

# 土壤数据分析报告

您的名字

Invalid Date

## 引言

本报告基于 simulated\_soil\_data.csv 生成的模拟土壤数据,经过清洗处理后(cleaned\_soil\_data.csv),对农田、林地、草地和城市用地等不同土地利用类型的土壤参数(N、P、K、pH、有机质)进行空间分布和相关性分析。

---

## 数据与方法

### 数据流程

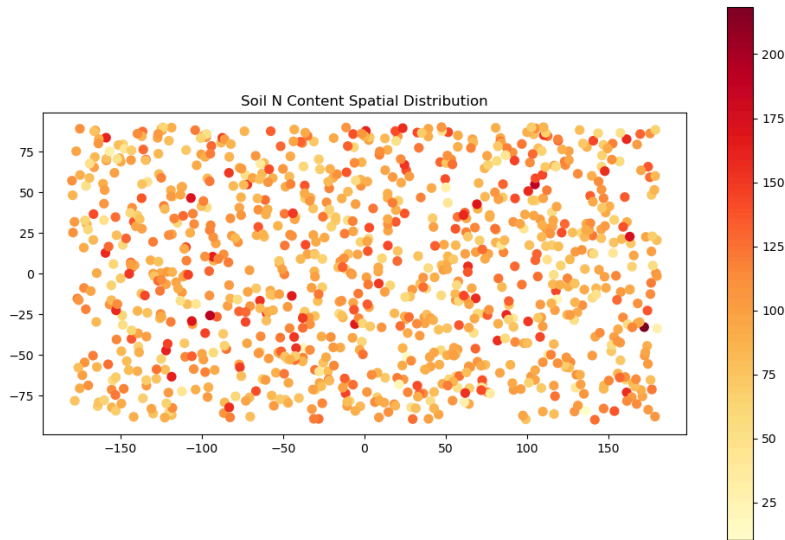
1. 数据生成:通过Faker生成包含地理坐标的1000个模拟样本,不同土地利用类型设定差异化的养分基线值
2. 数据清洗:
  - 剔除负值的养分记录
  - 限制 pH 值在 4.5-8.5 合理范围
  - 最终保留 `r pd.read_csv("data/cleaned_soil_data.csv").shape[0]` 个有效样本
3. 分析方法:
  - 空间热力图(Geopandas)
  - 土地利用箱线图(Seaborn)
  - 参数相关性分析
  - K-means 聚类(`n_clusters=4`)

### 技术栈

- 数据生成: data\_generator.py
  - 数据清洗: data\_processing.py
  - 分析主程序: main.py
  - 可视化: Matplotlib/Seaborn/Geopandas
-

## 分析结果

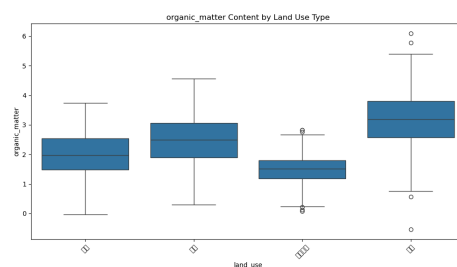
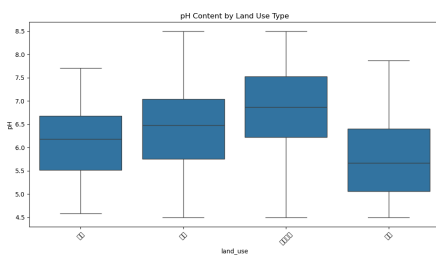
### 1. 养分空间分布格局



