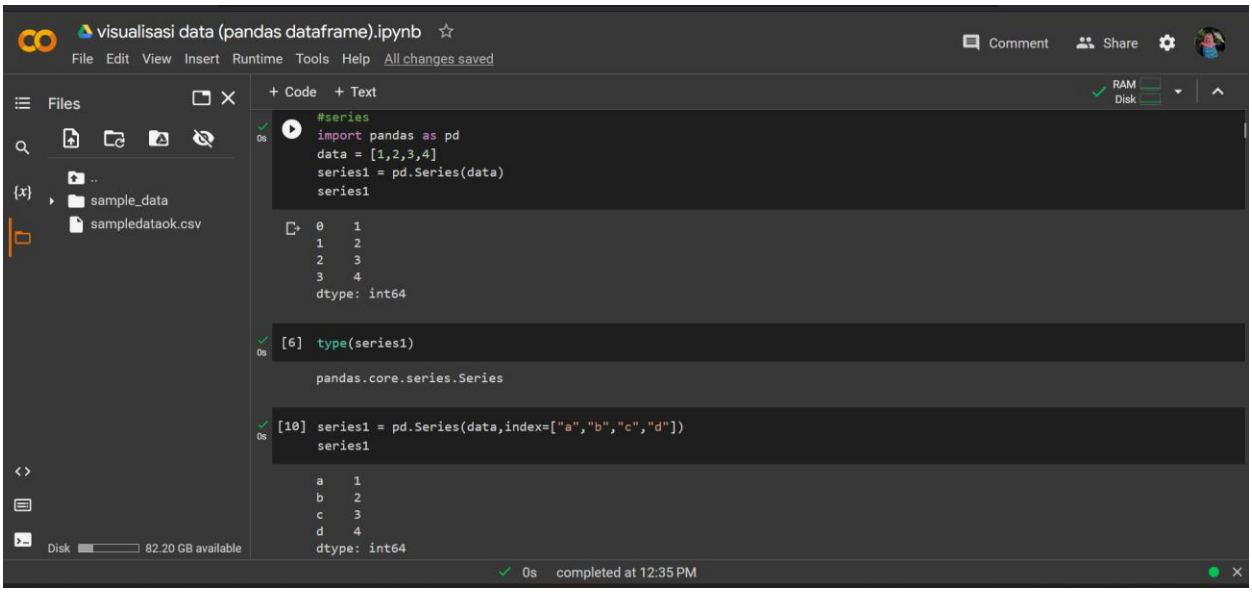
Nama : jahrawati

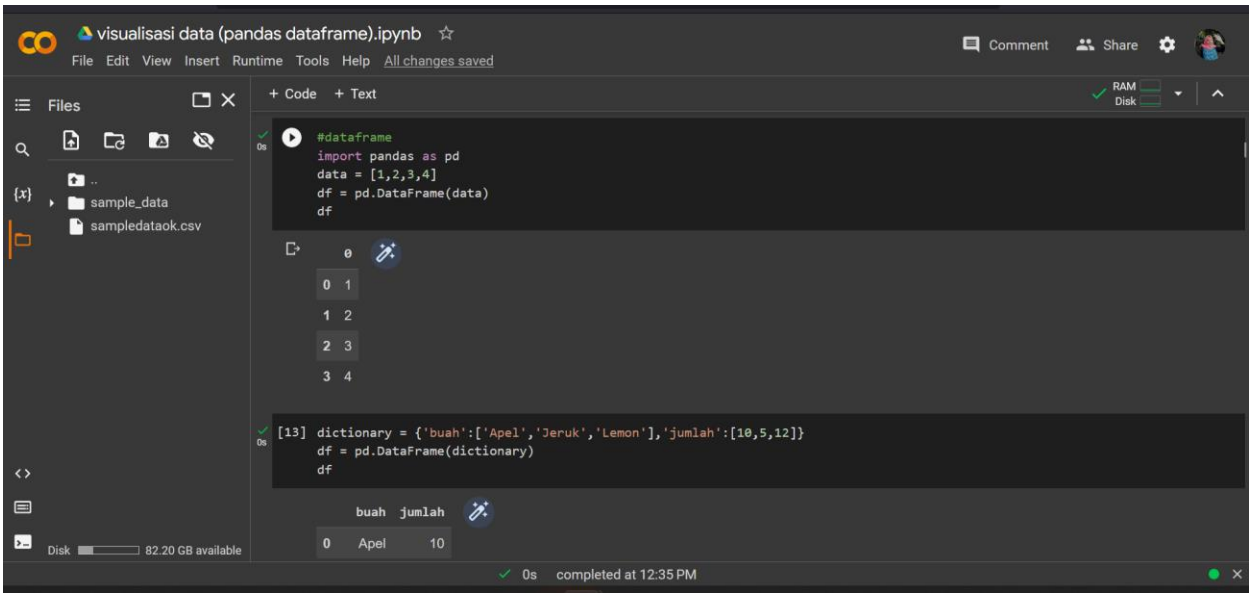
