

## 数据清理

### 1. 缺失值处理

- 1) 产生原因：测量设备、人为等因素导致的缺失/空白
- 2) 处理方法：删除数据或填充数据

### 2. 重复数据处理

- 1) 产生原因：重复记录、系统/操作问题等
- 2) 处理方法：通过确认后删除不必要的重复数据

### 3. 案例中数据清理的具体操作

- 1) 使用 `drop`、`dropna` 删除缺失值
- 2) 使用 `fillna` 填充缺失值
- 3) 使用 `duplicated` 和 `sum` 检查数据的重复情况
- 4) 使用 `drop_duplicates` 删除数据中的重复行