

# 肥料登记数据分析

## 摘要

为了更好地完成肥料登记工作，本文根据附件中提供的数据，解决以下四个主要任务。首先，进行数据预处理。利用 Excel 办公软件规范化处理产品通用名称不规范的情况，同时利用 Matlab 软件计算总无机养分百分比。然后，对肥料产品进行数据分析。对于任务 2.1、2.2，利用 Matlab 和 Excel 软件处理数据，筛选出复混肥料和有机肥料的信息。根据任务要求，找到相对应的等距分组并打上标签。为了更直观地分析出了复混肥料产品和有机肥料产品的分布特点，利用 TipDM-BI 数据分析和可视化平台绘制出了复混肥料产品和有机肥料的登记数量可视化图，从而分别看出两者按登记数量从大到小排列的前三名分组和相应的产品登记数量。对于任务 2.3，利用 Excel 软件获取所需信息。使用聚类算法按照氮、磷、钾养分的百分比进行分类并打上聚类标签。根据聚类标签绘制肥料产品的三维散点图和散点图矩阵和聚类结果的雷达图，分析出每个聚类的特征。接着，从省份、日期、生产商、肥料构成等维度对肥料产品进行对比分析。最后，设计了处理流程，借助 Excel 办公软件的功能，对附件 4 的信息进行人工处理，从而提取出氮、磷、钾养分百分比、有机质百分比以及肥料含氮的程度以及各种原料的名称及其百分比等信息。

**关键词：**聚类算法、可视化、多维度对比

## 一、问题重述

### 1.1 问题背景

肥料是农业生产中一种重要的生产资料，其生产销售必须遵循《肥料登记管理办法》，依法在农业行政管理部门进行登记。各省、自治区、直辖市人民政府农业行政主管部门主要负责本行政区域内销售的肥料登记工作。

### 1.2 问题描述

为了更好地完成肥料登记工作，本文根据附件中提供的数据，解决以下四个主要目标：

- (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。

## 二、问题分析

### 2.1 任务 1 分析

任务 1 要求对附件 1 所提供的肥料登记数据进行数据预处理操作。

### 2.2 任务 2 分析

任务 2 要求对附件 2 所提供的肥料登记数据进行肥料产品的数据分析操作。

### 2.3 任务 3 分析

任务 3.1 要求从文件“result2\_1.xlsx”中提取发证日期中的年份，通过分析比较各组别的复混肥料产品在不同年份时的登记数量变化趋势，并给出合适的图表结果来进行可视化操作。

任务 3.2 要求从文件“result2\_2.xlsx”中提取 2021 年 9 月 30 日，并仍有效的有机肥料产品，将所得的完整结果放入到“result3\_2.xlsx”文件中。根据

该表格的有效产品信息，利用正式登记证号区分出广西和湖北，同时分别筛选出两地的产品登记数量在前五名的组别，分析出两个省份在上述组别的分布差异。

任务 3.3 要求从附件 3 中提取产品登记数量大于 10 的肥料企业，并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，对企业之间的杰卡德相似系数矩阵进行计算，并将计算得到的结果放入到“result3\_3.xlsx”文件中，注意要保留 4 位小数。

2.4 任务 4 分析

任务 4.1 要求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，并将结果保存到文件“result4\_1.xlsx”中。

任务 4.2 要求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，并将结果保存到文件“result4\_2.xlsx”中。

