

## 目录

摘要 .....	1
1.任务一 数据检测与预处理 .....	1
1.1 缺失值处理 .....	1
1.2 异常值处理 .....	2
1.3 重复值处理 .....	2
1.3.1 重复值情况说明 .....	2
1.3.2 重复值处理过程描述 .....	2
1.3 数据处理——修改字段数据 .....	2
2.任务二 肥料产品的数据分析 .....	2
2.1 复混肥料产品分析.....	2
2.1.2 复混肥料分组特点分析.....	2
2.2 有机肥料产品分析.....	3
3. 任务三 肥料产品的多维度对比分析.....	4
3.1 复混肥料中各组别不同年份产品分析.....	4

## 摘要

肥料是农业生产中一种重要的生产资料。在农业肥料是指提供一种或一种以上植物必需的营养元素，改善土壤性质、提高土壤肥力水平的一类物质，是农业生产的物质基础之一。因此，可以对各种类型肥料生产量、购买量做不同数据分析。

针对任务一，对所给数据集的脏数据进行缺失值、异常值、重复值等方面的处理。针对任务二，在任务一的基础上，对数据集进行进一步的处理，为任务二的数据可视化奠定基础。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，并通过降序排序获得前三甲的登记数量。根据分组情况绘制有机肥料产品的分布热力图，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。根据 K-Means 函数聚类算法，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。

针对任务三，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。取 2021 年 9 月 30 日仍有效的有机肥料产品，选出广西和湖北产品登记数量在前 5 的组别，分析广西和湖北的分布差异。提取产品登记数量大于 10 的肥料企业。将这些企业所用到的原料集合（发酵菌剂除外）作为特征，使用计算公式原料直接的杰卡德相似系数矩阵。

针对任务四，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，分析肥料产品原料含量。

## 1.任务一 数据检测与预处理

### 1.1 缺失值处理

### 1.1.1 缺失值情况说明

### 1.1.2 缺失值处理过程描述

## 1.2 异常值处理

### 1.2.1 异常值情况说明

在本次数据分析过程中，针对已经进行缺失值处理后的数据进行异常值处理。这里的异常值仅指日期和数据不规范。

### 1.2.2 异常值处理过程描述

图 1.2.2.3 处理前数据示意图      图 1.2.2.4 处理后数据示意图

## 1.3 重复值处理

### 1.3.1 重复值情况说明

如果在数据集中，所有的字段数据均一致，则对进行重复值删除操作。

### 1.3.2 重复值处理过程描述

## 1.3 数据处理——修改字段数据

## 2.任务二 肥料产品的数据分析

### 2.1 复混肥料产品分析

#### 2.1.1 复混肥料分组描述

首先将产品通用名称字段中的复混肥料的所有数据选出，将总无机养分百分比的最高值和最小值，定义一个函数计算等距且分组，计算公式为**等距区间=总无机养分百分比最大值-总无机养分百分比最小值/组数**，循环数据表中的总无机养分百分比匹配等距区间的组数名。

#### 2.1.2 复混肥料分组特点分析

根据图 2.1.2.1 可知，在所有产品中，总无机养分占比含量越高，登记数量即购买数量越少。其中，湖北生产复混肥料数量最多且登记数量最大。有机质百分比基本为 0，无机养分接近占总养分 50%。

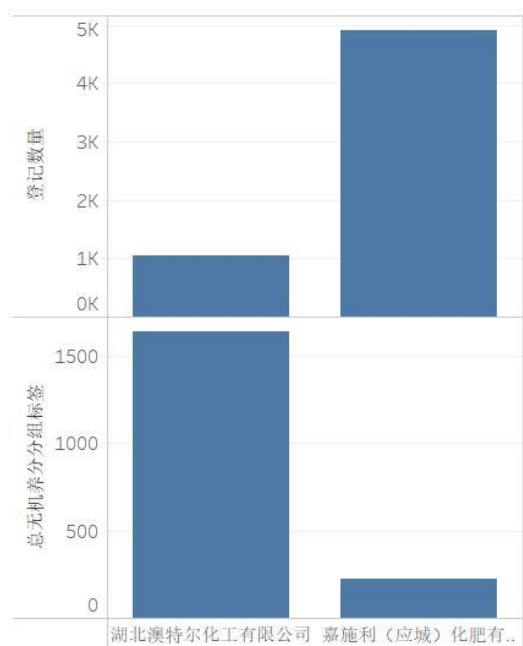


图 2.1.2.1 复混肥料原料成分分布图

#### 总无机养分分组标签

7 2098

6 1584

5 1040

排名	一	二	三
分组标签	7	6	5
产品登记数量	2098	1584	1040

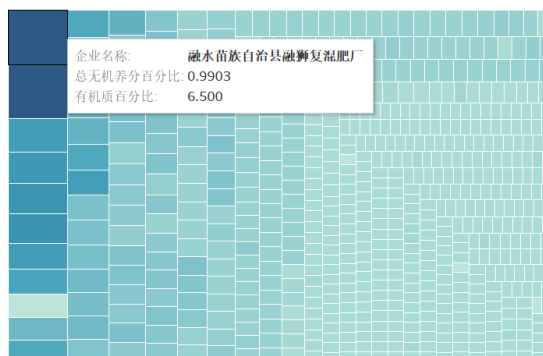
## 2.2 有机肥料产品分析

### 2.2.1 有机肥料分组描述

处理方法与 2.1.1 相似，不同之处在于将总无机养分百分和有机百分比均计算出等距。关键代码如下：

#### 2.2.2 有机肥料分组特点分析

从热力图中可知融水苗族自治县融狮复混肥厂占比最高，其总无机养分百分比 0.9903%，有机质百分比为 6.500%。



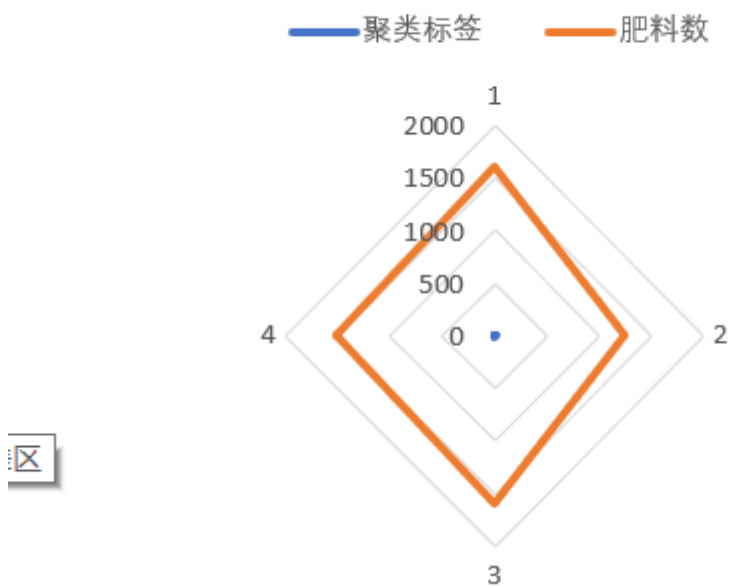
## 2.3 复混肥料分析——基于聚类算法

### 2.3.1 使用聚类算法分类过程描述

使用 K-Means 函数 `KMeans(n_clusters = 4, random_state = 50)` 将其分为 4 类，最后把结果并加入表中。

### 2.3.2 复混肥料分析

制聚类结果的雷达图



## 3. 任务三 肥料产品的多维度对比分析

### 3.1 复混肥料中各组别不同年份产品分析