

- 总结
 - 基本步骤

总结

基本步骤

1. 数据导入并查看基本信息

```
data=pd.read_csv('数据集')
```

2. 数据规整，修复各类字段名（按需求进行：重命名、类型转换、处理特殊字符）

3. 去重

```
initial_count = len(data)
data.drop_duplicates(keep="first",inplace=True)
final_count = len(data)
print(f"去重完成: {initial_count} → {final_count} 条记录")
```

4. 补缺

```
# 检查缺失值
missing_values=data.isnull().sum()
print(missing_values)
# 缺失值处理，例如：
data['product_rating'] =
data['product_rating'].fillna(data['product_rating'].median())
data['purchased_last_month'] = data['purchased_last_month'].fillna(0)
```

5. 检查并导出数据集

```
data = pd.to_csv('导出数据集名')
```