三、数据预处理

3.1 任务 1.1

对于任务 1.1，利用 Excel 办公软件中的数据自动筛选功能对附件 1 中的数据进行筛选，并对其不规范的产品通用名称进行名称替换和修改。

3.2 任务 1.2

复混肥料	颗粒	16.00%	10.00%	18.00%	中氯	0.00%	皖农肥(2016)准字4279号	2016-01-17	2021-01	0.440
复混肥料	颗粒	20.00%	8.00%	16.00%	中氯	0.00%	皖农肥(2016)准字4280号	2016-01-17	2021-01	0.440
复混肥料	颗粒	20.00%	6.00%	14.00%	中氯	0.00%	皖农肥(2016)准字4281号	2016-01-26	2021-01	0.400
复混肥料	颗粒	20.00%	6.00%	14.00%	低氯	0.00%	皖农肥(2016)准字4282号	2016-01-26	2021-01	0.400
复混肥料	颗粒	18.00%	24.00%	12.00%	低氯	0.00%	皖农肥(2016)准字4283号	2016-01-26	2021-01	0.540
复混肥料	颗粒	18.00%	18.00%	18.00%	低氯	0.00%	皖农肥(2016)准字4284号	2016-01-26	2021-01	0.540
复混肥料	颗粒	17.00%	17.00%	17.00%	低氯	0.00%	皖农肥(2016)准字4285号	2016-01-26	2021-01	0.510
复混肥料	颗粒	16.00%	20.00%	12.00%	中氯	0.00%	皖农肥(2016)准字4286号	2016-01-26	2021-01	0.480
复混肥料	颗粒	16.00%	20.00%	10.00%	中氯	0.00%	皖农肥(2016)准字4287号	2016-01-26	2021-01	0.460
复混肥料	颗粒	27.00%	8.00%	8.00%	中氯	0.00%	皖农肥(2016)准字4288号	2016-01-26	2021-01	0.430
复混肥料	颗粒	30.00%	6.00%	0.00%	低氯	0.00%	皖农肥(2016)准字4289号	2016-01-26	2021-01	0.360
复混肥料	颗粒	30.00%	0.00%	6.00%	无氯	0.00%	皖农肥(2016)准字4290号	2016-01-26	2021-01	0.360
复混肥料	颗粒	20.00%	10.00%	15.00%	高氯	0.00%	皖农肥(2016)准字4291号	2016-01-26	2021-01	0.450
复混肥料	颗粒	28.00%	0.00%	6.00%	中氯	0.00%	皖农肥(2016)准字4292号	2016-01-26	2021-01	0.340
复混肥料	颗粒	18.00%	10.00%	26.00%	中氯	0.00%	皖农肥(2016)准字4293号	2016-01-26	2021-01	0.540
复混肥料	颗粒	16.00%	8.00%	16.00%	高氯	0.00%	皖农肥(2016)准字4294号	2016-01-26	2021-01	0.400
复混肥料	颗粒	22.00%	8.00%	14.00%	中氯	0.00%	皖农肥(2016)准字4295号	2016-01-26	2021-01	0.440
复混肥料	颗粒	26.00%	11.00%	11.00%	低氯	0.00%	皖农肥(2016)准字4296号	2016-01-26	2021-01	0.480
复混肥料	颗粒	20.00%	6.00%	24.00%	无氯	0.00%	皖农肥(2016)准字4297号	2016-01-26	2021-01	0.500
复混肥料	颗粒	22.00%	5.00%	24.00%	无氯	0.00%	皖农肥(2016)准字4298号	2016-01-26	2021-01	0.510
复混肥料	颗粒	24.00%	10.00%	18.00%	低氯	0.00%	皖农肥(2016)准字4299号	2016-01-26	2021-01	0.520
有机-无机复混肥料	颗粒	10.00%	5.00%	5.00%	无氯	20.00%	皖农肥(2016)准字4300号	2016-01-26	2021-01	0.200
复混肥料	颗粒	25.00%	10.00%	10.00%	中氯	0.00%	皖农肥(2016)准字4301号	2016-02-5	2021-02	0.450
复混肥料	颗粒	26.00%	13.00%	8.00%	中氯	0.00%	皖农肥(2016)准字4302号	2016-02-5	2021-02	0.470

图 3-2 各肥料产品的氮、磷、钾养分百分比之和  
由于数据过多，图 1-2 中仅取了 50 行数据。

四、肥料产品的数据分析

4.1 任务 2.1

百分比区间	下临界值	分组
0%-7.3%	0	1
7.3%-14.6%	7.30%	2
14.6%-21.9%	14.60%	3
21.9%-29.2%	21.90%	4
29.2%-36.5%	29.20%	5
36.5%-43.8%	36.50%	6
43.8%-51.1%	43.80%	7
51.1%-58.4%	51.10%	8
58.4%-65.7%	58.40%	9
65.7%-73%	65.70%	10

图 4-1 分组区间和分组标签

为了更直观地分析出复混肥料产品的分布特点，利用 TipDM-BI 数据分析和可视化平台绘制出了复混肥料产品的登记数量可视化图形（包括了复混肥料产品的分布直方图），如图 4-3 所示。

