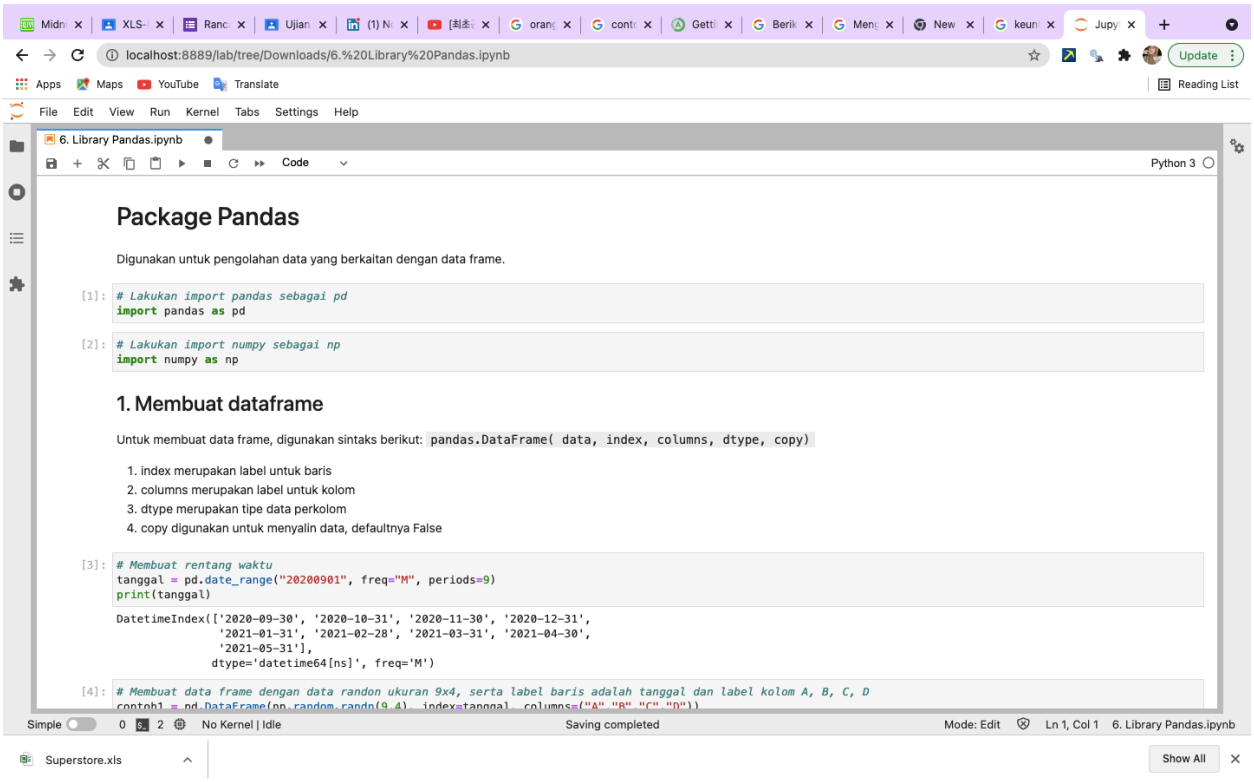
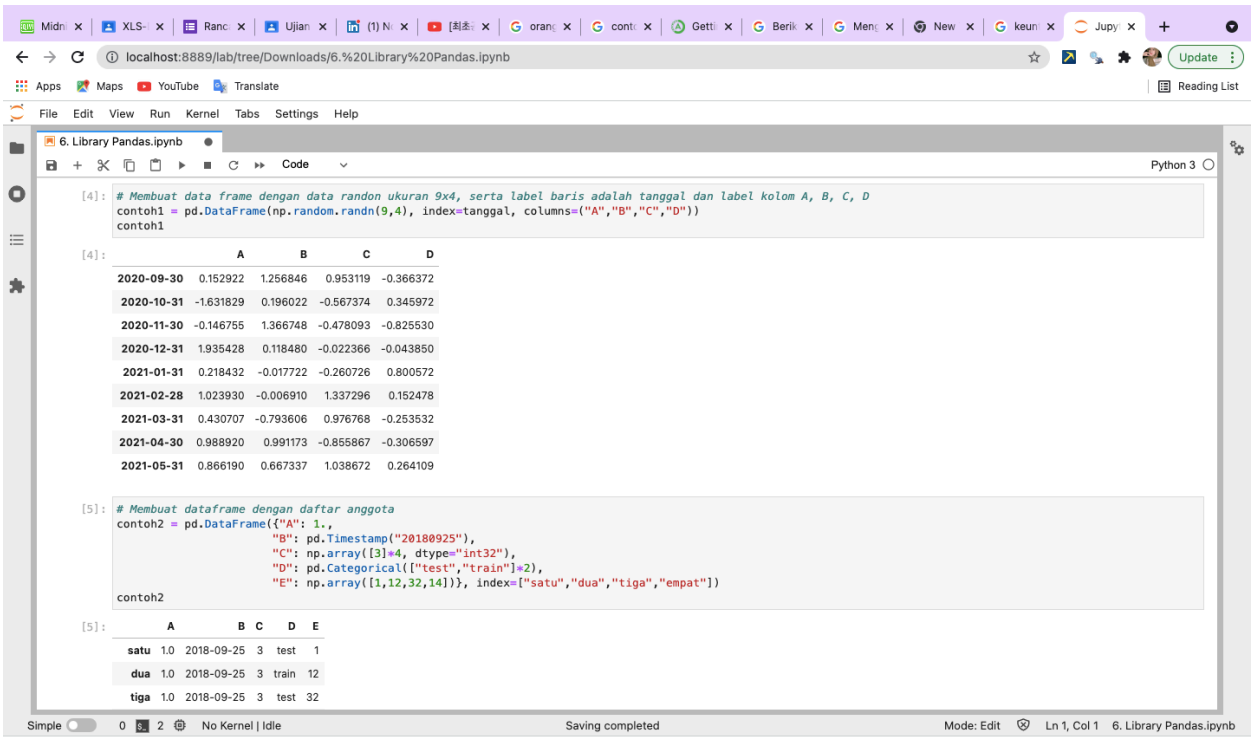
PDF-Hasil Running Notebook: Pandas





[4]:
```
# Membuat data frame dengan data randon ukuran 9x4, serta label baris adalah tanggal dan label kolom A, B, C, D
contoh1 = pd.DataFrame(np.random.randn(9,4), index=tanggal, columns=("A","B","C","D"))
contoh1
```

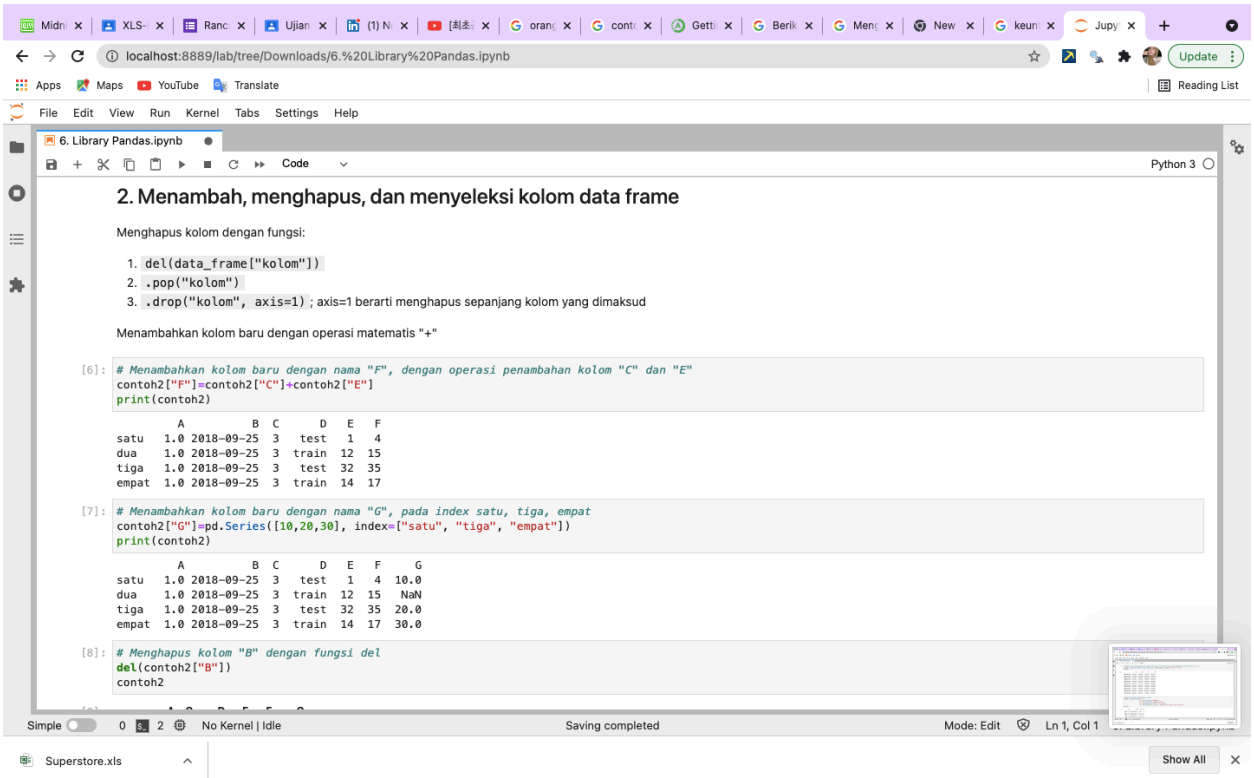
[4]:

