

# 目录

- 1.问题重述 .....2
  - 1.1 问题背景 .....2
  - 1.2 任务目标 .....2
- 2.任务 1 数据的预处理.....2
  - 2.1 任务 1.1 规范产品通用名称 .....2
  - 2.2 任务 1.2 各肥料产品总无机养分百分比 .....2
- 3.任务 2 肥料产品的数据分析 .....3
  - 3.1 任务 2.1 复混肥料的产品总无机养分百分比的分布特点 .....3
  - 3.2 任务 2.2 有机肥料的产品总无机养分百分比的分布特点 .....4
  - 3.3 任务 2.3 复混肥料产品的聚类分析 .....5
- 4.任务 3 肥料产品的多维度对比分析 .....5
  - 4.1 任务 3.1 复混肥料产品登记数量的变化趋势 .....5
  - 4.2 任务 3.2 筛选有效产品 .....5
  - 4.3 任务 3.3 企业的产品原料相似度探究 .....5

# 1.问题重述

## 1.1 问题背景

肥料产业服务于农业生产，国家经济水平、农业生产发展尤其生产体制变革又对肥料产业的发展、产品创新等产生深刻影响，对我国多方位的发展起到了重要作用。为了加强肥料管理，保护生态环境，保障人畜安全，促进农业生产，我国制定了《肥料登记管理办法》。所有肥料的生产销售都需要遵循该办法，并且依法在农业行政管理部门进行登记。各级人民政府都设置了相关的部门进行销售肥料的登记工作。因此，利用数据分析对肥料产品进行各项指标的细化分类十分必要。

## 1.2 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 1 数据的预处理

## 2.1 任务 1.1 规范产品通用名称

## 2.2 任务 1.2 各肥料产品总无机养分百分比

由题意可得，本题需要计算出肥料产品的氮、磷、钾养分百分比之和。利用 Pandas 中的 dataframe 创建新列“总无机养分百分比”，储存表格中“总氮百分比”“P2O5 百分比”“K2O 百分比”三列分别相加之和，并保留了 3 位小数。

