

目录

| | |
|-----------------------|---|
| 任务 1：数据预处理 | 2 |
| 任务 1.1 | 2 |
| 任务 1.2 | 2 |
| 任务 2： 肥料产品的数据分析 | 2 |
| 任务 2.1 | 2 |
| 任务 2.2 | 3 |
| 任务 2.3 | 4 |
| 任务三 | 6 |
| 任务 3.1 | 6 |
| 任务 3.2 | 8 |
| 任务 3.3 | 8 |
| 任务 4 | 8 |
| 任务 4.1 | 8 |
| 任务 4.2 | 9 |

任务 1：数据预处理

任务 1.1

题目复述：附件 1 的产品通用名称存在不规范的情况。请按照复混肥料（掺混肥料归入这一类）、有机-无机复混肥料、有机肥料和床土调酸剂这 4 种类别对附件 1 进行规范化处理。

任务 1.2

题目复述：计算附件 1 中各肥料产品的氮、磷、钾养分百分比之和，称为总无机养分百分比。

任务 2：肥料产品的数据分析

任务 2.1

题目复述：将所有复混肥料按照总无机养分百分比的取值等距分为 10 组。根据每个产品所在的分组，为其打上分组标签（标签用 1~10 表示）。分析复混肥料产品的分布特点，在报告中绘制产品登记数量的直方图，给出处理思路及过程，并按登记数量从大到小列出登记数量最大的前 3 个分组及相应的产品登记数量。

总无机养分分组标签登记柱状图

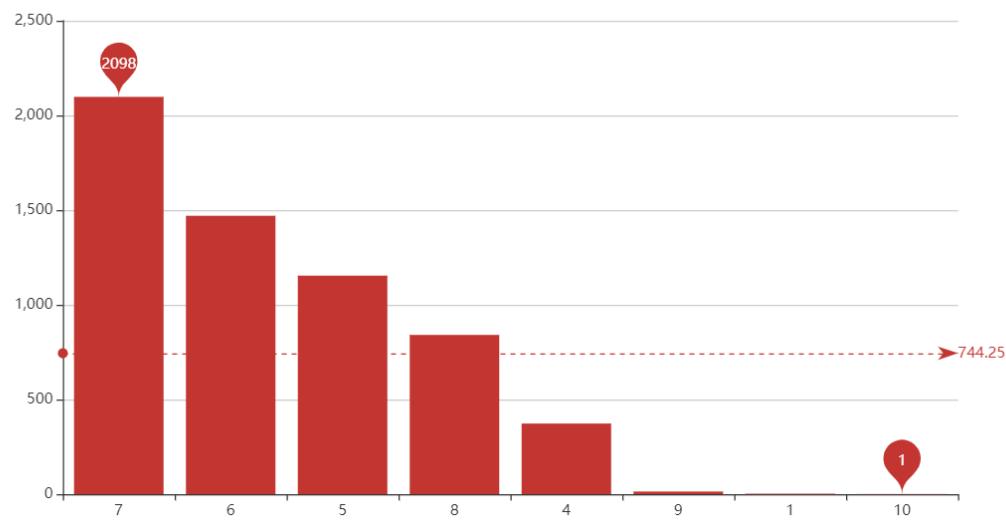


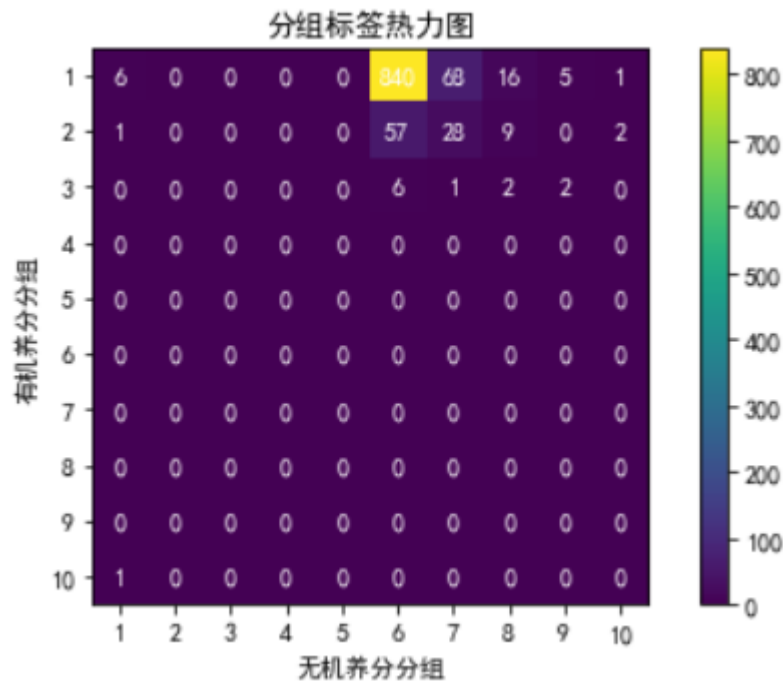
图 1

登记数量最大的前 3 个分组及相应的产品登记数量如下：

| | | | |
|--------|------|------|------|
| 排名 | 一 | 二 | 三 |
| 分组标签 | 7 | 6 | 5 |
| 产品登记数量 | 2098 | 1470 | 1154 |