	A	B	C	D
2020-09-30	0.152922	1.256846	0.953119	-0.366372
2020-10-31	-1.631829	0.196022	-0.567374	0.345972
2020-11-30	-0.146755	1.366748	-0.478093	-0.825530
2020-12-31	1.935428	0.118480	-0.022366	-0.043850
2021-01-31	0.218432	-0.017722	-0.260726	0.800572
2021-02-28	1.023930	-0.006910	1.337296	0.152478
2021-03-31	0.430707	-0.793606	0.976768	-0.253532
2021-04-30	0.988920	0.991173	-0.855867	-0.306597
2021-05-31	0.866190	0.667337	1.038672	0.264109

[5]:
```
# Membuat dataframe dengan daftar anggota
contoh2 = pd.DataFrame({"A": 1.,
                        "B": pd.Timestamp("20180925"),
                        "C": np.array([3]*4, dtype="int32"),
                        "D": pd.Categorical(["test","train"]*2),
                        "E": np.array([1,12,32,14])}, index=["satu","dua","tiga","empat"])
contoh2
```

[5]:

	A	B	C	D	E
satu	1.0	2018-09-25	3	test	1
dua	1.0	2018-09-25	3	train	12
tiga	1.0	2018-09-25	3	test	32



2. Menambah, menghapus, dan menyeleksi kolom data frame

Menghapus kolom dengan fungsi:

1. `del(data_frame["kolom"])`
2. `.pop("kolom")`
3. `.drop("kolom", axis=1)` ; axis=1 berarti menghapus sepanjang kolom yang dimaksud

Menambahkan kolom baru dengan operasi matematis "+"

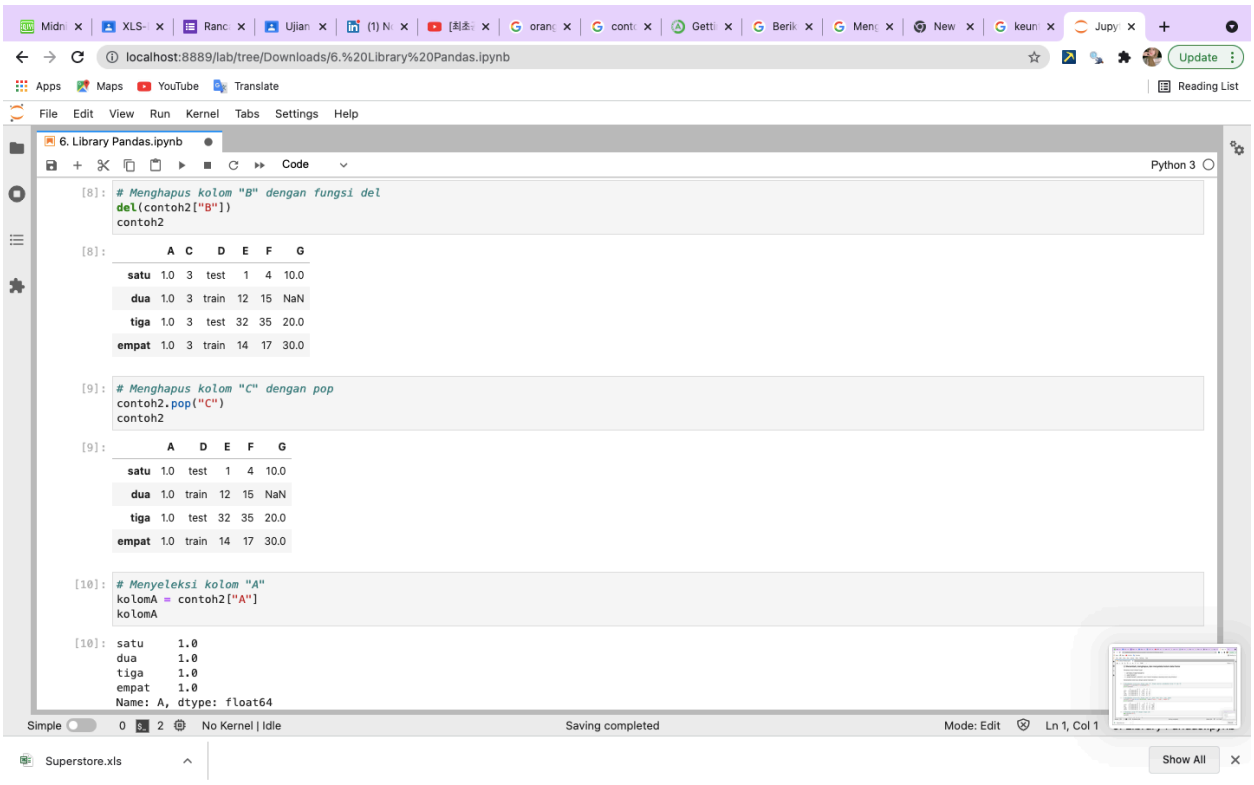
[6]:
```
# Menambahkan kolom baru dengan nama "F", dengan operasi penambahan kolom "C" dan "E"
contoh2["F"]=contoh2["C"]+contoh2["E"]
print(contoh2)
```

```
          A          B  C      D   E   F
satu    1.0 2018-09-25  3   test   1   4
dua     1.0 2018-09-25  3  train  12  15
tiga    1.0 2018-09-25  3   test  32  35
empat   1.0 2018-09-25  3  train  14  17
```

[7]:
```
# Menambahkan kolom baru dengan nama "G", pada index satu, tiga, empat
contoh2["G"]=pd.Series([10,20,30], index=["satu", "tiga", "empat"])
print(contoh2)
```

```
          A          B  C      D   E   F     G
satu    1.0 2018-09-25  3   test   1   4  10.0
dua     1.0 2018-09-25  3  train  12  15   NaN
tiga    1.0 2018-09-25  3   test  32  35  20.0
empat   1.0 2018-09-25  3  train  14  17  30.0
```

[8]:
```
# Menghapus kolom "B" dengan fungsi del
del(contoh2["B"])
contoh2
```



```
[8]: # Menghapus kolom "B" dengan fungsi del
     del(contoh2["B"])
     contoh2
```

