



DATAFRAME I





Topik diskusi hari ini

1. **Sorting in DataFrame**
2. **Filtering DataFrame**
3. **Creating Additional Column in DataFrame**



Sorting in Dataframe





Sorting in DataFrame

```
>>> df
   col1  col2  col3  col4
0     A     2     0     a
1     A     1     1     B
2     B     9     9     c
3  NaN     8     4     D
4     D     7     2     e
5     C     4     3     F
```

Sort → Mengurutkan

- Berdasarkan abjad
- Berdasarkan angka
- dsb..



Sorting in DataFrame

```
>>> df
   col1  col2  col3  col4
0     A     2     0     a
1     A     1     1     B
2     B     9     9     c
3  NaN     8     4     D
4     D     7     2     e
5     C     4     3     F
```

`DataFrame.sort_values(by, axis=0, ascending=True, inplace=False, kind='quicksort', na_position='last', ignore_index=False, key=None)`



Sorting in DataFrame

Case 1 : Mengurutkan berdasarkan abjad di col1 – part 1

```
>>> df
   col1  col2  col3  col4
0     A     2     0     a
1     A     1     1     B
2     B     9     9     c
3  NaN     8     4     D
4     D     7     2     e
5     C     4     3     F
```

```
>>> df.sort_values(by=['col1'])
   col1  col2  col3  col4
0     A     2     0     a
1     A     1     1     B
2     B     9     9     c
5     C     4     3     F
4     D     7     2     e
3  NaN     8     4     D
```



Sorting in DataFrame

Case 2 : Mengurutkan berdasarkan abjad di col1 – part 2

```
>>> df
   col1  col2  col3  col4
0     A     2     0     a
1     A     1     1     B
2     B     9     9     c
3  NaN     8     4     D
4     D     7     2     e
5     C     4     3     F
```

```
>>> df.sort_values(by='col1', ascending=False)
   col1  col2  col3  col4
4     D     7     2     e
5     C     4     3     F
2     B     9     9     c
0     A     2     0     a
1     A     1     1     B
3  NaN     8     4     D
```



Sorting in DataFrame

Case 3 : Mengurutkan berdasarkan abjad di col1 – part 3

```
>>> df
   col1  col2  col3 col4
0     A     2     0    a
1     A     1     1    B
2     B     9     9    c
3  NaN     8     4    D
4     D     7     2    e
5     C     4     3    F
```

```
>>> df.sort_values(by='col1', ascending=False, na_position='first')
   col1  col2  col3 col4
3  NaN     8     4    D
4     D     7     2    e
5     C     4     3    F
2     B     9     9    c
0     A     2     0    a
1     A     1     1    B
```




Sorting in DataFrame

Case 4 : Mengurutkan berdasarkan lebih dari 1 column

```
>>> df
   col1  col2  col3  col4
0     A     2     0     a
1     A     1     1     B
2     B     9     9     c
3  NaN     8     4     D
4     D     7     2     e
5     C     4     3     F
```

```
>>> df.sort_values(by=['col1', 'col2'])
   col1  col2  col3  col4
1     A     1     1     B
0     A     2     0     a
2     B     9     9     c
5     C     4     3     F
4     D     7     2     e
3  NaN     8     4     D
```

Filtering Dataframe





Filtering DataFrame

Melakukan seleksi terhadap dataframe untuk mendapatkan hanya informasi yang dibutuhkan/diinginkan

- Get only some columns
- Filter by single condition
- Filter by multiple conditions

	sepal_length	sepal_width	petal_length	petal_width	flower_class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa



Filtering DataFrame

- Get only some columns

```
df[['sepal_length', 'sepal_width']]
```

	sepal_length	sepal_width
0	5.1	3.5
1	4.9	3.0
2	4.7	3.2
3	4.6	3.1
4	5.0	3.6

```
df.filter(items=['sepal_length', 'sepal_width'])
```

	sepal_length	sepal_width
0	5.1	3.5
1	4.9	3.0
2	4.7	3.2
3	4.6	3.1
4	5.0	3.6



Filtering DataFrame

- Get some columns with filter using **.loc** & **.iloc**

.loc : label-based, perlu specify nama column & row

.iloc : integer index-based, perlu specify index dari column & row



Filtering DataFrame

- Get some columns with filter using **.loc** & **.iloc**

```
df.loc[df.flower_class == 'Iris-setosa']
```



	sepal_length	sepal_width	petal_length	petal_width	flower_class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa



Filtering DataFrame

- Get some columns with filter using **.loc** & **.iloc**

```
df.loc[(df.flower_class == 'Iris-setosa') & (df.sepal_length == 5.1)]
```

	sepal_length	sepal_width	petal_length	petal_width	flower_class
0	5.1	3.5	1.4	0.2	Iris-setosa
17	5.1	3.5	1.4	0.3	Iris-setosa
19	5.1	3.8	1.5	0.3	Iris-setosa



Filtering DataFrame

- Get some columns with filter using **.loc** & **.iloc**

```
[12] df.loc[(df.flower_class == 'Iris-setosa') & (df.sepal_length == 5.1), ['flower_class', 'sepal_length']]
```

	flower_class	sepal_length
0	Iris-setosa	5.1
17	Iris-setosa	5.1
19	Iris-setosa	5.1
21	Iris-setosa	5.1



Filtering DataFrame

- Get some columns with filter using **.loc** & **.iloc**

```
[16] df.iloc[[0,3]]
```

	sepal_length	sepal_width	petal_length	petal_width	flower_class
0	5.1	3.5	1.4	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa



Filtering DataFrame

- Get some columns with filter using **.loc** & **.iloc**

```
[19] df.iloc[[0,2],[1,3]]
```

	sepal_width	petal_width
0	3.5	0.2
2	3.2	0.2



Filtering DataFrame

- Get some columns with filter using **.loc** & **.iloc**

```
[18] df.iloc[0:3]
```

	sepal_length	sepal_width	petal_length	petal_width	flower_class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa



Filtering DataFrame

- Get some columns with filter using **.loc** & **.iloc**

```
df.iloc[1:3, 2:4]
```



	petal_length	petal_width
--	--------------	-------------

1	1.4	0.2
---	-----	-----

2	1.3	0.2
---	-----	-----

Creating Additional Column





Creating Additional Column

Memberikan column tambahan pada DataFrame

- Column tambahan : **value dari column yang lain**
- Column tambahan : **single value**
- Column tambahan : **other**



Creating Additional Column

- Column tambahan : value dari kolom-kolom yang lain

```
df['sepal_length_v2'] = df['sepal_length'] * 100  
df
```

	sepal_length	sepal_width	petal_length	petal_width	flower_class	sepal_length_v2
0	5.1	3.5	1.4	0.2	Iris-setosa	510.0
1	4.9	3.0	1.4	0.2	Iris-setosa	490.0
2	4.7	3.2	1.3	0.2	Iris-setosa	470.0
3	4.6	3.1	1.5	0.2	Iris-setosa	460.0
4	5.0	3.6	1.4	0.2	Iris-setosa	500.0

**Thank
YOU**

