

肥料登记数据分析报告

目录

一、 分析目的与内容.....	2
1.1 背景.....	2
1.2 任务目标.....	2
1.3 任务分析.....	2
1.3.1 任务一.....	2
1.3.2 任务二.....	3
1.3.3 任务三.....	3
1.3.4 任务四.....	3
二、 数据的预处理.....	4
2.1 产品通用名规范化处理.....	4
2.2 计算总无机养分百分比.....	4
三、 肥料产品的数据分析.....	4
3.1 筛选复混肥料的产品并细分.....	4
3.1.3 结果.....	4
3.2 筛选有机肥料的产品并细分.....	5
3.2.1 处理思路.....	5
3.2.2 过程.....	5
3.1.3 结果.....	5
3.3 筛选复混肥料的产品并聚类.....	5
3.3.1 处理思路.....	5
3.3.2 过程.....	5
四、 肥料产品的多维度对比分析.....	6
4.1 日期维度对复混肥料登记数据进行对比.....	6
4.1.1 处理思路.....	6

一、分析目的与内容

1.1 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，政府网站上可获得相关数据。

1.2 任务目标

根据附件 1~附件 4 中提供的数据，本文依次解决如下任务：

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。

1.3 任务分析

1.3.1 任务一

✧ 任务 1.1：附件 1 的产品通用名称存在不规范的情况，进行规范化处理。

任务 1.1 要求我们对附件 1 的产品通用名进行规范化处理。

✧ 任务 1.2：计算附件 1 中各肥料产品的总无机养分百分比。

任务 1.2 要求我们计算附件 1 中各肥料产品的氮、磷、钾养分百分比之和，由于肥料中氮肥成分以总氮的质量来计算含量，磷肥成分按磷元素的量折算成五氧化二磷(P_2O_5)的质量来计算含量，钾肥成分按钾元素的量折算成氧化钾(K_2O)的质量来计算含量，故考虑将各肥料产品对应的“总氮百分比”、“ P_2O_5 百分比”和“ K_2O 百分比”。

1.3.2 任务二

✧ 任务 2.1: 从附件 2 中筛选出复混肥料的产品, 根据总无机养分百分比对肥料产品进行细分, 并绘制直方图进行分析。

任务 2.1 要求我们根据总无机养分的百分比对复混肥料产品进行细分。挑选出产品登记数量最大的三个。

✧ 任务 2.2: 从附件 2 中筛选出有机肥料的产品, 根据总无机养分百分比和有机质百分比进行细分, 并绘制有机肥料产品的分布热力图进行分析。

任务 2.2 要求我们根据总无机养分百分比和有机质百分比对有机肥料产品进行细分。绘制热力图并进行描述性统计分析, 最后根据上述求解找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, 使用聚类算法将其分为 4 类。根据聚类结果为每个产品打上聚类标签, 同时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 2.3 要求我们根据氮、磷、钾养分的百分比用聚类算法将复混肥料的产品分为分类。绘制肥料产品的三维散点图、散点图矩阵和聚类结果的雷达图, 并运用描述性统计分析来分析每个聚类的特征。

1.3.3 任务三

✧ 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, 并使用合适的图表对结果进行可视化。

任务 3.1 要求我们从时间维度对复混肥料的等级数据进行对比分析。绘制折线图, 结合描述性统计分析比较各组别不同年份产品登记数量的变化趋势。

✧ 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, 并从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。

任务 3.2 要求我们提取 2021 年 9 月 30 日仍有效的有机肥料产品, 并从省份维度对有机肥料的等级数据进行对比分析。再根据“正式登记证号”中的省份简称来进行地域信息的筛选, 湖北简称为鄂, 广西简称为桂, 筛选出仍在有效产品中广西和湖北产品登记数量在前 5 的组别, 并利用描述性统计分析分析两个省份上述组别的分布差异。

✧ 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。

任务 3.3 要求我们提取产品登记数量大于 10 的肥料企业, 并以各企业用到的原料作为特征, 计算杰卡德相似系数矩阵。

1.3.4 任务四

✧ 任务 4.1: 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 要求我们从技术指标中提取出氮、磷、钾养分和有机质的百分比,