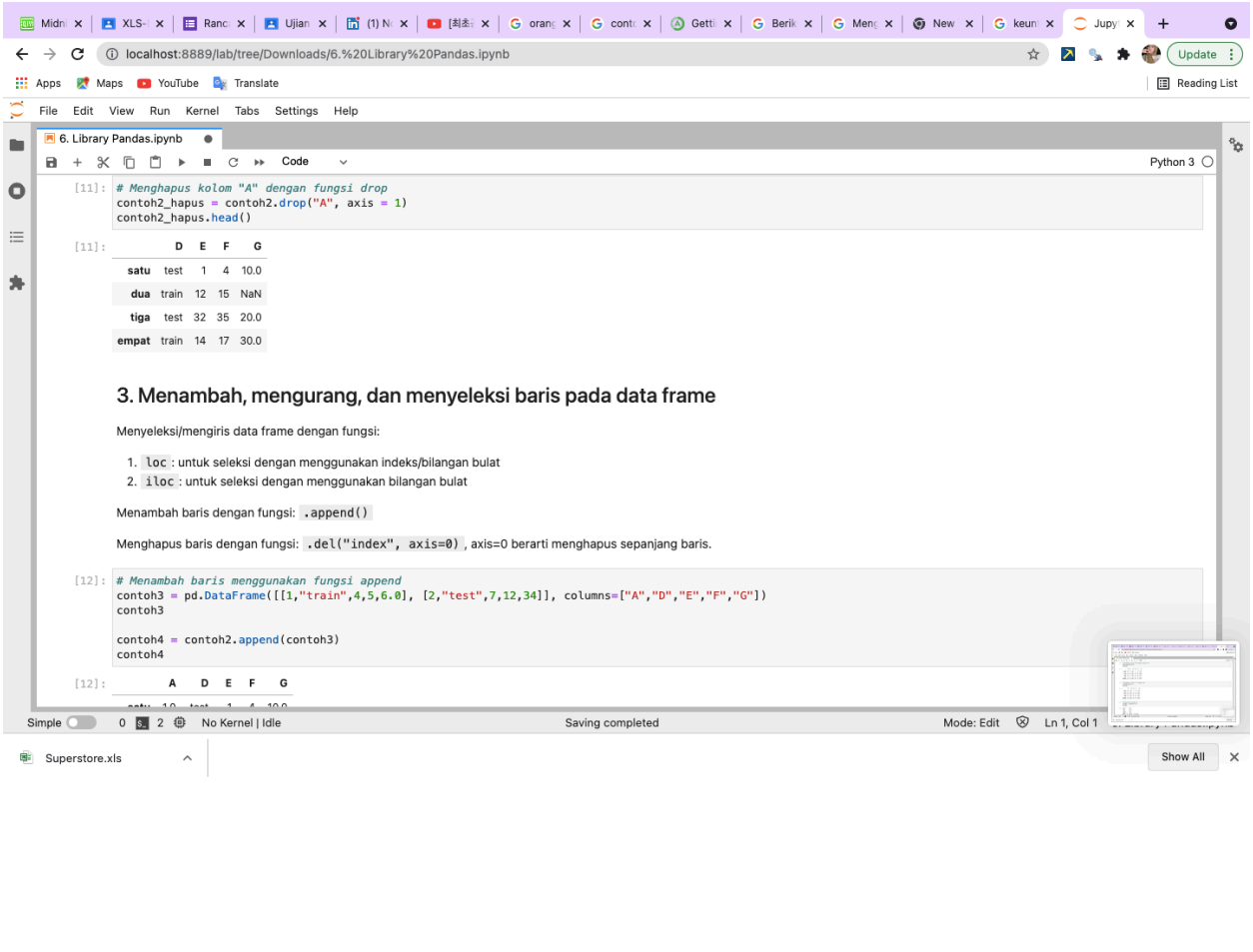
	A	C	D	E	F	G
satu	1.0	3	test	1	4	10.0
dua	1.0	3	train	12	15	NaN
tiga	1.0	3	test	32	35	20.0
empat	1.0	3	train	14	17	30.0

```
[9]: # Menghapus kolom "C" dengan pop
     contoh2.pop("C")
     contoh2
```

	A	D	E	F	G
satu	1.0	test	1	4	10.0
dua	1.0	train	12	15	NaN
tiga	1.0	test	32	35	20.0
empat	1.0	train	14	17	30.0

```
[10]: # Menyeleksi kolom "A"
      kolomA = contoh2["A"]
      kolomA
```

```
satu     1.0
dua      1.0
tiga     1.0
empat    1.0
Name: A, dtype: float64
```



```
[11]: # Menghapus kolom "A" dengan fungsi drop
      contoh2_hapus = contoh2.drop("A", axis = 1)
      contoh2_hapus.head()
```

	D	E	F	G
satu	test	1	4	10.0
dua	train	12	15	NaN
tiga	test	32	35	20.0
empat	train	14	17	30.0

3. Menambah, mengurang, dan menyeleksi baris pada data frame

Menyeleksi/mengiris data frame dengan fungsi:

1. `loc` : untuk seleksi dengan menggunakan indeks/bilangan bulat
2. `iloc` : untuk seleksi dengan menggunakan bilangan bulat

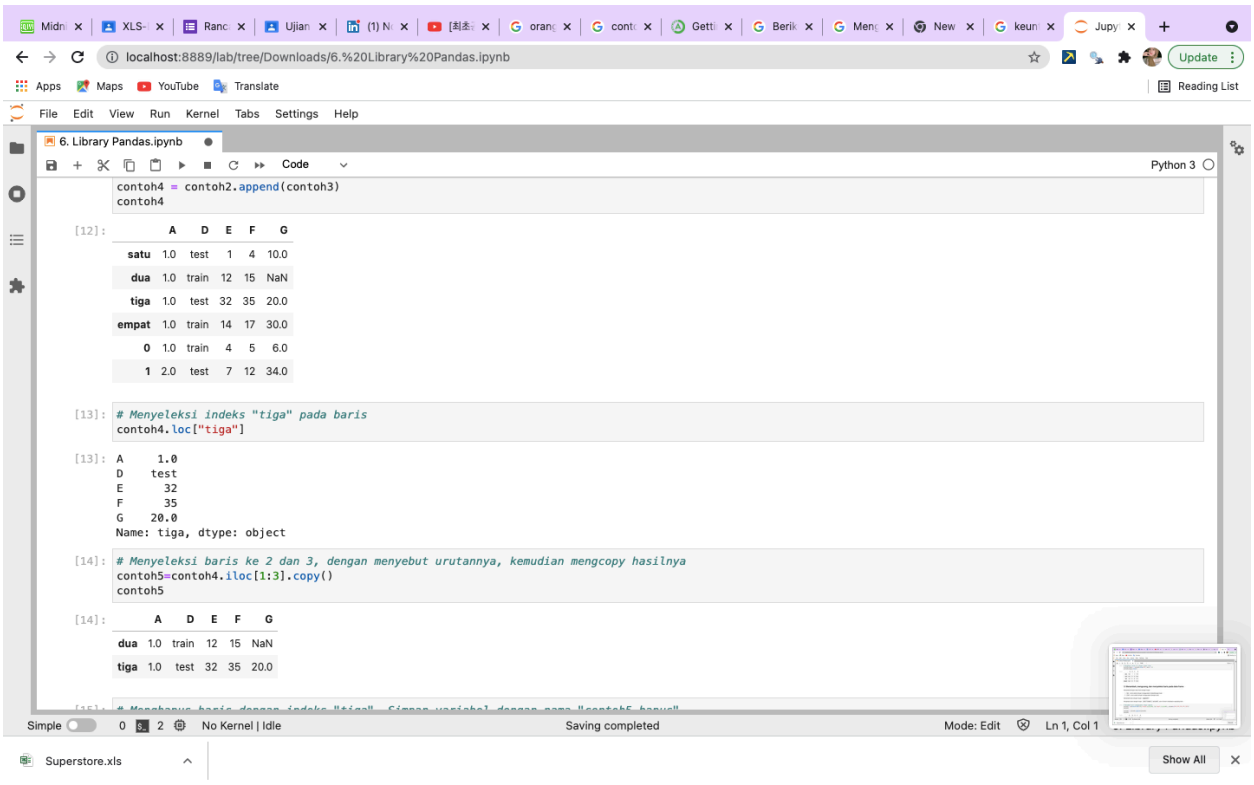
Menambah baris dengan fungsi: `.append()`

Menghapus baris dengan fungsi: `.del("index", axis=0)` , axis=0 berarti menghapus sepanjang baris.

```
[12]: # Menambah baris menggunakan fungsi append
      contoh3 = pd.DataFrame([[1,"train",4,5,6.0], [2,"test",7,12,34]], columns=["A","D","E","F","G"])
      contoh3

      contoh4 = contoh2.append(contoh3)
      contoh4
```

	A	D	E	F	G



```
contoh4 = contoh2.append(contoh3)
contoh4
```

[12]:

	A	D	E	F	G
satu	1.0	test	1	4	10.0
dua	1.0	train	12	15	NaN
tiga	1.0	test	32	35	20.0
empat	1.0	train	14	17	30.0
0	1.0	train	4	5	6.0
1	2.0	test	7	12	34.0

