

1. 问题重述

1.1 问题背景

从古至今，我国一直都是农业大国，而肥料是农业生产中一种重要的生产资料。对于肥料的生产销售，我国具有一定的规章制度，其生产销售必须遵循《肥料登记管理办法》，并且依法在农业行政管理部门进行登记。

1.2 问题提出

任务 1 数据的预处理

任务 1.1 按照复混肥料、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。请在报告中给出处理思路、过程及必要的结果，同时将完整的结果保存到文件“result1_1.xlsx”中。

任务 1.2 计算附件 1 中各肥料产品的总无机养分百分比。请在报告中给出处理思路、过程及必要的结果，同时将完整的结果保存到文件“result1_2.xlsx”中，结果保留 3 位小数（例如 1.0%，即 0.010）。

任务 2 肥料产品的数据分析

任务 2.1 从附件 2 中筛选出复混肥料的产品，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），将完整的结果保存到文件“result2_1.xlsx”中。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。

任务 2.2 从附件 2 中筛选出有机肥料的产品，将产品按照总无机养分百分比和有机质百分比分别等距分为 10 组，并为每个产品打上分组标签 (1,1), (1,2), ..., (10,10)，将完整的结果保存到文件“result2_2.xlsx”中。请在报告中给出处理思路及过程，并根据分组情况绘制有机肥料产品的分布热力图，其中横轴代表总无机养分分组，纵轴代表有机质分组。在此基础上，分析有机肥料产品的分布特点，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。

任务 2.3 从附件 2 中筛选出复混肥料的产品，按照氮、磷、钾养分的百分比，使用聚类算法将这些产品分为 4 类。根据聚类结果为每个产品打上聚类标签（标签用 1~4 表示），并将完整的结果保存到文件“result2_3.xlsx”中。请在报告中给出处理思路及过程，根据聚类标签绘制肥料产品的三维散点图和散点图矩阵，并通过绘制聚类结果的雷达图分析每个聚类的特征。

任务 3 肥料产品的多维度对比分析

任务 3.1 从文件“result2_1.xlsx”中提取发证日期中的年份，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。请在报告中给出处理思路及分析过程，使用合适的图表对结果进行可视化。

任务 3.2 从文件“result2_2.xlsx”中提取 2021 年 9 月 30 日仍有效的有机肥料产品，将完整的结果保存到文件“result3_2.xlsx”中。从有效产品中分别筛选出广西和湖北(根据正式登记证号区分)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。请在报告中给出处理过程及分析结果。

任务 3.3 从附件 3 中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）保存到文件“result3_3.xlsx”中。请在报告中给出处理思路、过程及相似系数矩阵。

任务 4 肥料产品的多维度对比分析

任务 4.1 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。请在报告中给出处理思路及过程，并将结果保存到文件“result4_1.xlsx”中。

任务 4.2 设计算法或处理流程，从附件 4 原料与百分比中提取各种原料的名称及其百分比。请在报告中给出处理思路及过程，并将结果保存到文件“result4_2.xlsx”中。

1.3 任务目标

1. 对肥料登记数据进行预处理。
2. 根据养分的百分比对肥料产品进行细分。
3. 从省份、日期、生产商、肥料构成等维度对肥料登记数据进行对比分析。
4. 对非结构化数据进行结构化处理。

