

第五课 数据分析工具Pandas基础

第12节 数据清洗--处理缺失数据

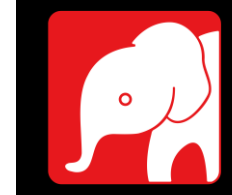
主 讲: Robin

数据清洗

- 是数据分析**关键**的一步，直接影响之后的处理工作
- 明确的问题：
 - 数据**需要**修改吗？
 - 有**什么**需要修改的吗？
 - 数据应该**怎么调整**才能适用于接下来的分析和挖掘？
- 是一个**迭代**的过程，实际项目中可能需要不止一次地执行这些清洗操作
- 操作：
 - 缺失数据
 - 重复数据
 - 无效数据

处理缺失数据

- 判断是否存在缺失值, `ser_obj.isnull()`, `df_obj.isnull()`
 - 可以结合 **any()** 判断行/列中是否存在缺失值
- `dropna()`, 丢弃缺失数据, 注意 **inplace** 参数
- `fillna()`, 填充缺失数据
- `df.ffill()`, 按之前的数据填充
- `df.bfill()`, 按之后的数据填充
- 项目中使用 `ffill()` 或 `bfill()` 时, 注意数据的 **排列顺序**



只为遇见明天更优秀的你！