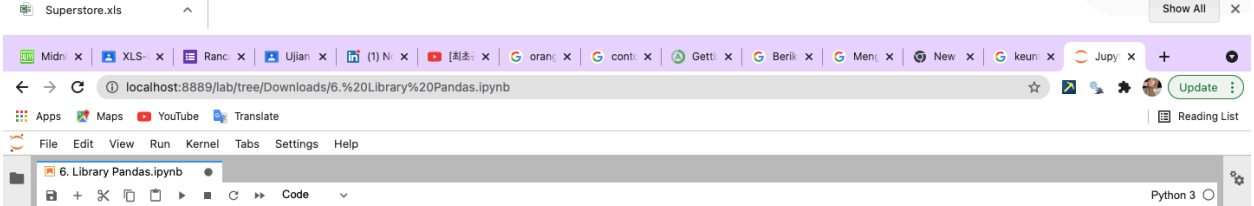
[13]:
```
# Menyeleksi indeks "tiga" pada baris
contoh4.loc["tiga"]
```

[13]:
```
A     1.0
D    test
E      32
F      35
G    20.0
Name: tiga, dtype: object
```

[14]:
```
# Menyeleksi baris ke 2 dan 3, dengan menyebut urutannya, kemudian mengcopy hasilnya
contoh5=contoh4.iloc[1:3].copy()
contoh5
```

[14]:

	A	D	E	F	G
dua	1.0	train	12	15	NaN
tiga	1.0	test	32	35	20.0



4. Menyeleksi/megiris data frame berdasar baris dan kolom

[16]:
```
contoh2
```

[16]:

	A	D	E	F	G
satu	1.0	test	1	4	10.0
dua	1.0	train	12	15	NaN
tiga	1.0	test	32	35	20.0
empat	1.0	train	14	17	30.0

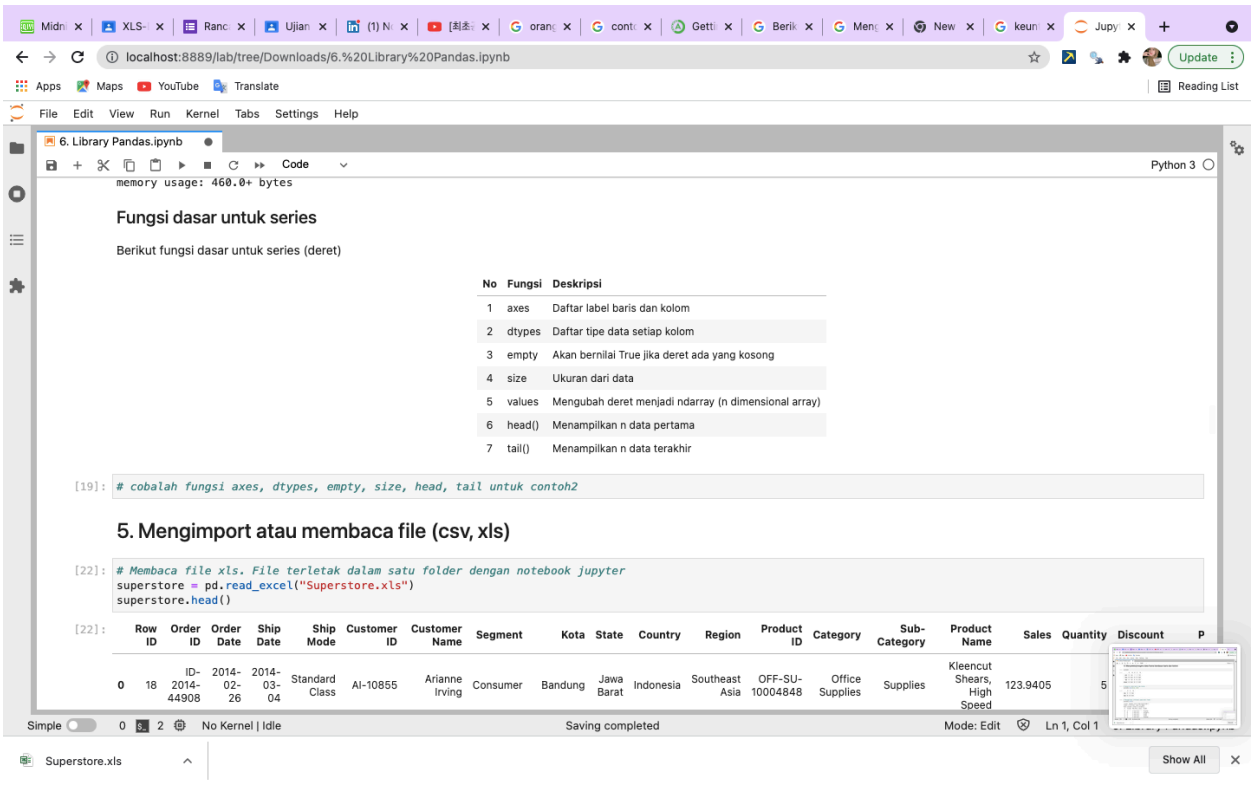
[17]:
```
# mengiris data baris dan kolom
contoh2.iloc[1:3, 2:5]
```

[17]:

	E	F	G
dua	12	15	NaN
tiga	32	35	20.0

[18]:
```
# Menampilkan informasi pada data frame
contoh2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 4 entries, satu to empat
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   A       4 non-null      float64
 1   D       4 non-null      category
 2   E       4 non-null      int64
 3   F       4 non-null      int64
 4   G       3 non-null      float64
```



memory usage: 460.0+ bytes

Fungsi dasar untuk series

Berikut fungsi dasar untuk series (deret)

No	Fungsi	Deskripsi
1	axes	Daftar label baris dan kolom
2	dtypes	Daftar tipe data setiap kolom
3	empty	Akan bernilai True jika deret ada yang kosong
4	size	Ukuran dari data
5	values	Mengubah deret menjadi ndarray (n dimensional array)
6	head()	Menampilkan n data pertama
7	tail()	Menampilkan n data terakhir

[19]:
```
# cobalah fungsi axes, dtypes, empty, size, head, tail untuk contoh2
```

5. Mengimport atau membaca file (csv, xls)

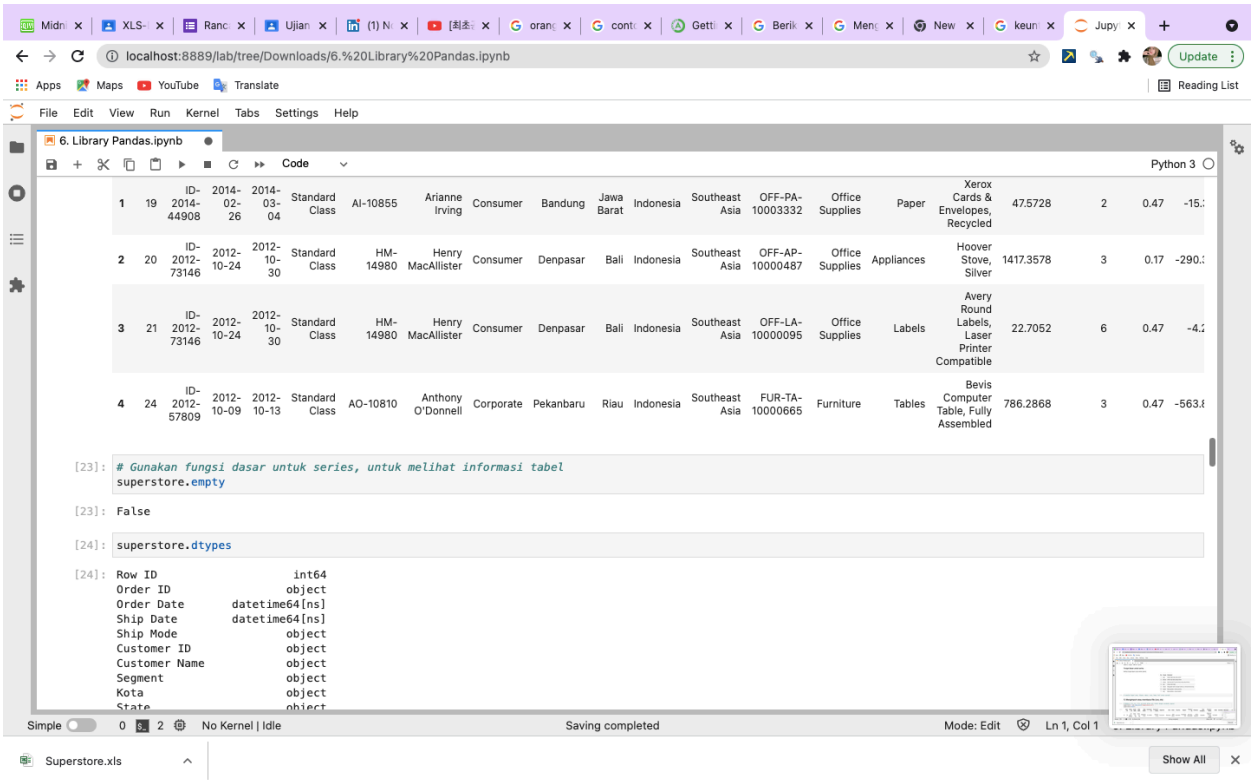
[22]:
```
# Membaca file xls. File terletak dalam satu folder dengan notebook jupyter
superstore = pd.read_excel("Superstore.xls")
superstore.head()
```