✧ 任务 4.2 : 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。

任务 4.2 要求我们从附件 4 原料与百分比中提取各种原料的名称及其百分比。

二、数据的预处理

2.1 产品通用名规范化处理

2.2 计算总无机养分百分比

三、肥料产品的数据分析

3.1 筛选复混肥料的产品并细分

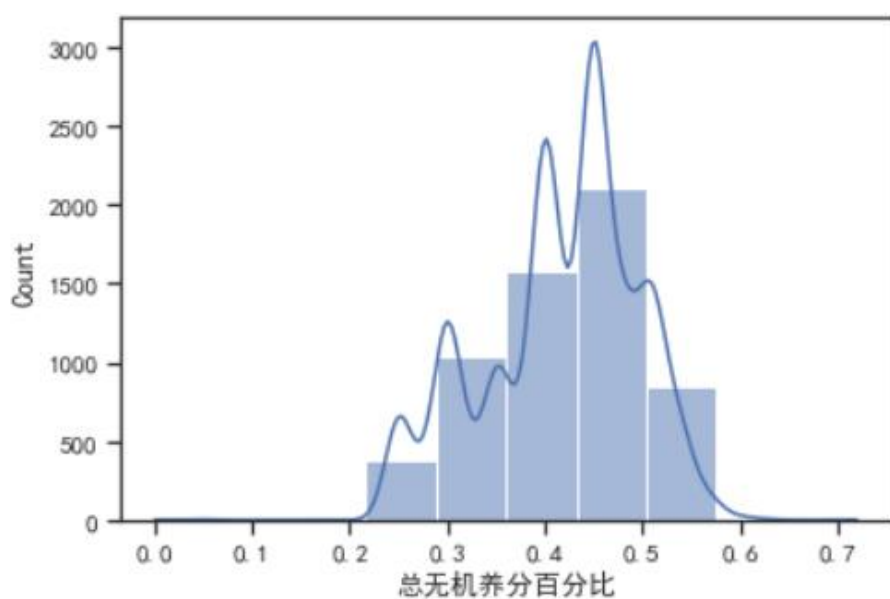


图 3- 1 总无机养分百分比直方图

从分布直方图中可以看出，复混肥料产品的总无机养分百分比的分布大致是一个左偏分布。总无机养分百分比较低的组（分组 0 - 0.072，分组 0.072 - 0.144，分组 0.144 - 0.216）和无机养分百分比较高的组（分组 0.576 - 0.648，分组 0.648 - 0.72）数量都特别少，大多数的复混肥料产品的总无机养分百分比都集中在 0.4-0.5 的范围内。从分组（0.216 - 0.288）开始，一直到分组（0.504 - 0.576），

