

# 肥料登记数据分析

## 摘要

肥料是指用于提供、保持或改善植物营养和土壤物理、化学性能以及生物活性，能提高农产品产量，或改善农产品品质，或增强植物抗逆性的有机、无机、微生物及其混合物料。有些肥料是肥料和农药的混合物或是农民自制的自用有机肥料。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。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，对质量不合格的产品，要限期改进。对质量连续不合格的产品，肥料登记证有效期满后不予续展。我们可以通过数据筛选和数据勘察来筛选出不合格的肥料。

关键词：k-mean 聚类算法；杰卡德相似系数矩阵；数据分析；数据透视

## 目录

1.问题分析 .....	3
1.1 问题一 .....	3
1. 对肥料登记数据进行预处理。 .....	3
1.2 问题二 .....	3
1.3 问题三 .....	3
1.4 问题四 .....	4
1.5 .....	4
2.数据预处理 .....	7
2.1 产品名称处理 .....	7
2.1.1 产品名称的数据分析 .....	7
2.1.2 产品名称的规范化处理 .....	8
2.2 无机养分计算 .....	8
2.2.1 无机养分的数据处理 .....	8
2.2.2 总无机养分的百分比计算处理结果 .....	8
3.肥料产品的数据分析 .....	8
3.1 复混肥料的数据分析 .....	8
3.1.1 复混肥料的分布特点分析 .....	8
3.1.2 数据处理 .....	9
3.2 有机肥料的数据分析 .....	9
3.2.1 数据处理 .....	9
3.2.2 有机肥料的分布分析 .....	10
3.3 复混肥料养分分类 .....	10

3.3.1 复混肥料的数据处理.....	10
3.3.2 复混肥料的数据分布 .....	10
3.3.3 分析聚类特征.....	10
4.肥料产品的多维度对比分析 .....	11
4.1 复混肥料的数据提取变化趋势 .....	11
4.1.1 数据可视化 .....	11
4.2 有机肥料的提取.....	12
4.2.1 有机肥料的数据提取.....	12
4.2.2 有机肥料的分布差异.....	12
4.3 肥料企业登记数据 .....	12
4.3.1 数据处理.....	12
4.3.2 杰卡德相似系数矩阵.....	12
5.1 肥料成分数据分析 .....	12
5.1.1 肥料氮、磷、钾养分和有机质的百分比及含氯的程度 .....	13
5.2 原料名称及百分比数据提取.....	13
5.2.1 原料名称及百分比提取 .....	13

## 1. 问题分析

### 1.1 问题一

1. 对肥料登记数据进行预处理。

### 1.2 问题二

2. 根据养分的百分比对肥料产品进行细分。

### 1.3 问题三

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。

1.4 问题四

4. 对非结构化数据进行结构化处理。

1.5

5. 各数据表结构展示

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2925 entries, 0 to 2924
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   序号            2925 non-null   int64
1   企业名称        2925 non-null   object
2   产品通用名称    2925 non-null   object
3   产品形态        2925 non-null   object
4   总氮百分比      2925 non-null   float64
5   P205百分比     2925 non-null   float64
6   K20百分比      2925 non-null   float64
7   含氯情况        2925 non-null   object
8   有机质百分比    2925 non-null   float64
9   正式登记证号    2925 non-null   object
10  发证日期        2925 non-null   object
11  有效期          2925 non-null   object
dtypes: float64(4), int64(1), object(7)
memory usage: 274.3+ KB
```

图一 附件 1 数据明细图

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7619 entries, 0 to 7618
Data columns (total 15 columns):
 #   Column                Non-Null Count  Dtype  
---  -
 0   序号                  7619 non-null  int64  
 1   企业名称              7619 non-null  object  
 2   产品通用名称          7619 non-null  object  
 3   产品形态              7618 non-null  object  
 4   总氮百分比            7619 non-null  float64 
 5   P205百分比           7619 non-null  float64 
 6   K20百分比            7619 non-null  float64 
 7   含氯情况              7619 non-null  object  
 8   有机质百分比          7619 non-null  float64 
 9   正式登记证号          7619 non-null  object  
10   发证日期              7619 non-null  object  
11   有效期                7619 non-null  object  
12   产品商品名称          1534 non-null  object  
13   适用作物              1 non-null     object  
14   总无机养分百分比      7619 non-null  float64 
dtypes: float64(5), int64(1), object(9)
memory usage: 893.0+ KB