[22]:

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Kota	State	Country	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount	P
0	18	ID-2014-44908	2014-02-26	2014-03-04	Standard Class	AI-10855	Arianne Irving	Consumer	Bandung	Jawa Barat	Indonesia	Southeast Asia	OFF-SU-10004848	Office Supplies	Supplies	Kleencut Shears, High Speed	123.9405	5		

Simple 0 2 No Kernel | Idle — Saving completed — Mode: Edit — Ln 1, Col 1

Superstore.xls — Show All



1	19	ID-2014-44908	2014-02-26	2014-03-04	Standard Class	AI-10855	Arianne Irving	Consumer	Bandung	Jawa Barat	Indonesia	Southeast Asia	OFF-PA-10003332	Office Supplies	Paper	Xerox Cards & Envelopes, Recycled	47.5728	2	0.47	-15.1
2	20	ID-2012-73146	2012-10-24	2012-10-30	Standard Class	HM-14980	Henry MacAllister	Consumer	Denpasar	Bali	Indonesia	Southeast Asia	OFF-AP-10000487	Office Supplies	Appliances	Hoover Stove, Silver	1417.3578	3	0.17	-290.1
3	21	ID-2012-73146	2012-10-24	2012-10-30	Standard Class	HM-14980	Henry MacAllister	Consumer	Denpasar	Bali	Indonesia	Southeast Asia	OFF-LA-10000095	Office Supplies	Labels	Avery Round Labels, Laser Printer Compatible	22.7052	6	0.47	-4.2
4	24	ID-2012-57809	2012-10-09	2012-10-13	Standard Class	AO-10810	Anthony O'Donnell	Corporate	Pekanbaru	Riau	Indonesia	Southeast Asia	FUR-TA-10000665	Furniture	Tables	Bevis Computer Table, Fully Assembled	786.2868	3	0.47	-563.8

[23]:
```
# Gunakan fungsi dasar untuk series, untuk melihat informasi tabel
superstore.empty
```

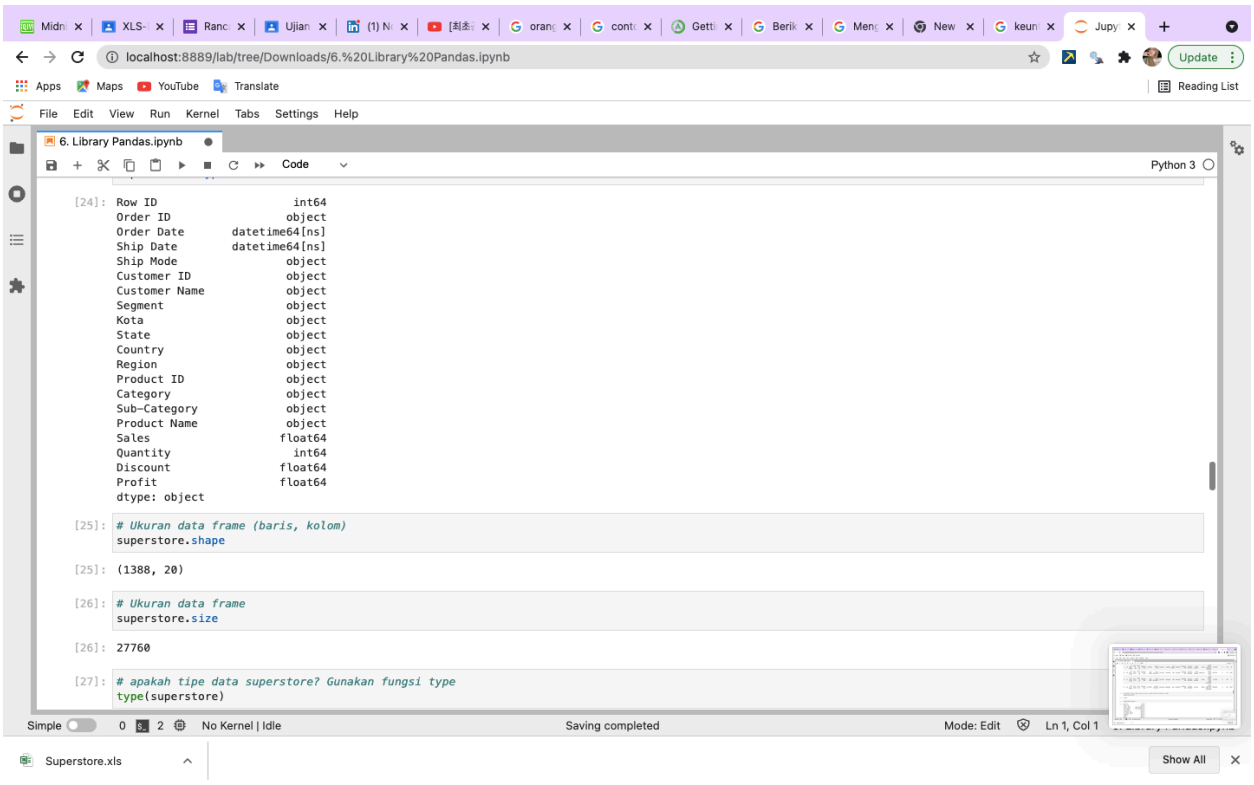
[23]:
```
False
```

[24]:
```
superstore.dtypes
```

[24]:
```
Row ID                  int64
Order ID               object
Order Date     datetime64[ns]
Ship Date      datetime64[ns]
Ship Mode              object
Customer ID            object
Customer Name          object
Segment                object
Kota                   object
State                  object
```

Simple 0 2 No Kernel | Idle — Saving completed — Mode: Edit — Ln 1, Col 1

Superstore.xls — Show All



```
[24]: Row ID                  int64
      Order ID                object
      Order Date      datetime64[ns]
      Ship Date       datetime64[ns]
      Ship Mode               object
      Customer ID             object
      Customer Name           object
      Segment                 object
      Kota                    object
      State                   object
      Country                 object
      Region                  object
      Product ID              object
      Category                object
      Sub-Category            object
      Product Name            object
      Sales                  float64
      Quantity                 int64
      Discount               float64
      Profit                 float64
      dtype: object
```