任务 2.2

绘制有机肥料产品的分布热力图



任务 2.3

此任务需要运用 k 均值聚类算法

k 均值聚类算法（k-means clustering algorithm）是一种迭代求解的聚类分析算法，其步骤是，预将数据分为 K 组，则随机选取 K 个对象作为初始的聚类中心，然后计算每个对象与各个种子聚类中心之间的距离，把每个对象分配给距离它最近的聚类中心。聚类中心以及分配给它们的对象就代表一个聚类。每分配一个样本，聚类的聚类中心会根据聚类中现有的对象被重新计算。这个过程将不断重复直到满足某个终止条件。终止条件可以是没有（或最小数目）对象被重新分配给不同的聚类，没有（或最小数目）聚类中心再发生变化，误差平方和局部最小。

先随机选取 K 个对象作为初始的聚类中心。然后计算每个对象与各个种

子聚类中心之间的距离，把每个对象分配给距离它最近的聚类中心。聚类中心以及分配给它们的对象就代表一个聚类。一旦全部对象都被分配了，每个聚类的聚类中心会根据聚类中现有的对象被重新计算。这个过程将不断重复直到满足某个终止条件。终止条件可以是以下任何一个：

- 1)没有（或最小数目）对象被重新分配给不同的聚类。
- 2)没有（或最小数目）聚类中心再发生变化。
- 3)误差平方和局部最小。

经过运行，得到肥料产品的三维散点图（图 6）和散点图矩阵（图 7）

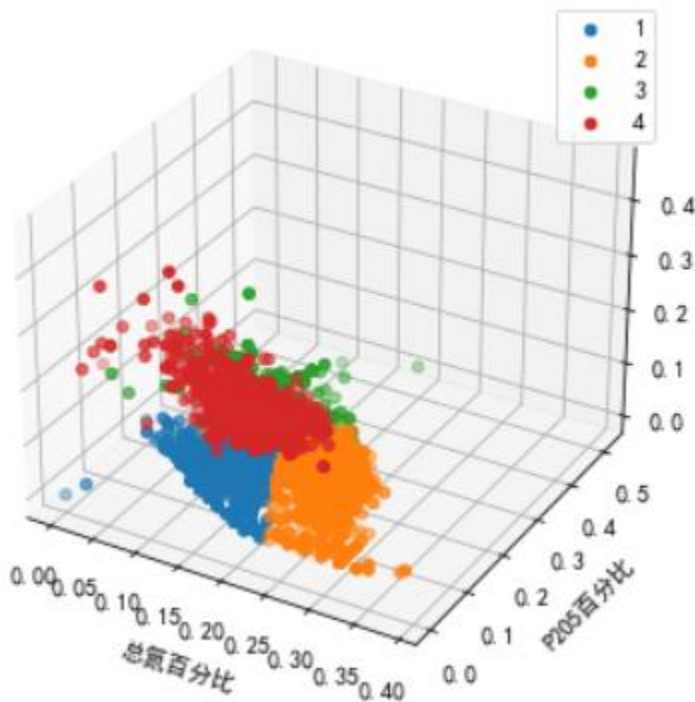


图 2

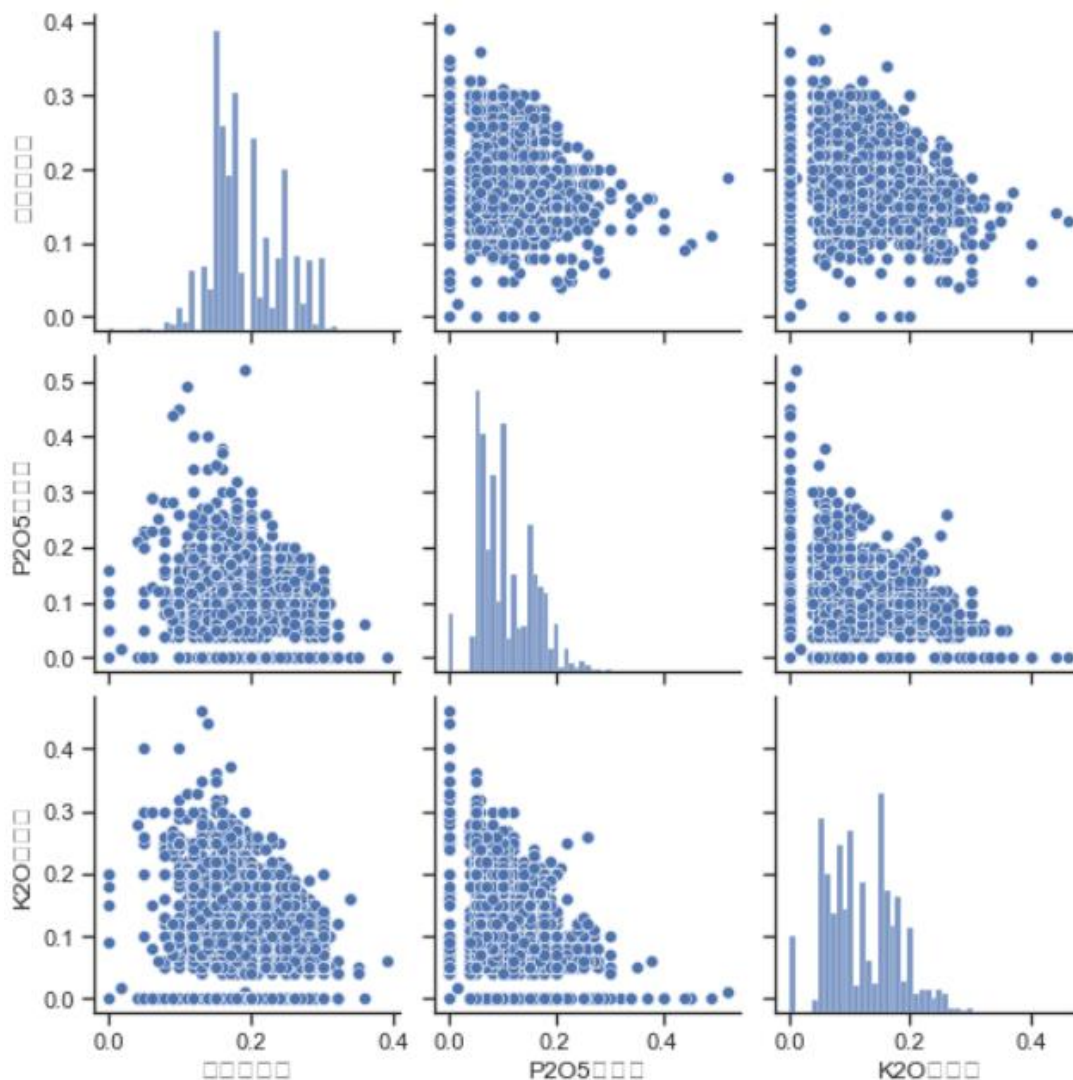


图 3

任务三

任务 3.1

题目复述：分析比较复混肥料中各组别不同年份产品登记数量的变化趋势，使用图表对结果进行可视化。

基于文件“result2_1.xlsx”中“发证日期”中的数据未统一(datetime)

政策影响

主要与实施《条例》和配套规章提高登记门槛有关，还可能受环保、安全生产及化肥生产许可证的颁发等政策限制，造成化肥企业转行、兼并重组等，影响了企业产品登记。