2. 任务解决

2.1 任务 1 数据的预处理

2.2 任务 2 肥料产品的数据分析

2.2.1 任务 2.1

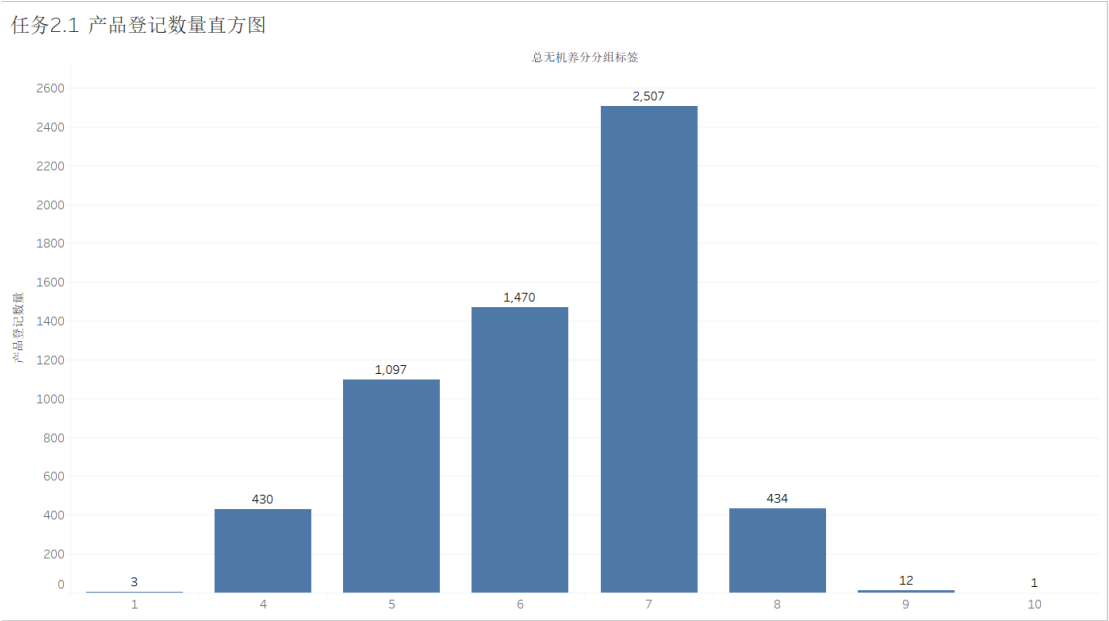


图 1 产品登记数量直方图

我们将直方图按产品登记数量进行降序排序，如图 2 所示，可得到登记数量最大的前 3 个分组为第 7 组、第 6 组和第 5 组，其产品登记数量分别为 2507, 1470, 1097。

排名	一	二	三
分组标签	7	6	5
产品登记数量	2507	1470	1097

任务2.1 产品登记数量直方图

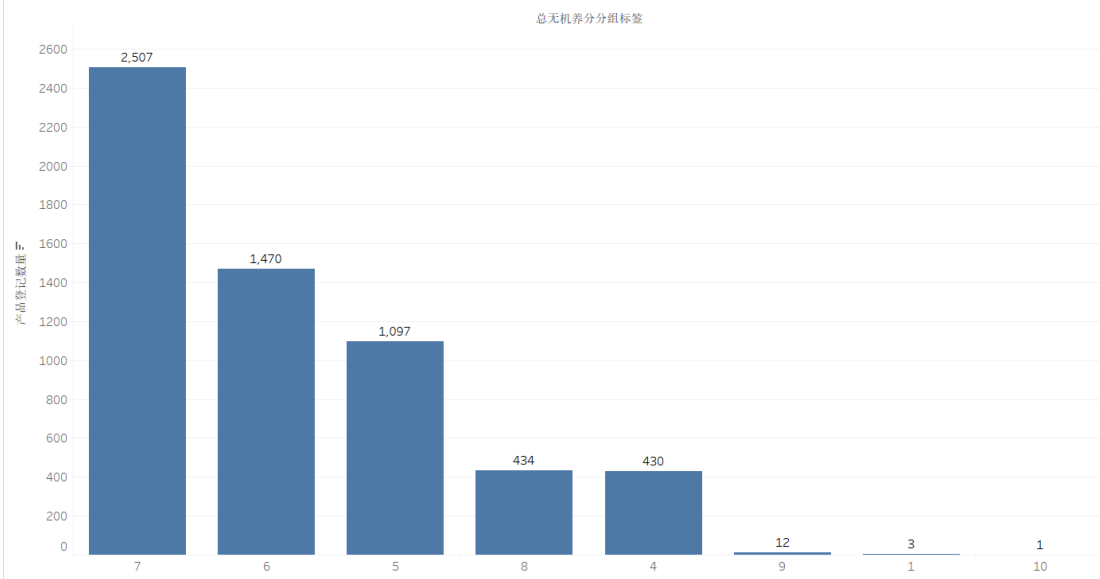


图 2 经降序排序的产品登记数量直方图