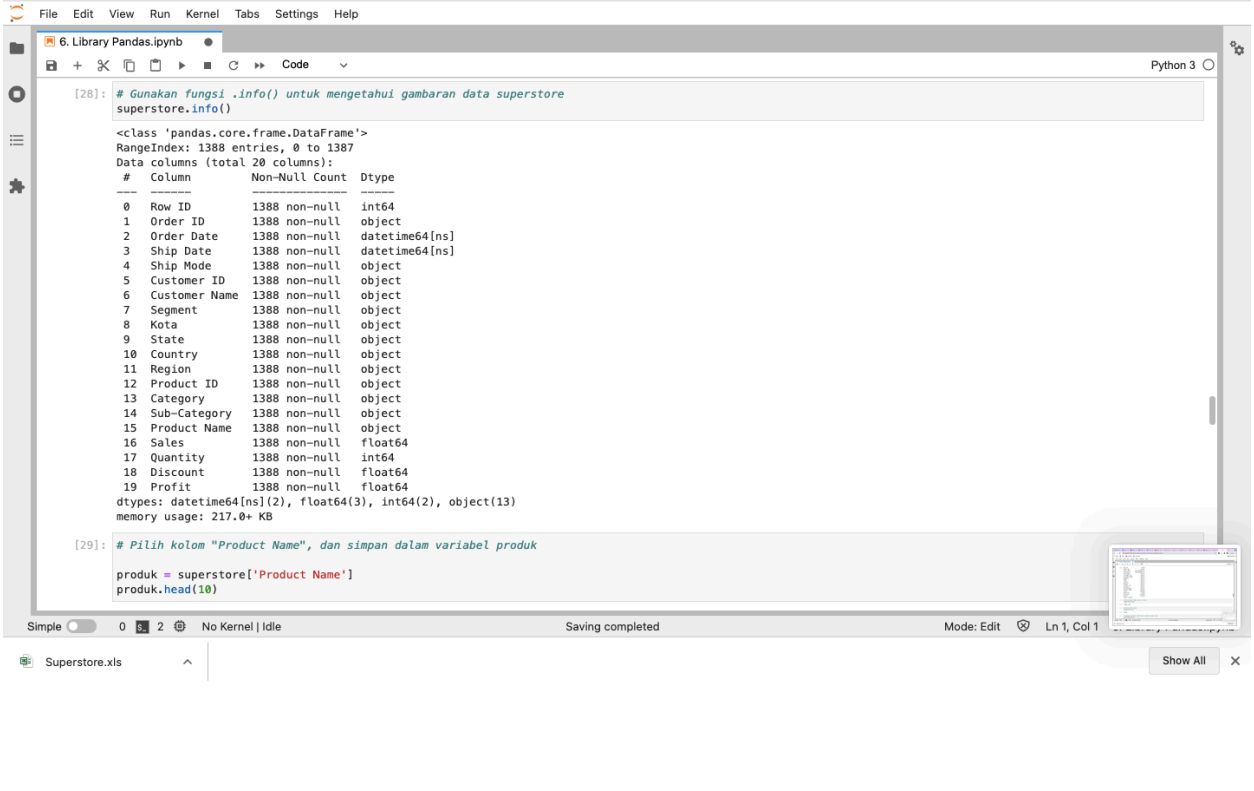
```
[25]: # Ukuran data frame (baris, kolom)
      superstore.shape
```

```
[25]: (1388, 20)
```

```
[26]: # Ukuran data frame
      superstore.size
```

```
[26]: 27760
```

```
[27]: # apakah tipe data superstore? Gunakan fungsi type
      type(superstore)
```

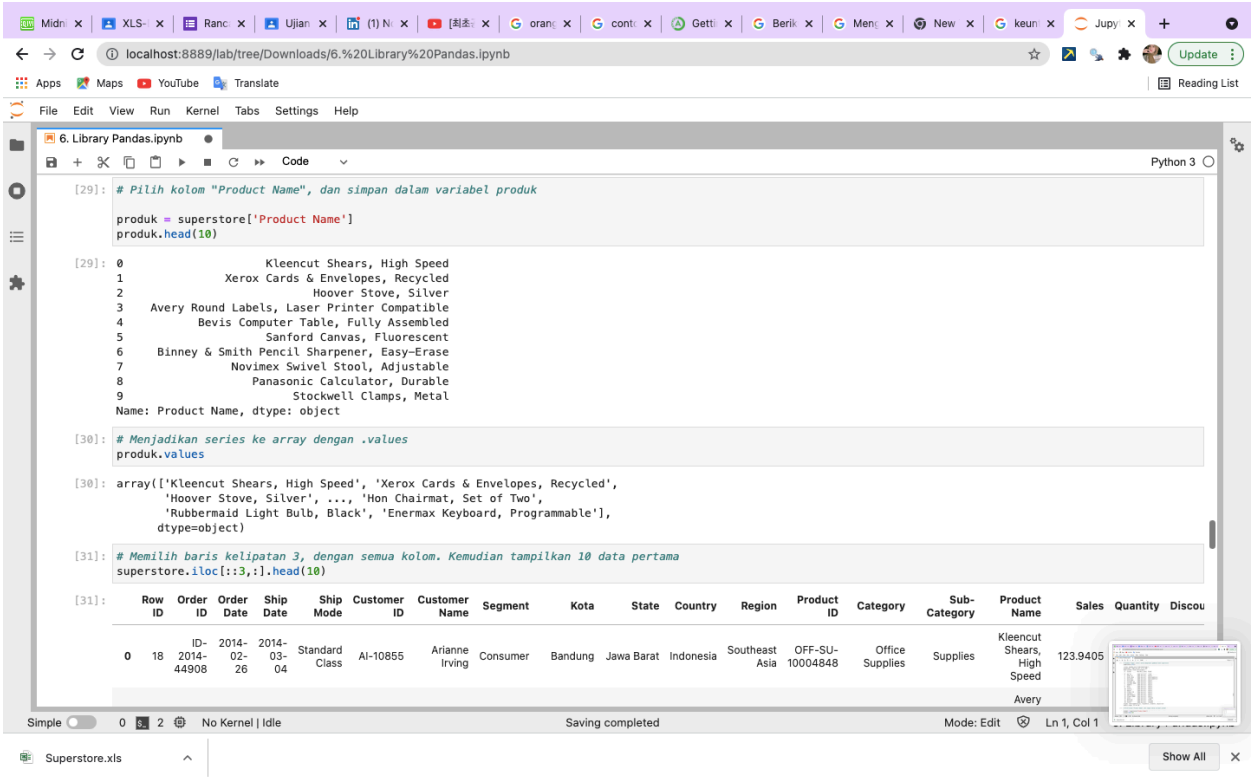


```
[28]: # Gunakan fungsi .info() untuk mengetahui gambaran data superstore
      superstore.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1388 entries, 0 to 1387
Data columns (total 20 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Row ID         1388 non-null   int64
 1   Order ID       1388 non-null   object
 2   Order Date     1388 non-null   datetime64[ns]
 3   Ship Date      1388 non-null   datetime64[ns]
 4   Ship Mode      1388 non-null   object
 5   Customer ID    1388 non-null   object
 6   Customer Name  1388 non-null   object
 7   Segment        1388 non-null   object
 8   Kota           1388 non-null   object
 9   State          1388 non-null   object
 10  Country        1388 non-null   object
 11  Region         1388 non-null   object
 12  Product ID     1388 non-null   object
 13  Category       1388 non-null   object
 14  Sub-Category   1388 non-null   object
 15  Product Name   1388 non-null   object
 16  Sales          1388 non-null   float64
 17  Quantity       1388 non-null   int64
 18  Discount       1388 non-null   float64
 19  Profit         1388 non-null   float64
dtypes: datetime64[ns](2), float64(3), int64(2), object(13)
memory usage: 217.0+ KB
```

```
[29]: # Pilih kolom "Product Name", dan simpan dalam variabel produk

      produk = superstore['Product Name']
      produk.head(10)
```



```
# Pilih kolom "Product Name", dan simpan dalam variabel produk

produk = superstore['Product Name']
produk.head(10)
```

```
0                  Kleencut Shears, High Speed
1              Xerox Cards & Envelopes, Recycled
2                         Hoover Stove, Silver
3    Avery Round Labels, Laser Printer Compatible
4          Bevis Computer Table, Fully Assembled
5                    Sanford Canvas, Fluorescent
6     Binney & Smith Pencil Sharpener, Easy-Erase
7                 Novimex Swivel Stool, Adjustable
8                    Panasonic Calculator, Durable
9                         Stockwell Clamps, Metal
Name: Product Name, dtype: object
```

```
# Menjadikan series ke array dengan .values
produk.values
```