表 3- 1 登记数量最大的前 3 个分组及相应的产品登记数量

3.1.3 结果

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

3.2 筛选有机肥料的产品并细分

3.2.1 处理思路

3.2.2 过程

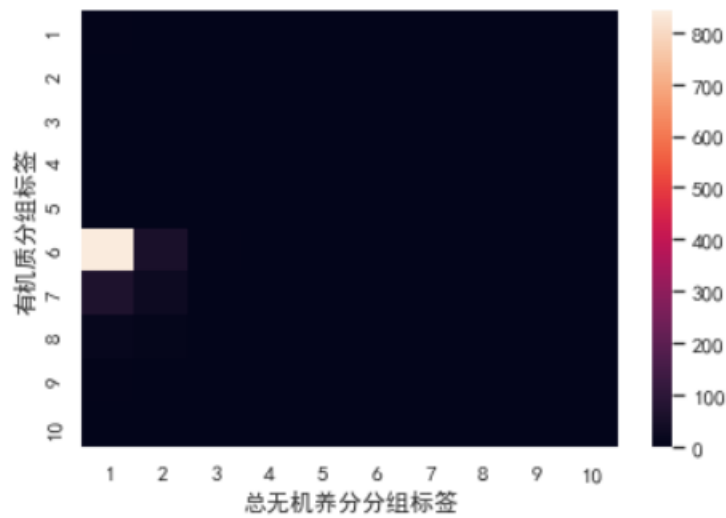


图 3-2 热力图

3.1.3 结果

3.3 筛选复混肥料的产品并聚类

3.3.1 处理思路

3.3.2 过程

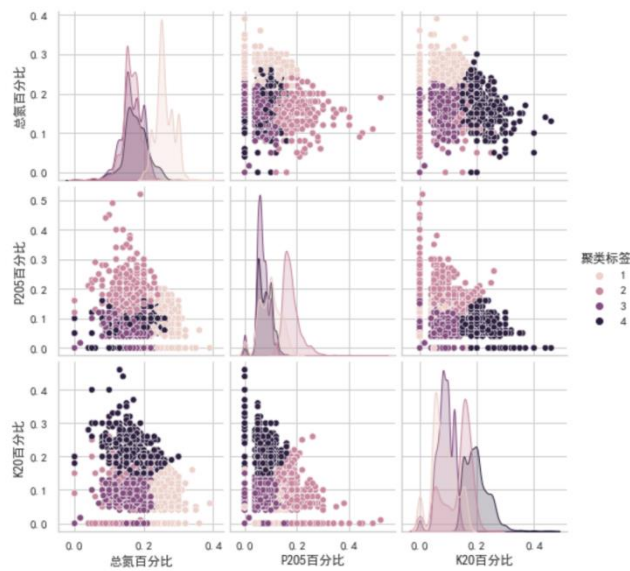


图 3-3 散点图矩阵

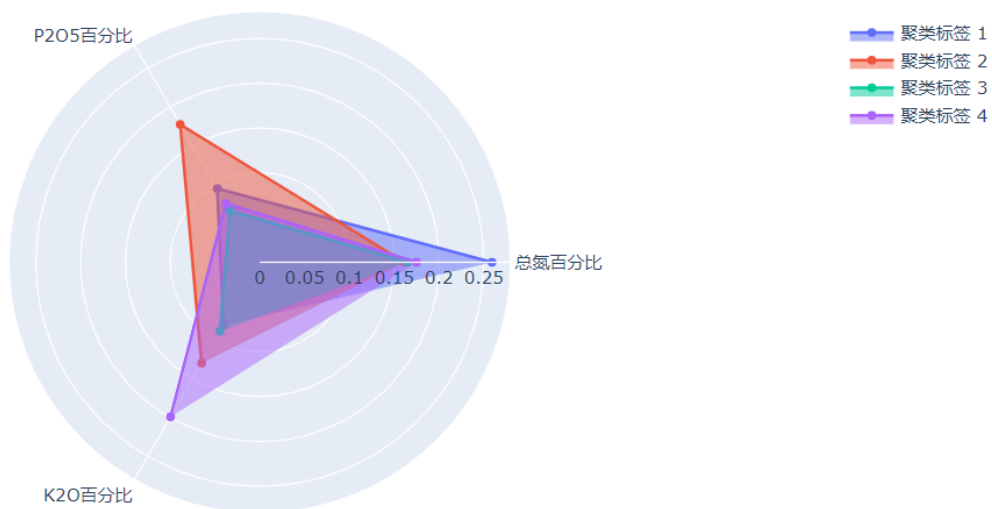


图 3-4 雷达图

四、肥料产品的多维度对比分析

4.1 日期维度对复混肥料登记数据进行对比

4.1.1 处理思路

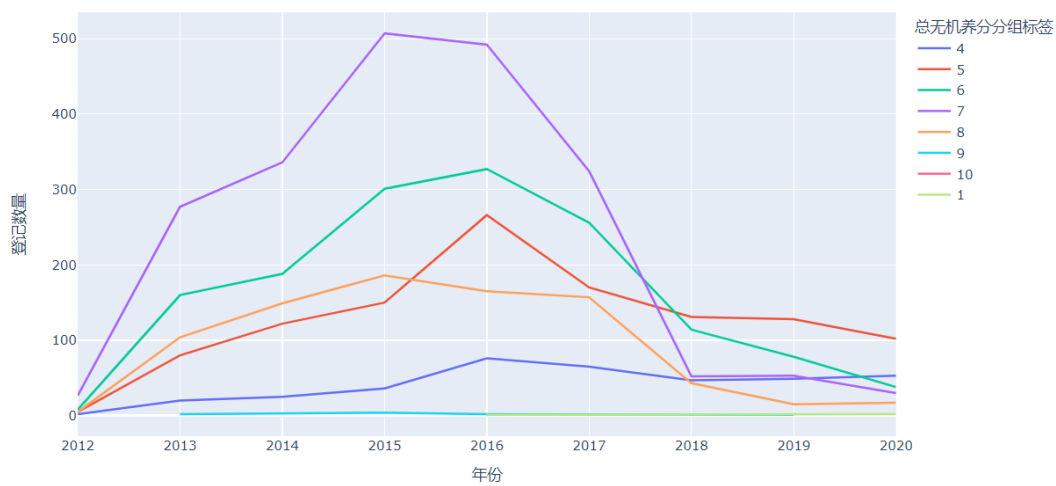


图 4-1 折线图