W2. Summary

2020年3月22日 星期日 下午4:59

1. create Series
   - list - data: type - array - index
   - dict - data: type - key - index; value
     - missing value - None / NaN
   - value, index) - list of values

2. Query a series
   - iloc / loc [label] - value
   - attri
   - work with data: each value - operation
     - program approach: iterate - invoke
     - ① vectorization: np. function (iterable) method / stream
   - test:
     - %% timeit -n
     - modify code, function, no. loops
     - ② broadcasting: operation → value
   - iloc/loc [label] - change values
     - add new values - new entry → | row
   - Series.append (series) → add new series.

3. Dataframe
   - create
     - series - row
     - dict - key - index - col. names
       - value - cols. values
   - extract - iloc/loc [index] → series (col name, col. value)
     - (new entry) row selection
     - 1 row / multi series.
   - loc [index, col. name] → multi axis
     - x chain operation
   - [col. [col. name]
   - [[ ]] loc[:, [col name1, col. name2]]
   - drop. - copy
   - del - inplace.

4 Dataframe indexing & loading
   - read data - Py create: series - dict
     - external file: pd. read_csv (file. + shell; delimited data, index_col ↓, headers —, skiprows ←)
   - into df.
   - reduce df → 1 col / row - df[ ]
     - view of df.

5. Query a Dataframe
   - Boolean Mask
   - 1-D Series, 2-D DF) - data structure → specific row/col.
   - overlay - comparison operator - T | F
     - Broadcasting
   - 2. overlay Boolean mask on df.
     - ① → condition = where
       - apply on df / Series = df.where (Boolean mask)
       - return new df / series: shape. index (T - ✓, F - NaN)
     - ② → value : index operator [ ]
       - df [Boolean mask]
   - 3. 2 Boolean mask outputs
     - chain. logical operators and | or
     - another Boolean mask: value / condition

6 Dataframes Index
   - set_index (col
   - reset_index
     - multi-level index: a list of cols
       - search in order
       - find distinct data. .unique
       - reduce data →
       - cols keep ↓ ↓ .
       - set composite keys
       - extract data - level

7 missing data
   - set_index - single col
   - sort_index → x unique
   - reset_index
   - set_index - composite
   - fillna (method = ffill)