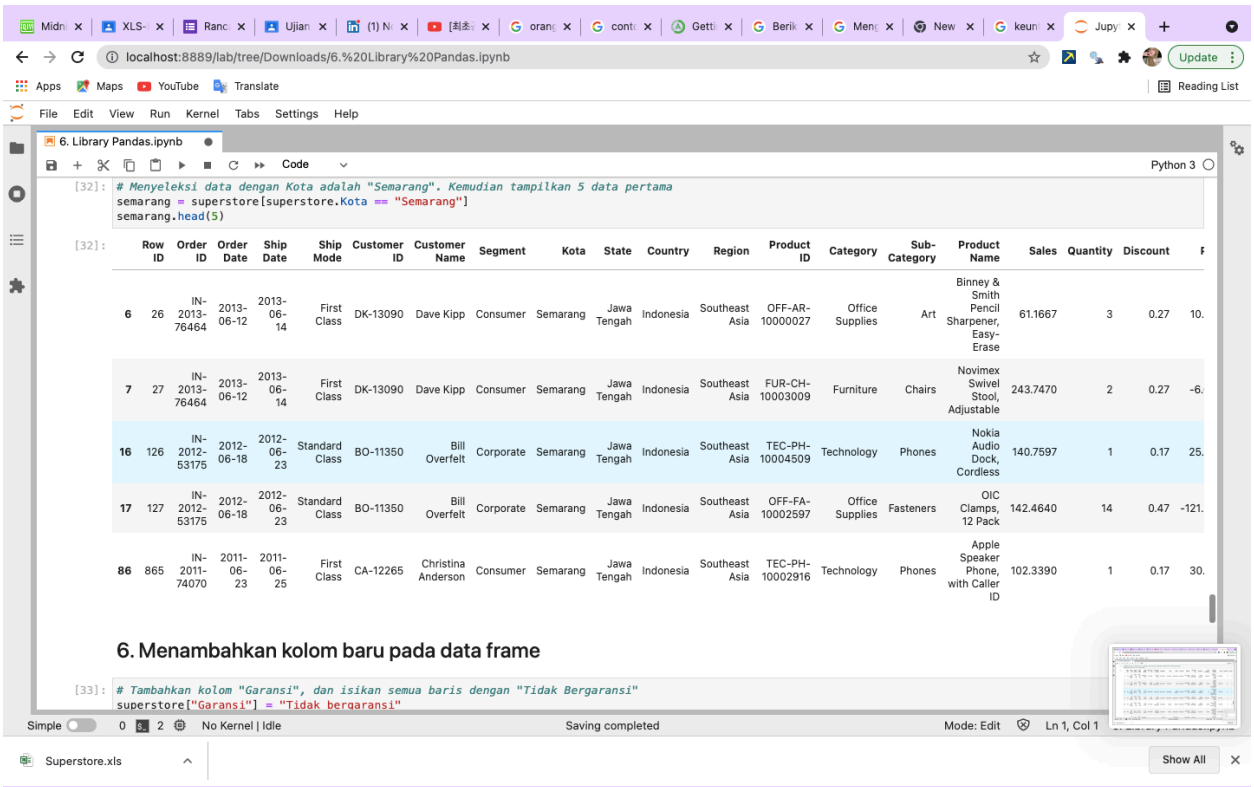
```
array(['Kleencut Shears, High Speed', 'Xerox Cards & Envelopes, Recycled',
       'Hoover Stove, Silver', ..., 'Hon Chairmat, Set of Two',
       'Rubbermaid Light Bulb, Black', 'Enermax Keyboard, Programmable'],
      dtype=object)
```

```
# Memilih baris kelipatan 3, dengan semua kolom. Kemudian tampilkan 10 data pertama
superstore.iloc[::3,:].head(10)
```

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Kota	State	Country	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discou
0	18	ID-2014-44908	2014-02-26	2014-03-04	Standard Class	AI-10855	Arianne Irving	Consumer	Bandung	Jawa Barat	Indonesia	Southeast Asia	OFF-SU-10004848	Office Supplies	Supplies	Kleencut Shears, High Speed	123.9405		
																Avery			


```
# Memilih baris kelipatan 3, dengan semua kolom. Kemudian tampilkan 10 data pertama
superstore.iloc[::3,:].head(10)
```


	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Kota	State	Country	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discou
0	18	ID-2014-44908	2014-02-26	2014-03-04	Standard Class	AI-10855	Arianne Irving	Consumer	Bandung	Jawa Barat	Indonesia	Southeast Asia	OFF-SU-10004848	Office Supplies	Supplies	Kleencut Shears, High Speed	123.9405	5	0.
3	21	ID-2012-73146	2012-10-24	2012-10-30	Standard Class	HM-14980	Henry MacAllister	Consumer	Denpasar	Bali	Indonesia	Southeast Asia	OFF-LA-10000095	Office Supplies	Labels	Avery Round Labels, Laser Printer Compatible	22.7052	6	0.
6	26	IN-2013-76464	2013-06-12	2013-06-14	First Class	DK-13090	Dave Kipp	Consumer	Semarang	Jawa Tengah	Indonesia	Southeast Asia	OFF-AR-10000027	Office Supplies	Art	Binney & Smith Pencil Sharpener, Easy-Erase	61.1667	3	0.
9	97	ID-2011-74721	2011-05-04	2011-05-05	First Class	HG-14845	Harry Greene	Consumer	Yogyakarta	Yogyakarta	Indonesia	Southeast Asia	OFF-FA-10001010	Office Supplies	Fasteners	Stockwell Clamps, Metal	82.0440	8	0.
12	110	ID-2012-10230	2012-07-06	2012-07-09	First Class	JC-16105	Julie Creighton	Corporate	Jakarta	Jakarta	Indonesia	Southeast Asia	OFF-LA-10004062	Office Supplies	Labels	Hon Removable Labels, 5000 Label Set	16.3134	3	0.
15	115	ID-2011-36011	2011-01-26	2011-01-27	First Class	HK-14890	Heather Kirkland	Corporate	Depok	Jawa Barat	Indonesia	Southeast Asia	FUR-CH-10003950	Furniture	Chairs	Novimex Executive Leather Armchair, Black	334.3035		
		ID-	2012-	2012-	Standard			Home		Sumatera		Southeast	TEC-AC-			SanDisk			



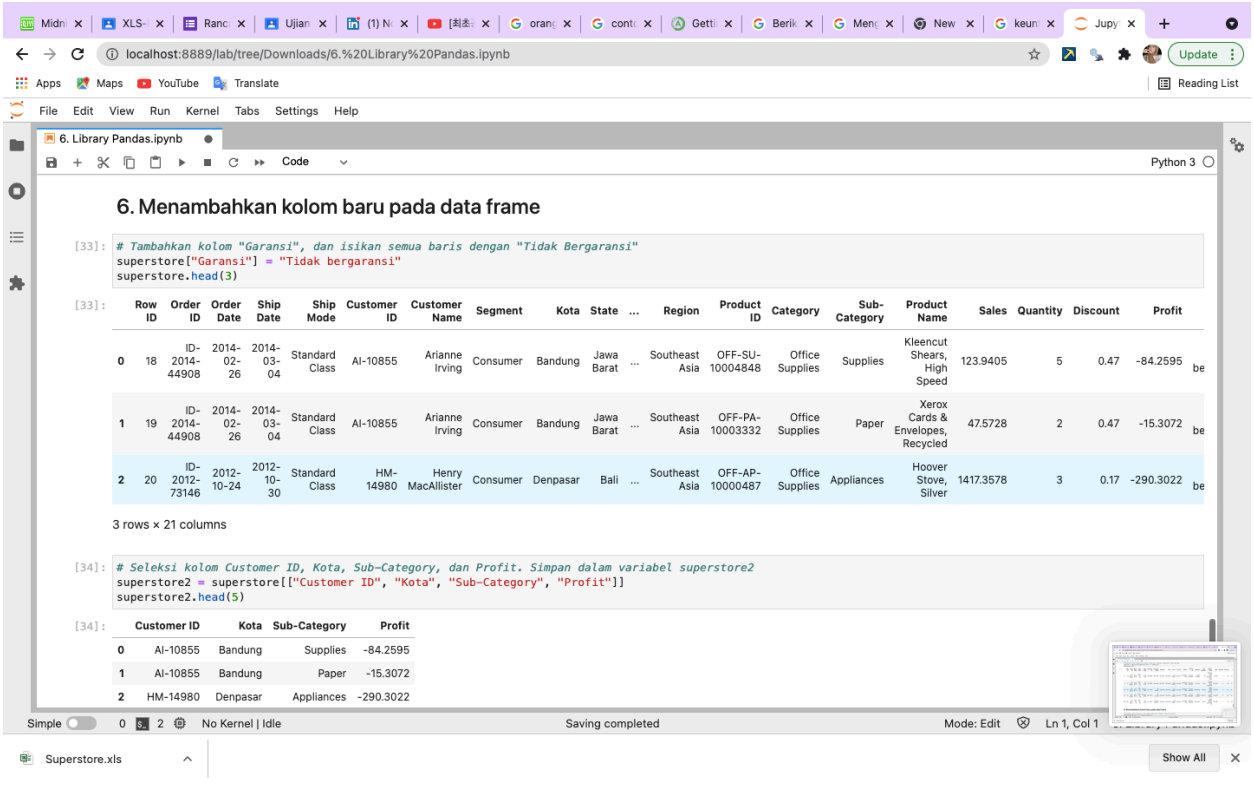
```
[32]: # Menyeleksi data dengan Kota adalah "Semarang". Kemudian tampilkan 5 data pertama
semarang = superstore[superstore.Kota == "Semarang"]
semarang.head(5)
```

