

manipulating data

adam okulicz-kozaryn

`adam.okulicz.kozaryn@gmail.com`

this version: Wednesday 29th January, 2025 01:41

outline

intuition

outline

intuition

general idea, intuition

- data management is mostly about:
- manipulating eg `gen`, `map`, `[:]`—today!
- (and related understanding, des sta/vis
eg `describe`, `value_counts`—last week)
- (and `merge`—next week)
- pretty easy—no complicated code, no fancy things
 - but also lame, boring, and tedious
 - yet necessary!

basic coding rules

- simplicity, clarity: drop the clutter, be clean
 - drop everything not necessary
 - efficiency: have “tight” code:
 - as few lines as possible that do as much as possible
- be lazy/work smart not hard
 - do only necessary work, eg do ren vars BUT not all just key ones like couple and then play with these
 - copy from others and use AI, not >80%!
- transparency/readability (eg meaningful names for df, var, etc)
- more rules later