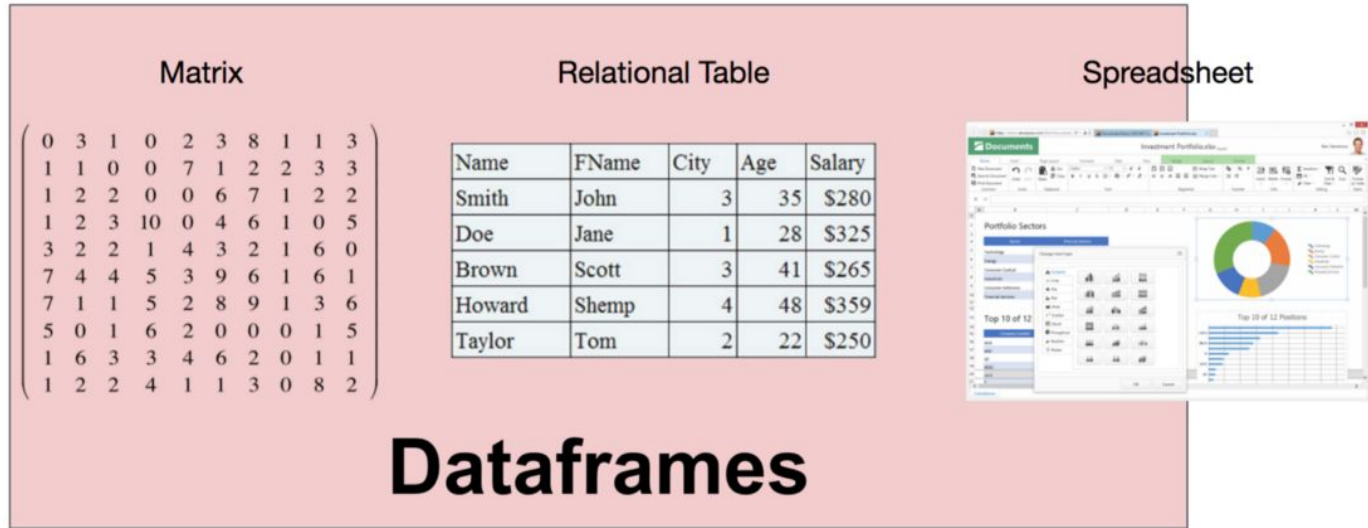


HAS Tools:

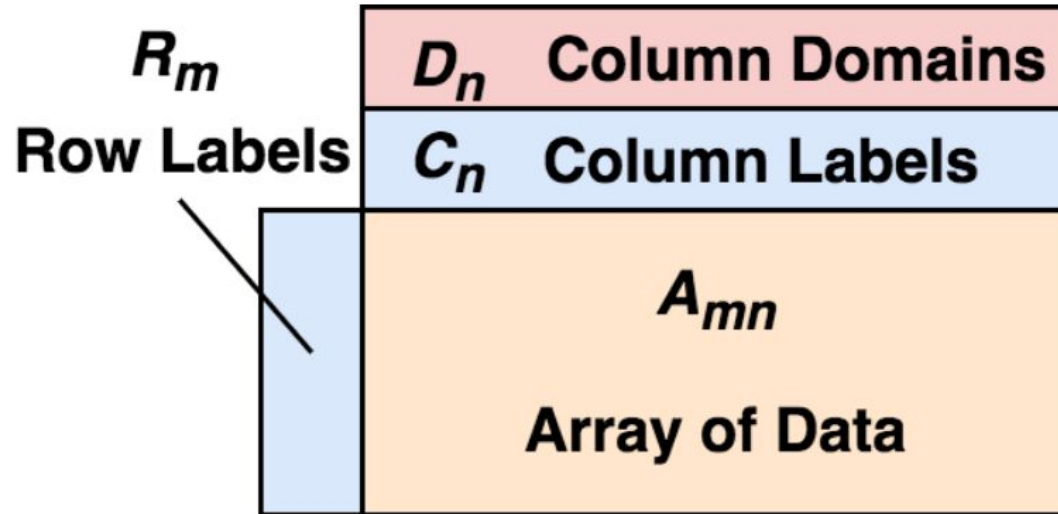
pandas stuff

September 23, 2024

DataFrames are “like” arrays, but also familiar if you are used to spreadsheets



DataFrames are “like” arrays, but also familiar if you are used to spreadsheets



pandas has 2 main data types: pd.Series & pd.DataFrame

pd.Series is *like* a one dimensional array, but has a key addition: the index

The index has a label attached to it

The index is used to select, align, and merge data

Series 1

INDEX	DATA
0	A
1	B
2	C
3	D
4	E
5	F

Series 2

INDEX	DATA
A	1
B	2
C	3
D	4
E	5
F	6

Series 3

INDEX	DATA
0	[1, 2]
1	A
2	1
3	(4, 5)
4	{"a": 1}
5	6

Series 4

INDEX	DATA
Jan-18	11
Feb-18	23
Mar-18	43
Apr-18	21
May-18	17
Jun-18	6

Pandas simplifies calculations

Like numpy operations
broadcast/align on
axes/dimensions, series

Series 1

INDEX	DATA
A	0
B	1
C	2
D	3
-	-

Pandas simplifies calculations

Like numpy operations
broadcast/align on
axes/dimensions, series

Alignment happens
along the index rather
than location

Series 1			Series 2	
INDEX	DATA		INDEX	DATA
A	0		-	-
B	1	+	B	10
C	2	+	C	11
D	3	+	D	12
-	-		E	13

Pandas simplifies calculations

Like numpy operations
broadcast/align on
axes/dimensions, series

Alignment happens
along the index rather
than location

So what happens when
there are missing values
or index discrepancies?

Series 1			Series 2	
INDEX	DATA		INDEX	DATA
A	0		-	-
B	1	+	B	10
C	2	+	C	11
D	3	+	D	12
-	-		E	13

Pandas simplifies calculations

Like numpy operations
broadcast/align on
axes/dimensions, series

Alignment happens
along the index rather
than location

So what happens when
there are missing values
or index discrepancies?

Series 1			Series 2			Series 1 + Series 2	
INDEX	DATA		INDEX	DATA		INDEX	DATA
A	0		-	-		A	NaN
B	1	+	B	10	=	B	11
C	2	+	C	11	=	C	13
D	3	+	D	12	=	D	15
-	-		E	13		E	NaN

Pandas also simplifies aligning and building datasets

Series, when combined
in multiple columns
become a DataFrame

Series are aligned
separately by name as
columns

Series 1		Series 2		Series 3	
INDEX	DATA	INDEX	DATA	INDEX	DATA
0	A	0	1	0	[1, 2]
1	B	1	2	1	A
2	C	2	3	2	1
3	D	3	4	3	(4, 5)
4	E	4	5	4	{"a": 1}
5	F	5	6	5	6

Pandas also simplifies aligning and building datasets

Series, when combined
in multiple columns
become a DataFrame

Series are aligned
separately by name as
columns

Series 1		Series 2		Series 3		Dataframe			
INDEX	DATA	INDEX	DATA	INDEX	DATA	INDEX	SERIES 1	SERIES 2	SERIES 3
0	A	0	1	0	[1, 2]	0	A	1	[1, 2]
1	B	1	2	1	A	1	B	2	A
2	C	2	3	2	1	2	C	3	1
3	D	3	4	3	(4, 5)	3	D	4	(4, 5)
4	E	4	5	4	{"a": 1}	4	E	5	{"a": 1}
5	F	5	6	5	6	5	F	6	6

Let's see how it works, jump to your codespaces