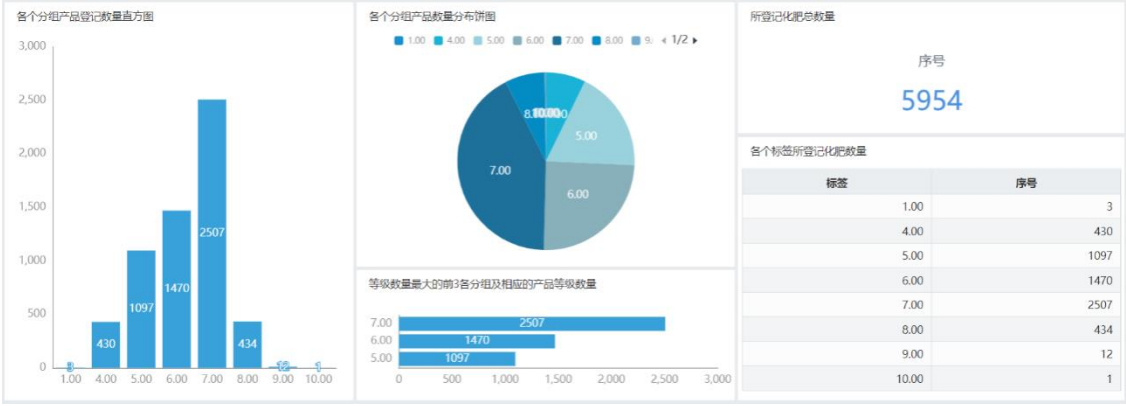


图 4-3 复混肥料产品的登记数量可视化图形

观察该图可看出按登记数量从大到小列出登记数量最大的前三个分組分别为标签 7、6 和 5，且相对应的产品登记数量分别为 2098、1470 和 1154。则说明总无机养分百分比在 28.9%到 50.4%范围内的复混肥料产品，其登记数量更高。通过计算可知，该范围的产品占了所有复混肥料产品中的 79%。

排名	一	二	三
分組标签	7	6	5
产品登记数量	2098	1470	1154

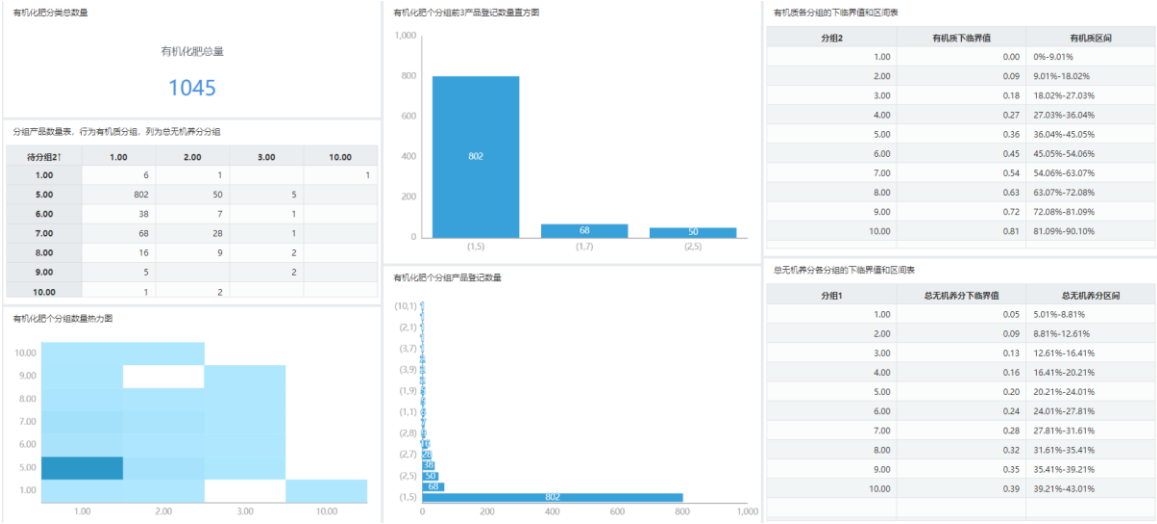


图 4-6 有机肥料产品的登记数量可视化图形