```

图二 附件 2 数据明细图

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 550 entries, 0 to 549
Columns: 296 entries, 序号 to 55磷酸一铵
dtypes: float64(294), int64(1), object(1)
memory usage: 1.2+ MB

```

图三 附件 3 数据明细图

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  -
0   序号         200 non-null    int64
1   产品通用名称  200 non-null    object
2   技术指标     200 non-null    object
3   原料与占比   200 non-null    object
dtypes: int64(1), object(3)
memory usage: 6.4+ KB
```

图四 附件 4 数据明细图

序号	0	
企业名称	0	
产品通用名称	0	
产品形态	1	
总氮百分比	0	
P205百分比	0	
K20百分比	0	
含氯情况	0	
有机质百分比	0	
正式登记证号	0	
发证日期	0	
有效期	0	
产品商品名称	6085	
适用作物	7618	
总无机养分百分比		0
dtype: int64		

图 5 附件 2 空值缺失明细表

## 2. 数据预处理

### 2.1 产品名称处理

#### 2.1.1 产品名称的数据分析

### 2.1.2 产品名称的规范化处理

## 2.2 无机养分计算

### 2.2.1 无机养分的数据处理

### 2.2.2 总无机养分的百分比计算处理结果

## 3. 肥料产品的数据分析

### 3.1 复混肥料的数据分析

#### 3.1.1 复混肥料的分布特点分析

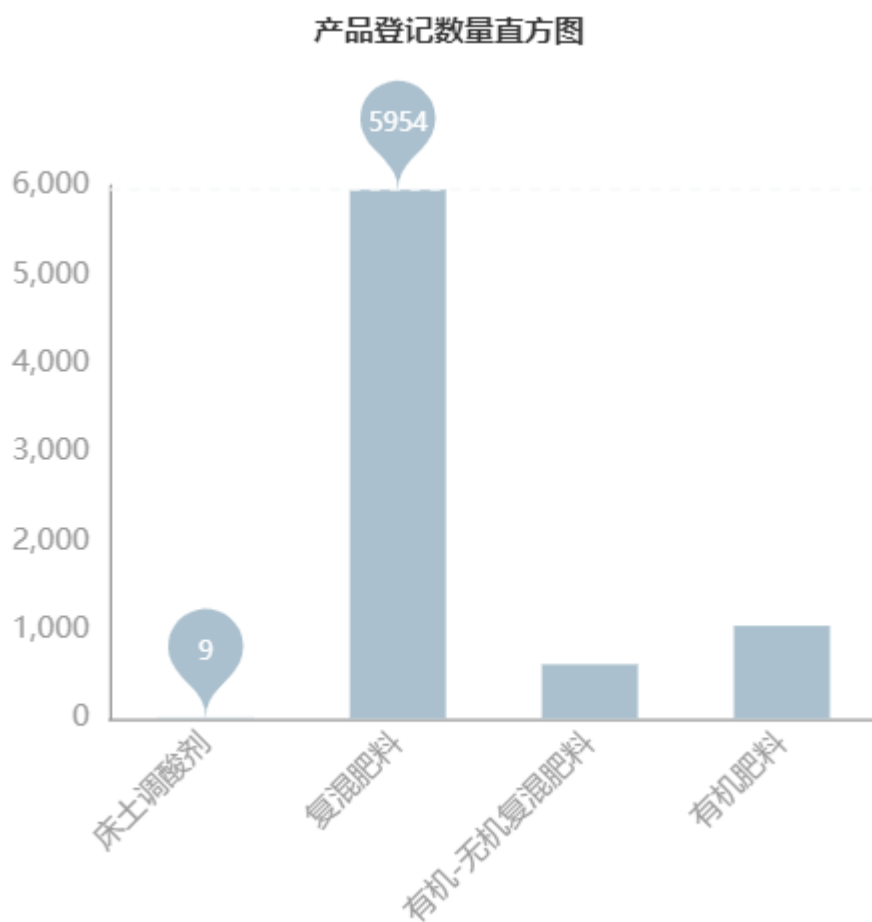




图 7 产品登记数量直方图

3.1.2 数据处理

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

图 8 登记数量最大前 3 分组

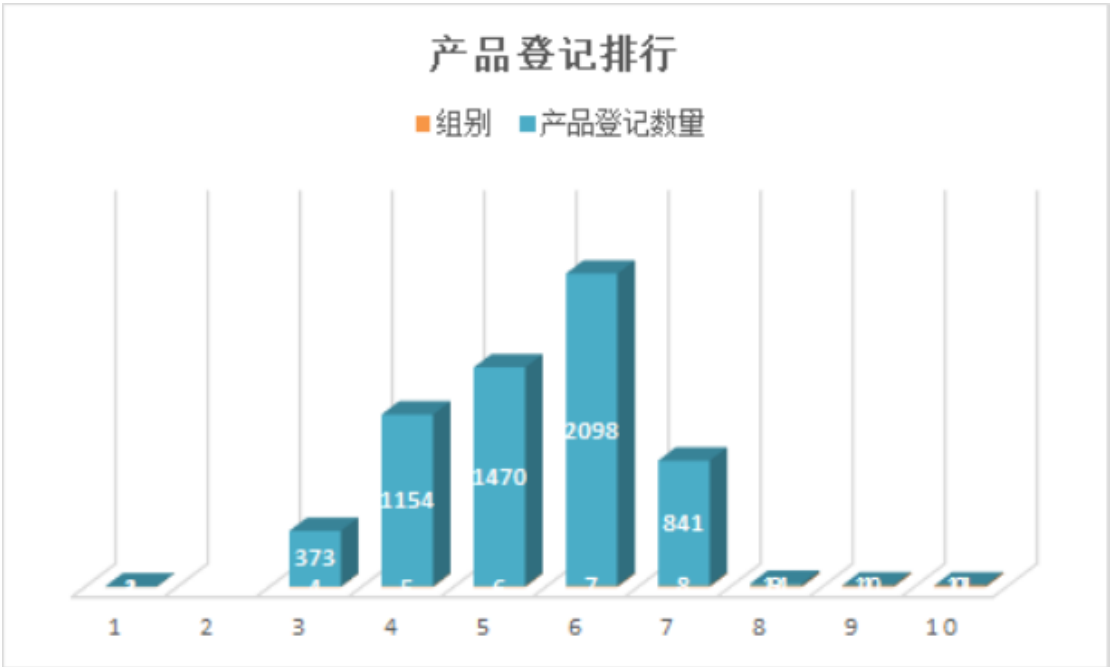


图 9 产品登记排行可视化

从图表中可以看出登记数量最大的前 3 个分组分别为第 6 组，第 5 组和第 4 组以及其相应的产品登记数量为 2098，1470 和 1154.

