

## 1. Issues to Check with Data Sources

- Missing values
- Duplicate entries
- Inconsistent categories
- Ambiguous units or measures
- Unclear variable definitions
- Date/time inconsistencies
- Data collection scope

## 2. Issues I Must Handle Myself

- Text encoding issues
- Sampling bias
- Non-random samples
- Data locked in PDFs or images
- Time comparability
- Granularity mismatch
- Small sample sizes
- Contextual knowledge

## 3. Issues Requiring Expert or Domain Knowledge

- Data credibility
- Collection methodology
- Too-perfect numbers
- Outliers
- Statistical misuse
- Hidden assumptions
- Ethical considerations

## 4. Issues Needing Programming or Technical Solutions

- Data aggregation errors
- Merging and joins
- File format conversions
- Scanned documents
- Automated cleaning scripts
- Version control

## 5. Visualization and Reporting Concerns

- Geometric vs. optical accuracy
- Chart junk
- Label clarity
- Color use
- Comparability across visuals
- Transparency about uncertainty

## 6. Personal Reminders

- Document cleaning process
- Keep raw data unchanged
- Check dataset vs. research question
- Ask for clarification when in doubt