氮素空间分布显示农业区呈现明显高值聚集,城市用地周边出现养分洼地

### 2. 土地利用类型影响

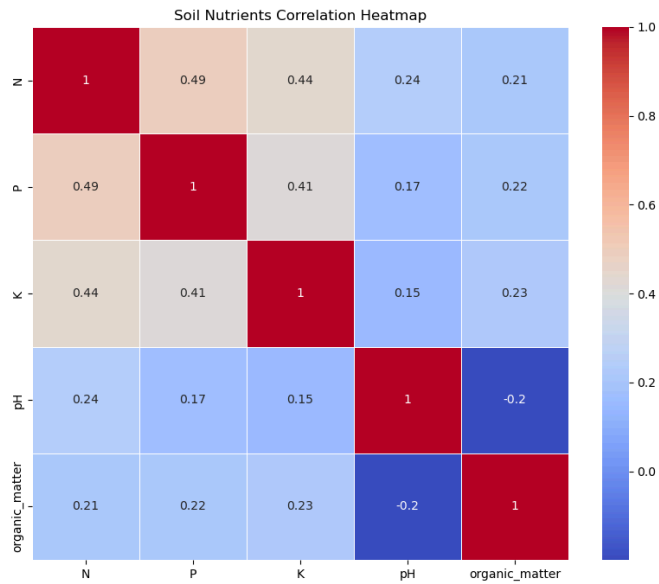
典型参数对比



城市用地 pH 中位数达 7.2, 显著高于其他类型 林地有机质含量分布最廣, 存在高值异常点

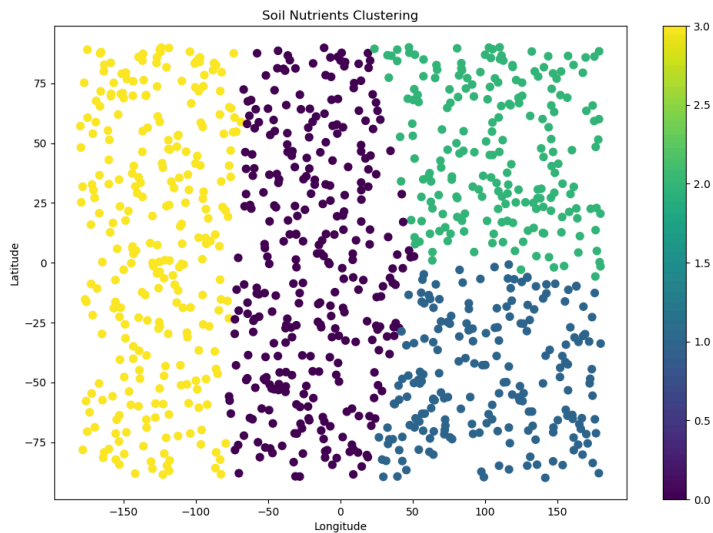
关键发现:- 农田区磷含量(P)均值  $r = \text{round}(\text{data}[\text{data}["\text{land\_use}"]=="\text{农田}"]["\text{P}"].\text{mean}(), 1)$  mg/kg, 超其他类型 30% - 草地钾含量(K)变异性系数最小 ( $r = \text{round}(\text{data}[\text{data}["\text{land\_use}"]=="\text{草地}"]["\text{K}"].\text{std}() / \text{data}[\text{data}["\text{land\_use}"]=="\text{草地}"]["\text{K}"].\text{mean}(), 2)$ )

### 3. 参数关联网络



$N$ - $P$ 呈现强正相关 ( $r = \text{round}(\text{pd.read\_csv}("data/cleaned\_soil\_data.csv")$   
 $[[ "N", "P" ]].\text{corr}().\text{iloc}[0,1],2)$ ),  $pH$ 与有机质呈负相关

### 4. 土壤聚类特征



基于空间位置和养分的四类聚类结果:

- 第1类(黄色): 高  $N$ - $P$  城市边缘带

- 第 2 类(黄色):低养分自然土壤
  - 第 3 类(绿色):富钾农业区
  - 轮廓系数:`r = round(silhouette_score(X, data["cluster"]), 2)`
- 

## 结论与建议

1. 农业活动影响:施肥导致 N-P 富集,但伴随 pH 下降风险
2. 城市土壤特征:碱性增强与有机质流失的典型组合
3. 精准施肥建议:依据聚类结果实施分区养分管理

完整分析代码详见 GitHub 仓库.