Data Indexing and Selection

Data Indexing and Selection

What is Data Indexing and Selection?

Accessing and modifying values in pandas Series and DataFrame objects.

➡ Data Selection in Series

➡ Data Selection in DataFrame

Data Indexing and Selection

Data Selection in Series

The indexing and selection on Series can be done when:

Series Behaves As

One dimensional Dictionary

Series Behaves As

One dimensional array

Data Indexing and Selection

Data Selection in DataFrame

The indexing and selection on DataFrame can be done when:

DataFrame Behaves As

Dictionary

DataFrame Behaves As

Two Dimensional Array

Data Indexing and Selection

Indexing And Selection In General Can Be Divided In To:

- Implicit

- Explicit

Data Indexing and Selection

Implicit Indexing and Selection

Implicit

It is generally used to access the values by integers

It can be used when the indexing and selection is based on Pythonic style, i.e, by using integers index values

Implicit index is useful when Series Or DataFrame doesn't have explicit labels(or Index)

It is also helpful when Series Or DataFrame has integers as labels in both rows and columns starting with '0'

Data Indexing and Selection

Explicit Indexing and Selection

Explicit

It is generally used to access the values by labels instead of integers

It can be used when the indexing and selection is based on Dictionary style, i.e, by using Dictionary 'keys' as labels

It is useful when Series Or DataFrame both have explicit labels(Row Index and Column Index)

It is also helpful when masking, fancy indexing and slicing on Series Or DataFrame is done using the labels of rows and columns

Data Indexing and Selection

Explicit Is Always Better Than Implicit

Implicit Index	Explicit Index
Final index is excluded in the slice	Final index is included in the slice
It always refers to integers as index values	It always refers to labels as the index values

Data Indexing and Selection

The List Of Indexing And Selection Options in Pandas DataFrame(df):

Type	Descriptions
df[v]	This method selects single column or sequence of columns from the DataFrame
df.loc[v]	This method selects single row or subset of rows from the DataFrame by label
df.loc[:, v]	This method Selects single column or subset of columns by label
df.loc[v1, v2]	This method Selects both rows and columns by label
df.iloc[var]	This method Selects single row or subset of rows from the DataFrame by integer position
df.iloc[:, var]	This method Selects single column or subset of columns by integer position
df.iloc[var_i, var_j]	This method Selects both rows and columns by integer position
df.at[var_i, var_j]	This method Selects a single scalar value by row and column label
reindex method	This method Selects either rows or columns by labels