

## 2024年第三届钉钉杯大学生大数据挑战赛复赛题目

### 复赛A：烟草营销案例数据分析

#### 一、问题背景:

烟草是我国重要的国家税收和财政收入来源。在现有市场中占有极为稳定的根基。据统计，近年来我国卷烟销售收入逐年增长。现在国家对烟草实行专卖制度，对烟草专卖品的生产、销售、进出口依法实行专卖管理，并实行烟草专卖许可制度。因为烟草是一种特殊的商品，国家对其实行专卖专营政策，对烟草及其制品的生产 and 流通进行严格管理，所以卷烟不能像其它商品那样随意生产和流通。由于烟草产品的特殊属性，必须实行行政管制，集中统一生产销售才有更好的控制力和经营秩序，故实行由国家管制发展的烟草专卖制度不是权宜之计，而是力求社会目标和经济目标最优的制度安排。国家管制、计划生产、许可经营和政企合一体制成为烟草专卖制度的主要内容。

我国的烟草产业链由烟叶种植与购销、烟标制造、烟叶加工与卷烟制造、卷烟批发、卷烟零售五个主要环节构成。上游以烟叶种植与购销为核心，我国卷烟制造所使用烟叶基本依靠国产，由中国烟草总公司下属的各省烟草公司集中采购。上游另一生产活动为卷烟包装，主要包括卷烟工业用纸和烟标两大产品的产销经营。中游为各省级中烟工业公司负责卷烟的加工和生产，将集中采购的原材料按类型分配至各卷烟复烤企业和卷烟材料生产企业，最后由卷烟生产企业制成品。下游为成烟销售活动，由国家烟草专卖局统筹规划，再由各省级烟草专卖公司通过颁发烟草专卖许可证管控批发与零售渠道。

该数据来自某地区近些年多种品牌的烟草销售情况，已经做了脱敏和数据变换处理。每种烟草的数据为1个excel文件，共5个文件，每个文件内共5列，记录了各类烟草品牌各月的销售情况等。具体说明见表1。

表1

变量编号	变量名	释义
1	月份	销售的自然月
2	样品代码	产品身份标识
3	名称	产品名称
4	销量（箱）	产品当月销售箱数
5	金额（元）	产品当月销售总金额

## 二. 解决问题

### 1. 聚类分析

(1) 通过数据分析选取合适的指标，这些指标的目的在于能够描述不同香烟品牌的画像特征，说明选择这些指标的理由和合理性。

(2) 结合选取的特征指标，使用2种聚类算法，对多个香烟品牌的销售数据进行聚类分析，找出具有相似销售特征的品牌。自行选择聚类算法并设计模型的结构和参数。自行选择或设计美观、合理的可视化图表展示聚类结果，并对聚类结果的机理、原因、特点进行详细的分析。

### 2. 影响分析

针对上述5个香烟品牌的销售数据进行分析，定性和定量的分析价格、年份、时间对销售数据的影响，请提供模型的设计思路、实现过程、完整代码和分析结果。