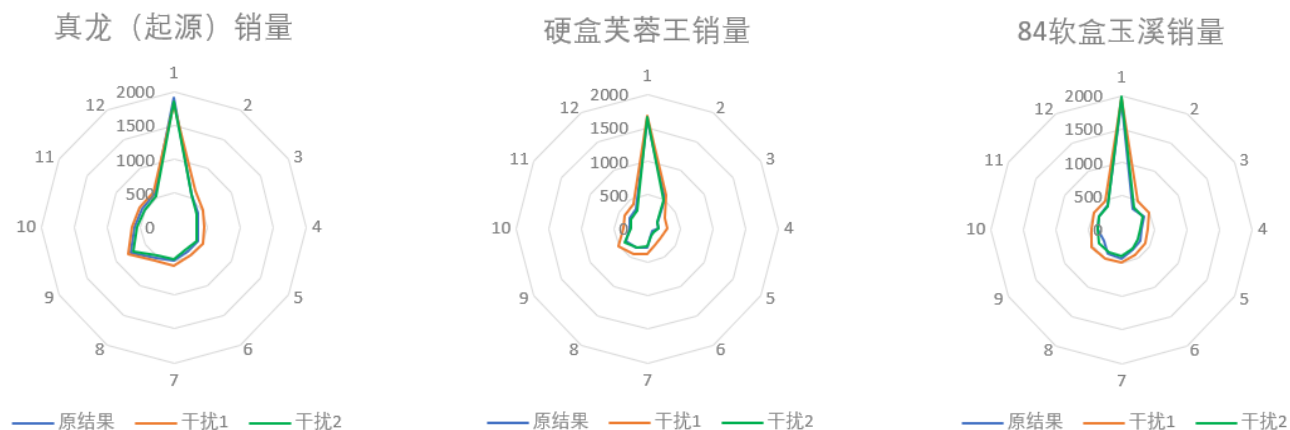


图 6：灵敏度分析

二、局限性

- (1) 只能基于 2019-2021 共 3 年的数据进行预测，可参考的数据较少，预测结果准确性略显不足；
- (2) 做出了一些假设，导致一定的理想化。

三、结论与建议

3.1 结论

- (1) 我国大部分烟民日常买的卷烟都属于中低端范围，节假日时可能会购买中高端卷烟作为送礼；
- (2) 三种烟的销量随时间的变化趋势类似，说明这三者在一年中的

相同月份里营业状况类似，生意兴旺程度相似，也即三者有着较强的竞争关系；而三者总体变化趋势相同，不存在一种卷烟的销量逐渐变少直至趋于 0 的情况，说明三者之中并不存在替代关系；

（3）未来 1 年各个价位的前 3 种主销商品的占比变化为：

对于低端卷烟：真龙（软祥云）的占比波动上升，真龙（轩云）的占比在 8%附近波动，84 软盒红河（99）的占比上升；

对于中高端卷烟：真龙（海韵）的占比波动上升，硬盒中华的占比缓慢下降，芙蓉王（硬蓝新版）的占比在 7%附近波动；

对于高端卷烟：软盒中华的占比下降，真龙（海韵细支）的占比缓慢上升，真龙（软海韵）的占比波动上升。

3.2 建议

（1）加大年初及节假日的客户联系，加大销售力度，旺季进行扩大销售，并且可以通过大客户的介绍扩大销售渠道，联系其他的客户和商家，以大带大。淡季销售以中低端卷烟主体，在小客户之间增强自己的宣传，提升自己的质量，提高口碑，从而增大自己的客户基数；

（2）将真龙品牌打造成优势品牌，提高质量，增大销售的渠道，进一步地打响公司的名号，从而带动其他的产品的销售量；

（3）做好一定的风险预案，合理控制卷烟的产量，做好相关的货储管理，以便能够及时的应对出现的销售变化，做出合理的调整。

附录：

一、参考文献

[1] 百度百科，总体标准偏差

网址：<https://www.baike.baidu.com/item/总体标准偏差>

[2] CSDN 作者卖山楂啦 prss，CRITIC 权重法

网 址 :

https://blog.csdn.net/qq_42374697/article/details/106742082

[3] CSDN 作者风度 78，时间序列预测模型,

网址：<https://blog.csdn.net/fengdu78/article/details/121347188>

二、Python 绘图柱形图代码

```
import matplotlib
import matplotlib.pyplot as plt
import numpy as np

plt.rcParams['font.sans-serif']=['SimHei'] #解决中文乱码

xlabel=[i for i in range(1,13)]
y1=[1904.1000,552.3880,423.5280,353.3160,419.1640,406.0880,492.49
60,
525.0920,740.9280,588.0760,543.0880,566.7800]
y2=[1621.7680,516.3520,168.6840,162.8520,85.2640,110.6600,275.732
0,
334.6000,382.0040,266.2160,306.8320,329.2360]
y3=[1936.2040,349.9640,387.9040,302.8640,321.0640,348.6680,428.70
80,
407.9960,318.2760,367.5920,388.3880,415.5520]
x=np.arange(len(xlabel))
width=0.25 #柱状图的宽度
fig,ax=plt.subplots()
p1=ax.bar(x-width,y1,width,label='真龙(起源)')
p2=ax.bar(x,y2,width,label='硬盒芙蓉王')
p3=ax.bar(x+width,y3,width,label='软盒玉溪')
ax.set_ylabel('销量',fontsize=16)
ax.set_xlabel('月份',fontsize=16)
ax.set_xticks(x)
ax.set_xticklabels(xlabel)
ax.legend()
fig.tight_layout()
plt.show()
```