### 3.任务 2 肥料产品的数据分析

#### 3.1 任务 2.1 复混肥料的产品总无机养分百分比的分布特点

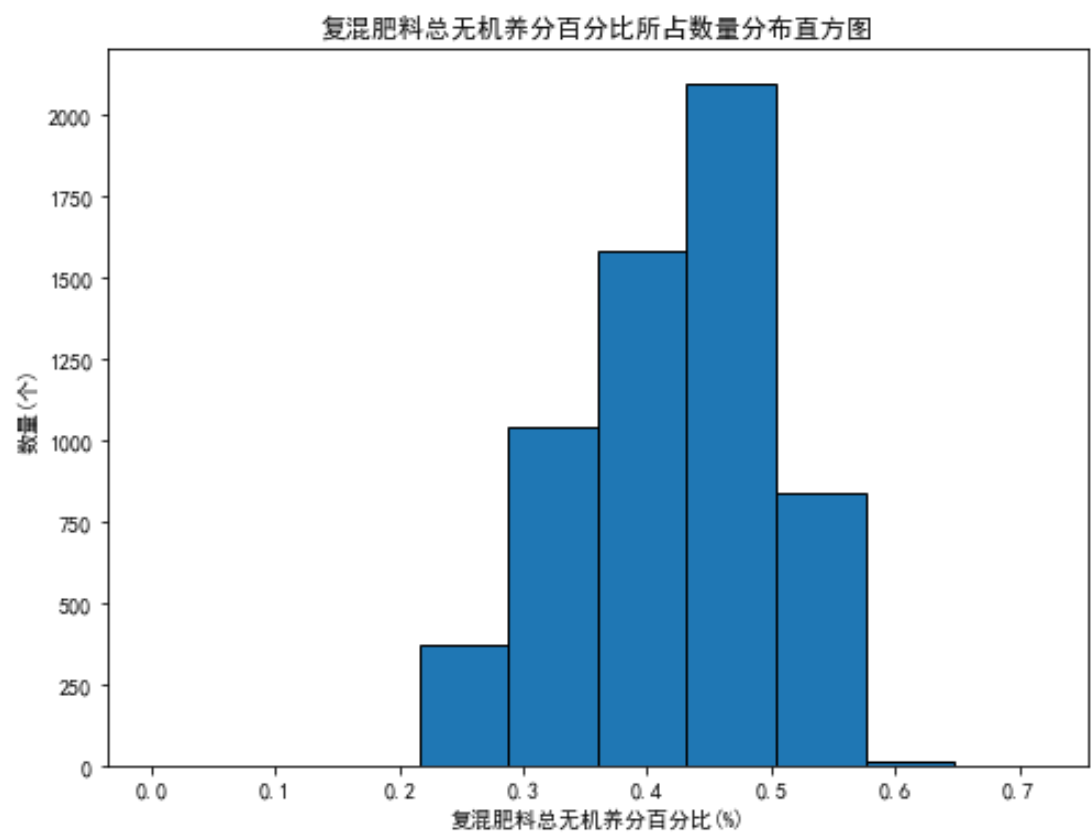


图 1-复混肥料产品总无机养分百分比的数量分布直方图

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

图 2-登记数量最大的前 3 个分组及相应的产品登记数量表

3.2 任务 2.2 有机肥料的产品总无机养分百分比的分布特点

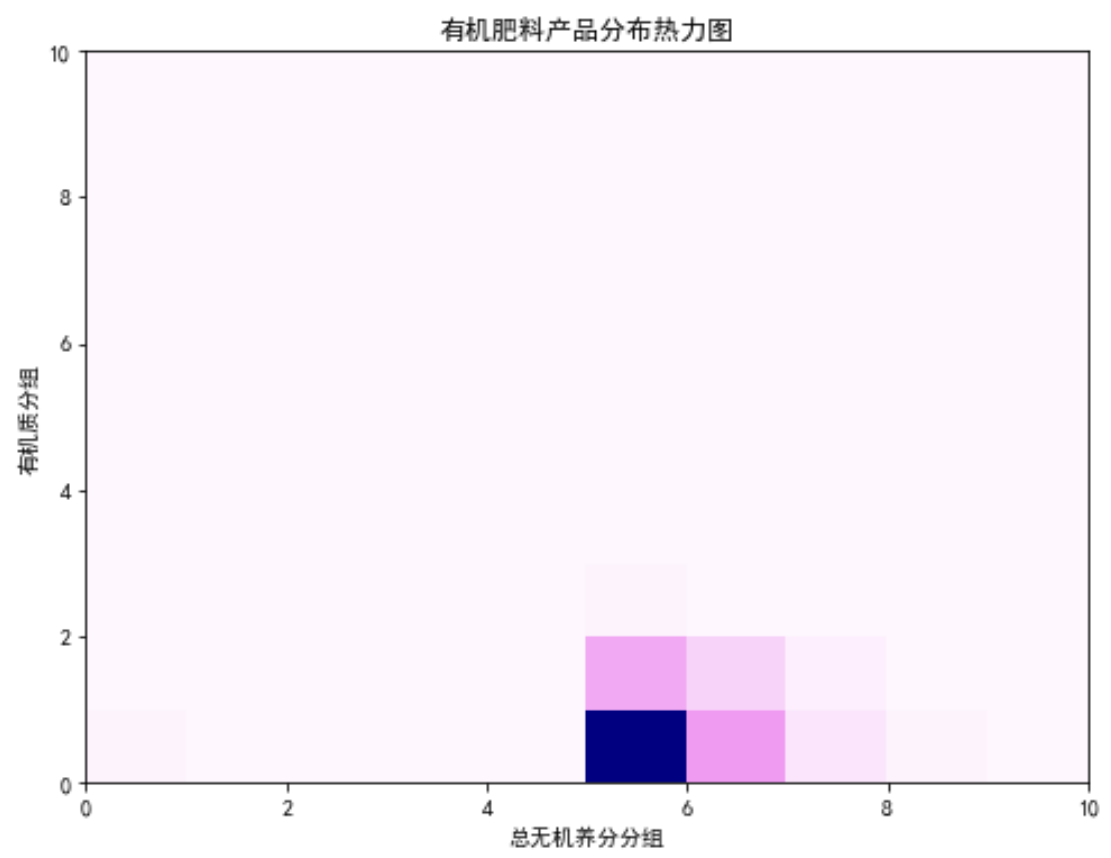


图 3-有机肥料产品总无机养分百分比的数量分布热力图

### 3.3 任务 2.3 复混肥料产品的聚类分析

## 4.任务 3 肥料产品的多维度对比分析

### 4.1 任务 3.1 复混肥料产品登记数量的变化趋势

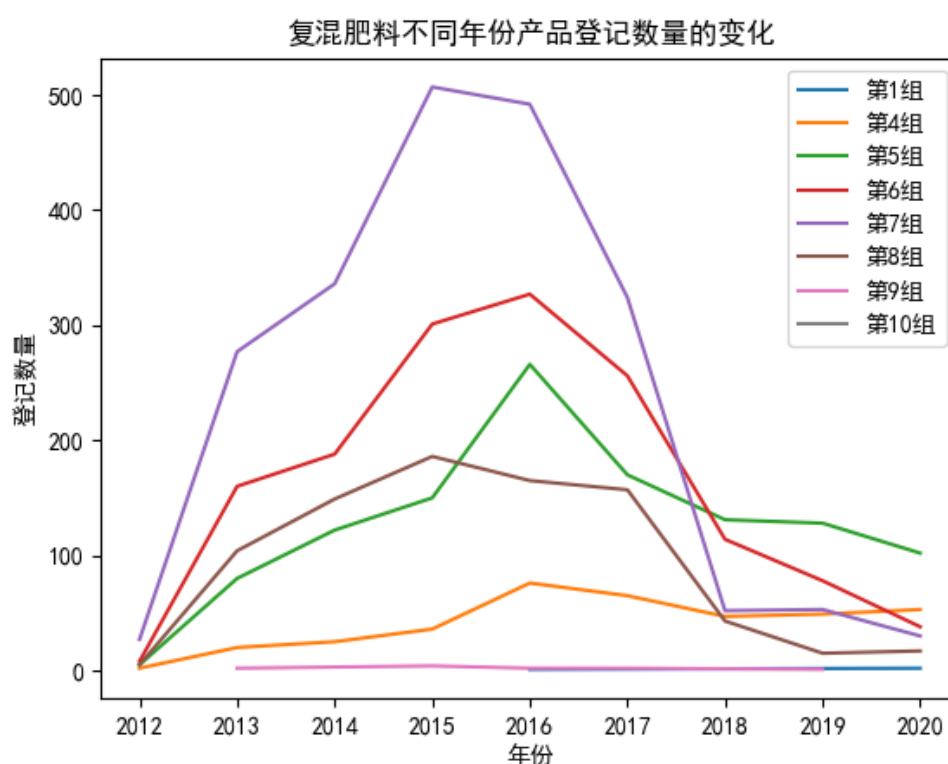


图 6-各组别复混肥料不同年份产品登记数量的折线图

### 4.2 任务 3.2 筛选有效产品

### 4.3 任务 3.3 企业的产品原料相似度探究

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。