Nim : 20.01.013.050

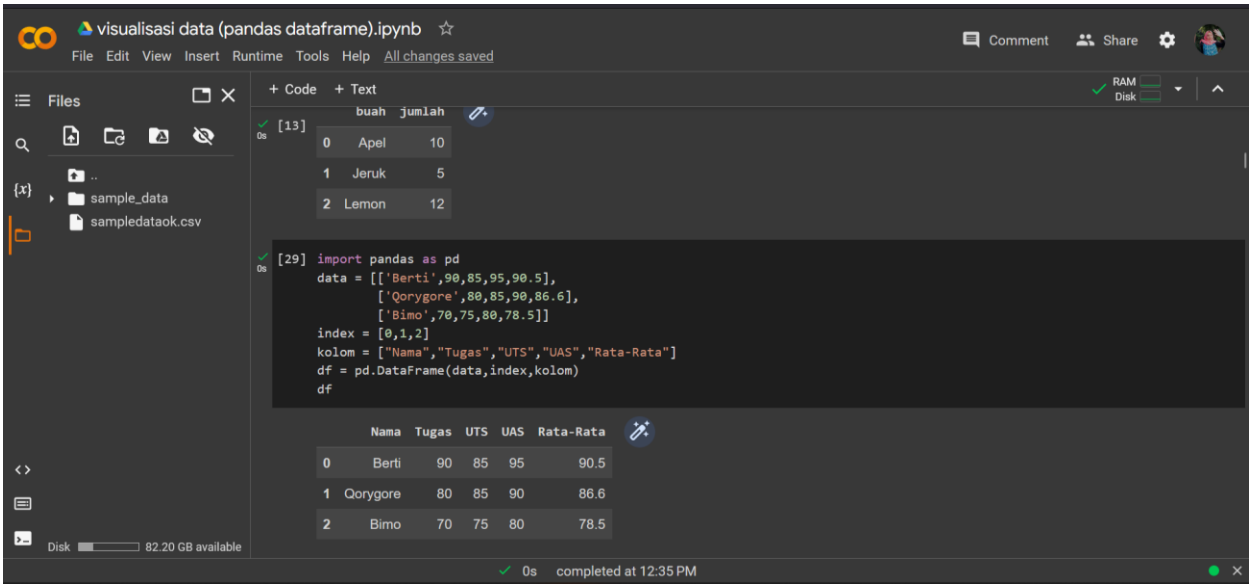
04. visualisasi data (pandas dataframe)