3.2 有机肥料的数据分析

3.2.1 数据处理

3. 2. 2 有机肥料的分布分析

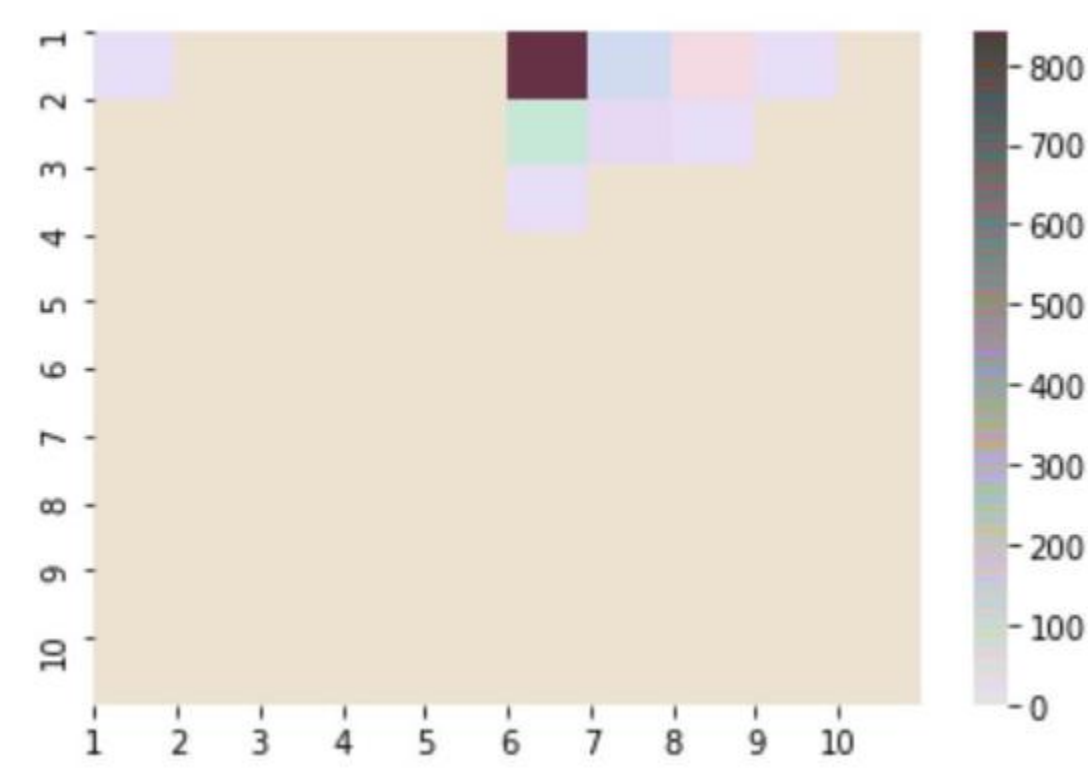


图 10 总无机养分的分组和有机质的热力图

3. 3 复混肥料养分分类

3. 3. 1 复混肥料的数据处理

3. 3. 2 复混肥料的数据分布

3. 3. 3 分析聚类特征

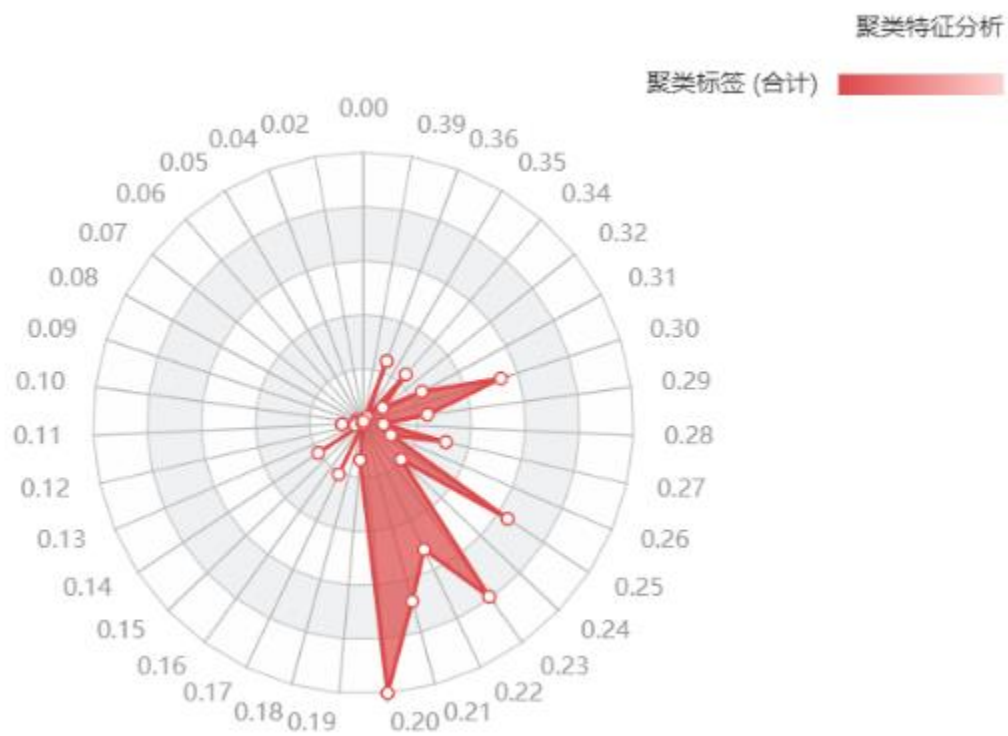


图 13 总无机养分占比与聚类雷达图