原材料高压

复混肥作为二次加工产品，受原材料行情波动的影响尤其明显，在原材料集体高位上涨前提下，复混肥也被推到了风口浪尖，但相比单质肥的涨幅，复合肥涨幅相对较低，且涨价开启滞后。

经销商抵触

“长期看供应，短期看需求”，就复混肥产能严重过剩的行业现状来讲，供应紧张基本是不存在的，那下游需求来看，一直是不温不火的状态。近几年原材料行情持续高位，一路喊涨，复混肥价格也涨至高位，低价货源难寻，近几年尤其今年，市场集中备货高峰期很难再现，并且基层零售价格倒挂成为普遍现象。近阶段，复混预收方面经销商打款积极性有限，企业的预收进度普遍较低，部分企业直呼收款困难，苦不堪言，下游经销商的抵触心态也从未放松过。

市场假货横行

农资市场五花八门、鱼目混珠的假肥料从来没有缺席过。假货横行，专业人士都真假难辨，更何况是善良朴实的农民，但在化肥价格居高不下行情之下，假化肥对复混化肥市场的冲击也是不容小觑的。

新型肥挤压

行情疲软，企业可操作空间较小，部分企业重心转向生产差异化肥料。另外在国家政策导向之下，化肥产业结构不断调整，化肥需求稳中有降。复混化肥进一步受到新型肥挤压，市场占有率持续下行。

任务 3.2

题目复述：提取 2021 年 9 月 30 日仍有效的有机肥料产品，从有效产品中分别筛选出广西和湖北产品登记数量在前 5 的组别，分析两个省份上述组别的分布差异。

任务 3.3

题目复述：提取产品登记数量大于 10 的肥料企业，给出这些企业所用到的原料集合（发酵菌剂除外）。以各企业用到的原料作为特征，计算企业之间的杰卡德相似系数矩阵，并将结果（保留 4 位小数）

任务 4

任务 4.1

题目复述：设计算法或处理流程，从附件 4 技术指标中提取出氮、磷、钾养分和有机质的百分比，以及肥料含氯的程度。

任务 4.2

题目复述：提取各种原料的名称及其百分比。