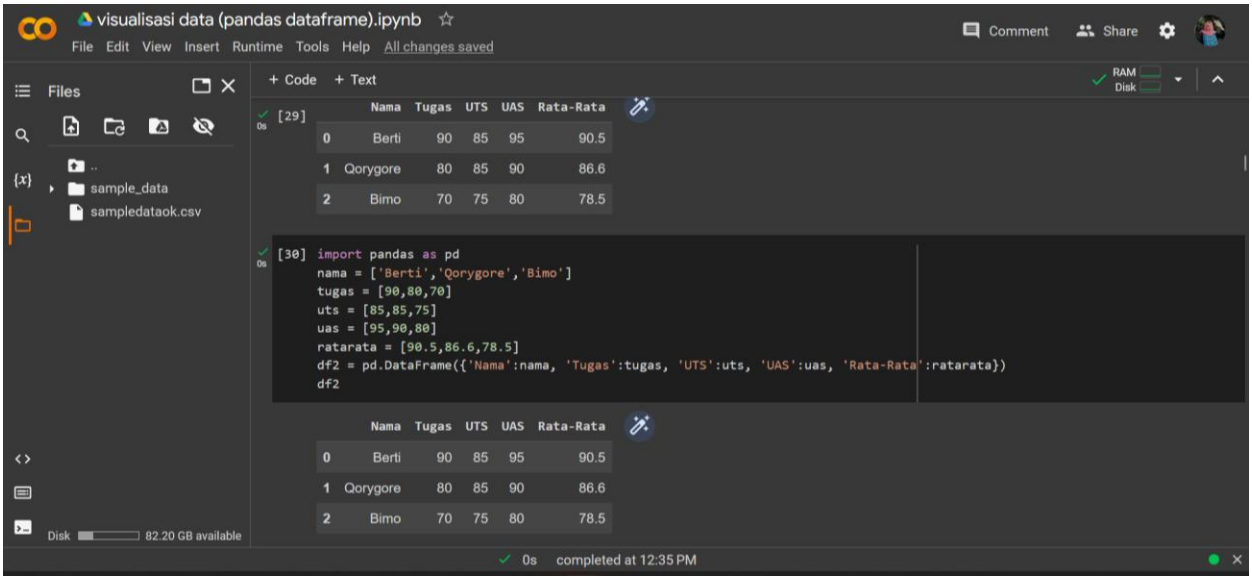
1. Series



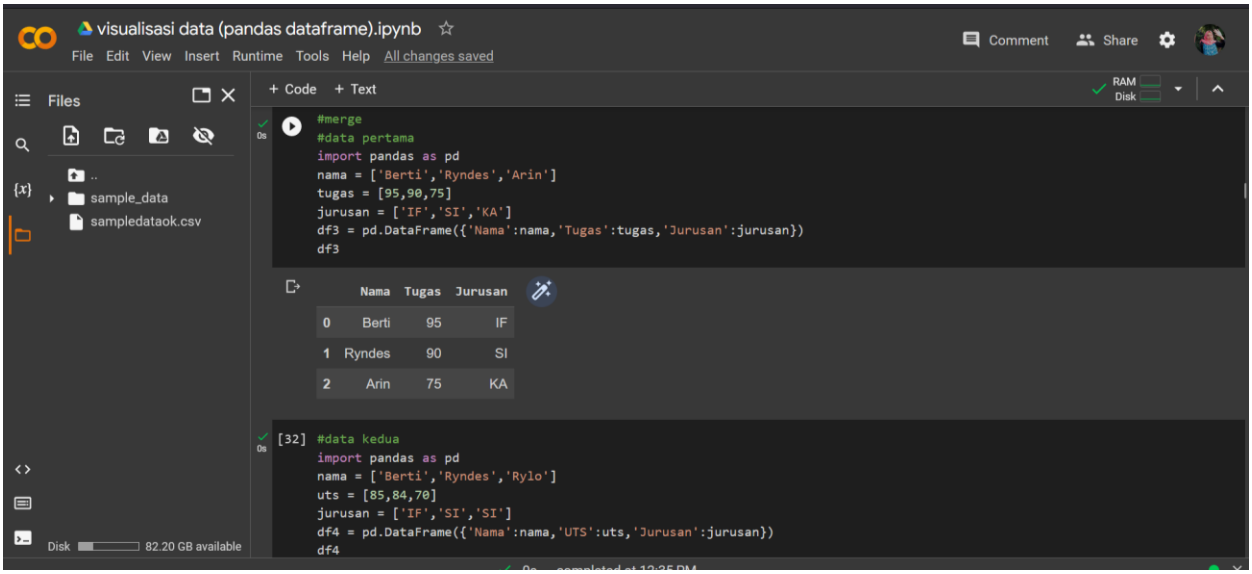

2. DataFrame

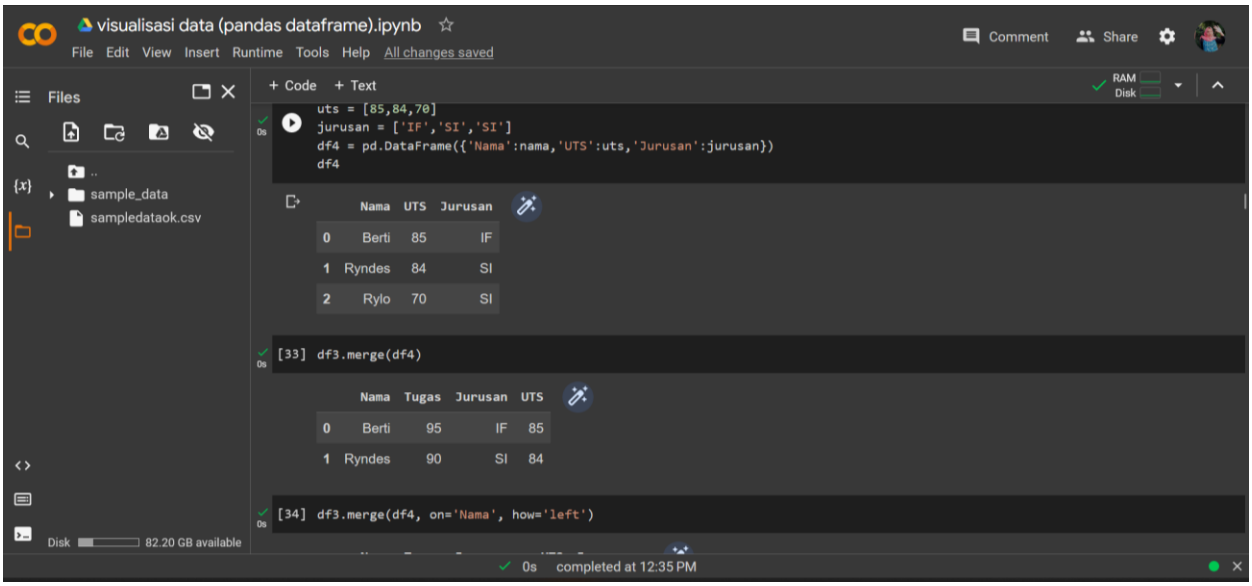


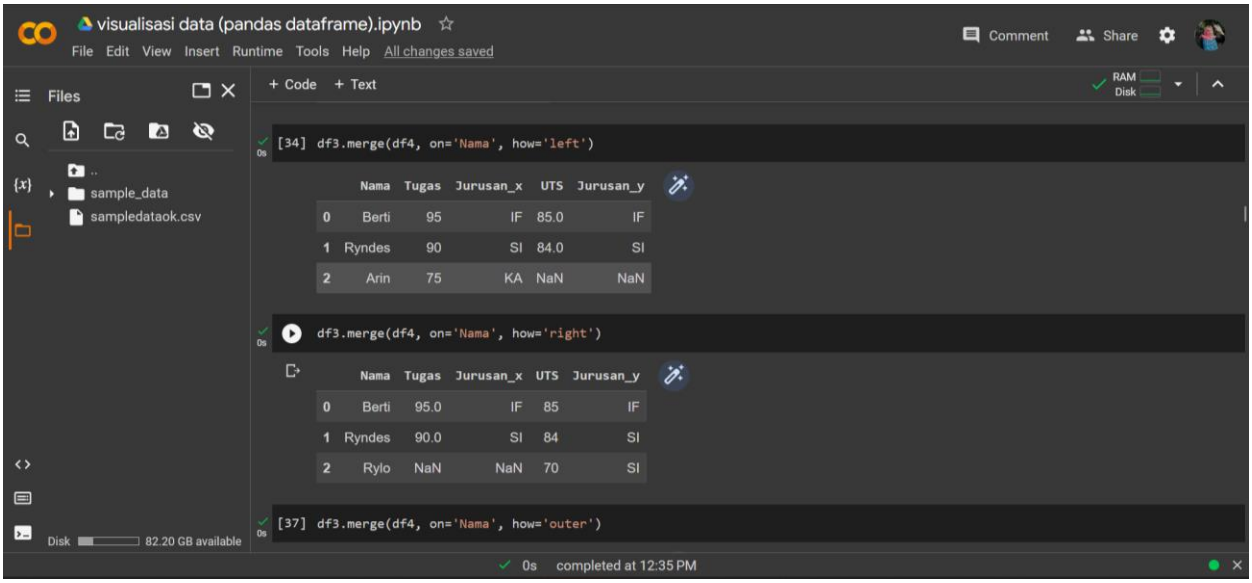

visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code
+ Text
RAM
Disk
..
sample_data
sampledataok.csv
Disk 82.20 GB available
[13]
buah jumlah
0 Apel 10
1 Jeruk 5
2 Lemon 12
[29] import pandas as pd
data = [['Berti',90,85,95,90.5],
['Qorygore',80,85,90,86.6],
['Bimo',70,75,80,78.5]]
index = [0,1,2]
kolom = ["Nama","Tugas","UTS","UAS","Rata-Rata"]
df = pd.DataFrame(data,index,kolom)
df
Nama Tugas UTS UAS Rata-Rata
0 Berti 90 85 95 90.5
1 Qorygore 80 85 90 86.6
2 Bimo 70 75 80 78.5
0s completed at 12:35 PM