[32]:

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Kota	State	Country	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount	P
6	26	IN-2013-76464	2013-06-12	2013-06-14	First Class	DK-13090	Dave Kipp	Consumer	Semarang	Jawa Tengah	Indonesia	Southeast Asia	OFF-AR-10000027	Office Supplies	Art	Binney & Smith Pencil Sharpener, Easy-Erase	61.1667	3	0.27	10.
7	27	IN-2013-76464	2013-06-12	2013-06-14	First Class	DK-13090	Dave Kipp	Consumer	Semarang	Jawa Tengah	Indonesia	Southeast Asia	FUR-CH-10003009	Furniture	Chairs	Novimex Swivel Stool, Adjustable	243.7470	2	0.27	-6.
16	126	IN-2012-53175	2012-06-18	2012-06-23	Standard Class	BO-11350	Bill Overfelt	Corporate	Semarang	Jawa Tengah	Indonesia	Southeast Asia	TEC-PH-10004509	Technology	Phones	Nokia Audio Dock, Cordless	140.7597	1	0.17	25.
17	127	IN-2012-53175	2012-06-18	2012-06-23	Standard Class	BO-11350	Bill Overfelt	Corporate	Semarang	Jawa Tengah	Indonesia	Southeast Asia	OFF-FA-10002597	Office Supplies	Fasteners	OIC Clamps, 12 Pack	142.4640	14	0.47	-121.
86	865	IN-2011-74070	2011-06-23	2011-06-25	First Class	CA-12265	Christina Anderson	Consumer	Semarang	Jawa Tengah	Indonesia	Southeast Asia	TEC-PH-10002916	Technology	Phones	Apple Speaker Phone, with Caller ID	102.3390	1	0.17	30.

6. Menambahkan kolom baru pada data frame

```
[33]: # Tambahkan kolom "Garansi", dan isikan semua baris dengan "Tidak Bergaransi"
superstore["Garansi"] = "Tidak bergaransi"
```



```
[33]: # Tambahkan kolom "Garansi", dan isikan semua baris dengan "Tidak Bergaransi"
superstore["Garansi"] = "Tidak bergaransi"
superstore.head(3)
```

[33]:

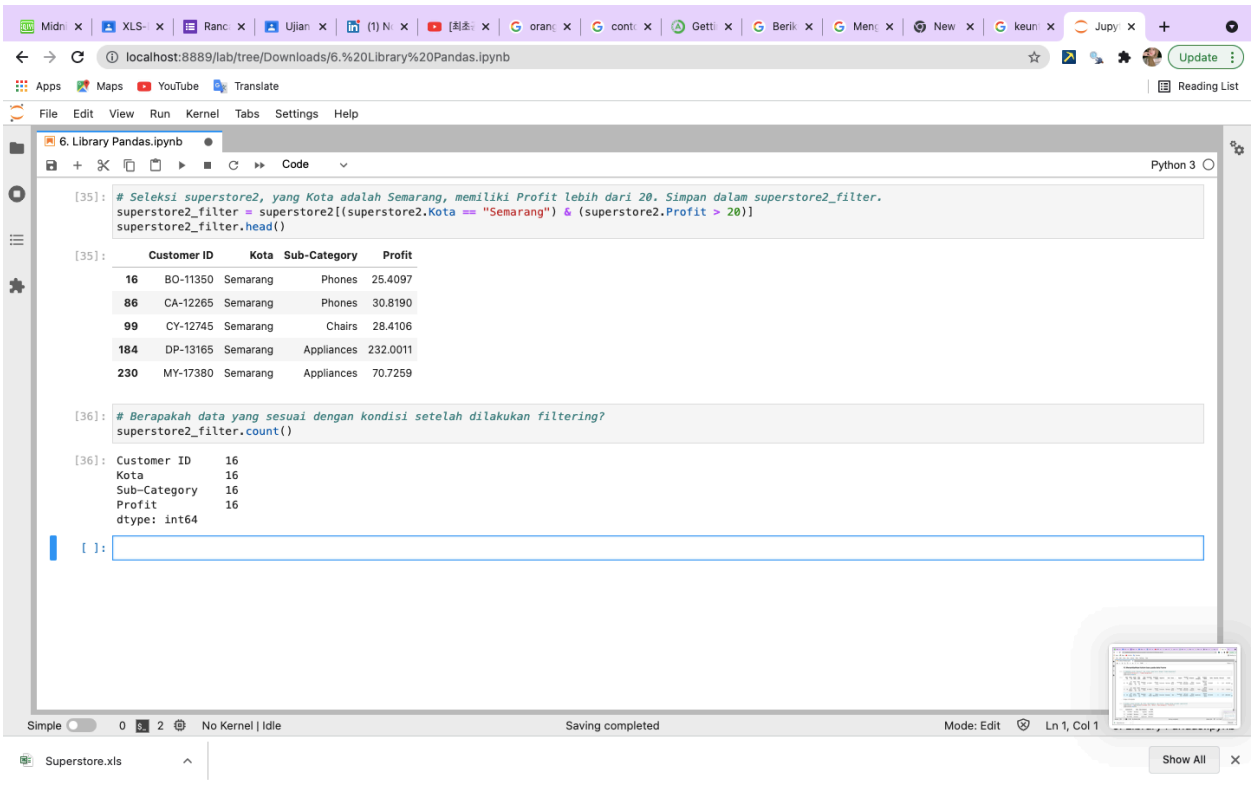
	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Kota	State	...	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount	Profit	
0	18	ID-2014-44908	2014-02-26	2014-03-04	Standard Class	AI-10855	Arianne Irving	Consumer	Bandung	Jawa Barat	...	Southeast Asia	OFF-SU-10004848	Office Supplies	Supplies	Kleencut Shears, High Speed	123.9405	5	0.47	-84.2595	be
1	19	ID-2014-44908	2014-02-26	2014-03-04	Standard Class	AI-10855	Arianne Irving	Consumer	Bandung	Jawa Barat	...	Southeast Asia	OFF-PA-10003332	Office Supplies	Paper	Xerox Cards & Envelopes, Recycled	47.5728	2	0.47	-15.3072	be
2	20	ID-2012-73146	2012-10-24	2012-10-30	Standard Class	HM-14980	Henry MacAllister	Consumer	Denpasar	Bali	...	Southeast Asia	OFF-AP-10000487	Office Supplies	Appliances	Hoover Stove, Silver	1417.3578	3	0.17	-290.3022	be

3 rows × 21 columns

```
[34]: # Seleksi kolom Customer ID, Kota, Sub-Category, dan Profit. Simpan dalam variabel superstore2
superstore2 = superstore[["Customer ID", "Kota", "Sub-Category", "Profit"]]
superstore2.head(5)
```

[34]:

	Customer ID	Kota	Sub-Category	Profit
0	AI-10855	Bandung	Supplies	-84.2595
1	AI-10855	Bandung	Paper	-15.3072
2	HM-14980	Denpasar	Appliances	-290.3022


localhost:8889/lab/tree/Downloads/6.%20Library%20Pandas.ipynb
Update
Apps
Maps
YouTube
Translate
Reading List
File Edit View Run Kernel Tabs Settings Help
6. Library Pandas.ipynb
Code
Python 3
[35]:
# Seleksi superstore2, yang Kota adalah Semarang, memiliki Profit lebih dari 20. Simpan dalam superstore2_filter.
superstore2_filter = superstore2[(superstore2.Kota == "Semarang") & (superstore2.Profit > 20)]
superstore2_filter.head()
[35]:
Customer ID Kota Sub-Category Profit
16 BO-11350 Semarang Phones 25.4097
86 CA-12265 Semarang Phones 30.8190
99 CY-12745 Semarang Chairs 28.4106
184 DP-13165 Semarang Appliances 232.0011
230 MY-17380 Semarang Appliances 70.7259
[36]:
# Berapakah data yang sesuai dengan kondisi setelah dilakukan filtering?
superstore2_filter.count()
[36]:
Customer ID 16
Kota 16
Sub-Category 16
Profit 16
dtype: int64
[ ]:
Simple
No Kernel | Idle
Saving completed
Mode: Edit
Ln 1, Col 1
Superstore.xls
Show All