## 4. 肥料产品的多维度对比分析

### 4.1 复混肥料的数据提取变化趋势

#### 4.1.1 数据可视化

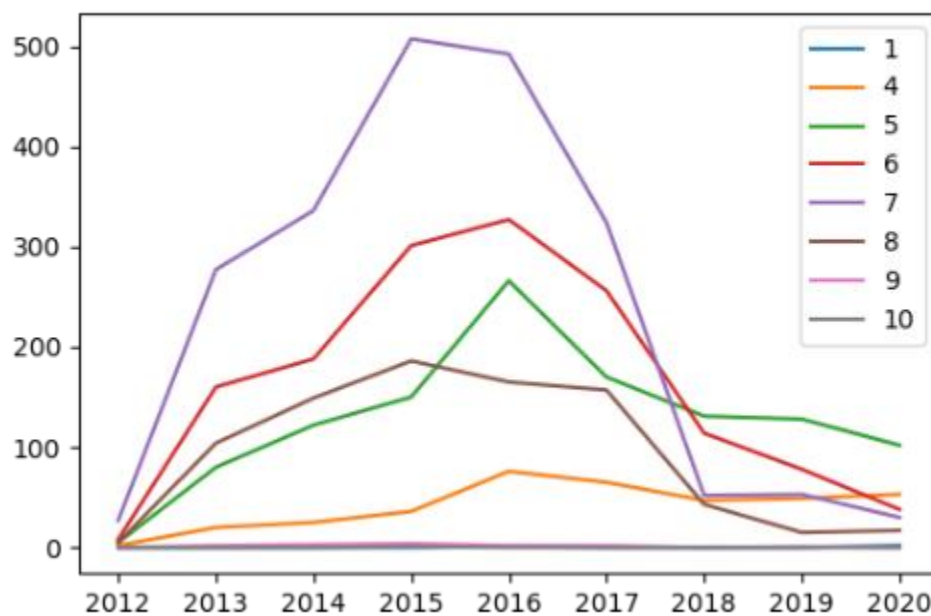


图 14 不同年份的产品登记数量

## 4. 2 有机肥料的提取

### 4. 2. 1 有机肥料的数据提取

### 4. 2. 2 有机肥料的分布差异

## 4. 3 肥料企业登记数据

### 4. 3. 1 数据处理

### 4. 3. 2 杰卡德相似系数矩阵

## 5. 1 肥料成分数据分析

5.1.1 肥料氮、磷、钾养分和有机质的百分比及含氯的程度

5.2 原料名称及百分比数据提取

5.2.1 原料名称及百分比提取