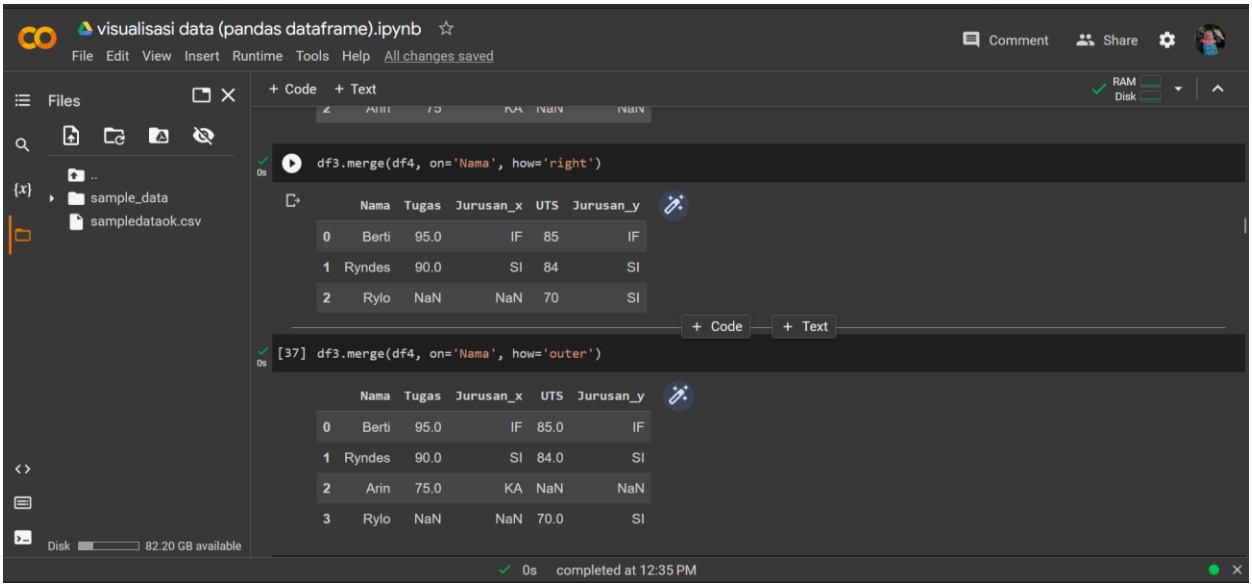

visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code
+ Text
RAM
Disk
..
sample_data
sampledataok.csv
Disk 82.20 GB available
[29]
Nama Tugas UTS UAS Rata-Rata
0 Berti 90 85 95 90.5
1 Qorygore 80 85 90 86.6
2 Bimo 70 75 80 78.5
[30] import pandas as pd
nama = ['Berti','Qorygore','Bimo']
tugas = [90,80,70]
uts = [85,85,75]
uas = [95,90,80]
ratarata = [90.5,86.6,78.5]
df2 = pd.DataFrame({'Nama':nama, 'Tugas':tugas, 'UTS':uts, 'UAS':uas, 'Rata-Rata':ratarata})
df2
Nama Tugas UTS UAS Rata-Rata
0 Berti 90 85 95 90.5
1 Qorygore 80 85 90 86.6
2 Bimo 70 75 80 78.5
0s completed at 12:35 PM

3. Merge

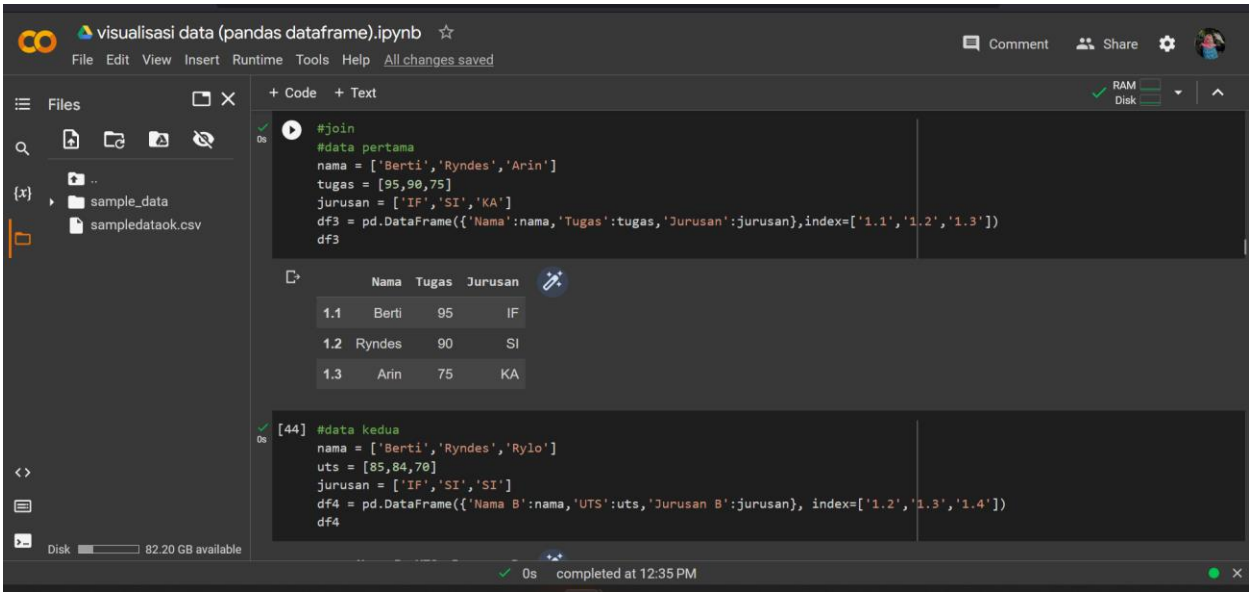

visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
sample_data
sampledataok.csv
Disk 82.20 GB available
#merge
#data pertama
import pandas as pd
nama = ['Berti','Ryndes','Arin']
tugas = [95,90,75]
jurusan = ['IF','SI','KA']
df3 = pd.DataFrame({'Nama':nama,'Tugas':tugas,'Jurusan':jurusan})
df3
Nama Tugas Jurusan
0 Berti 95 IF
1 Ryndes 90 SI
2 Arin 75 KA
[32] #data kedua
import pandas as pd
nama = ['Berti','Ryndes','Rylo']
uts = [85,84,70]
jurusan = ['IF','SI','SI']
df4 = pd.DataFrame({'Nama':nama,'UTS':uts,'Jurusan':jurusan})
df4

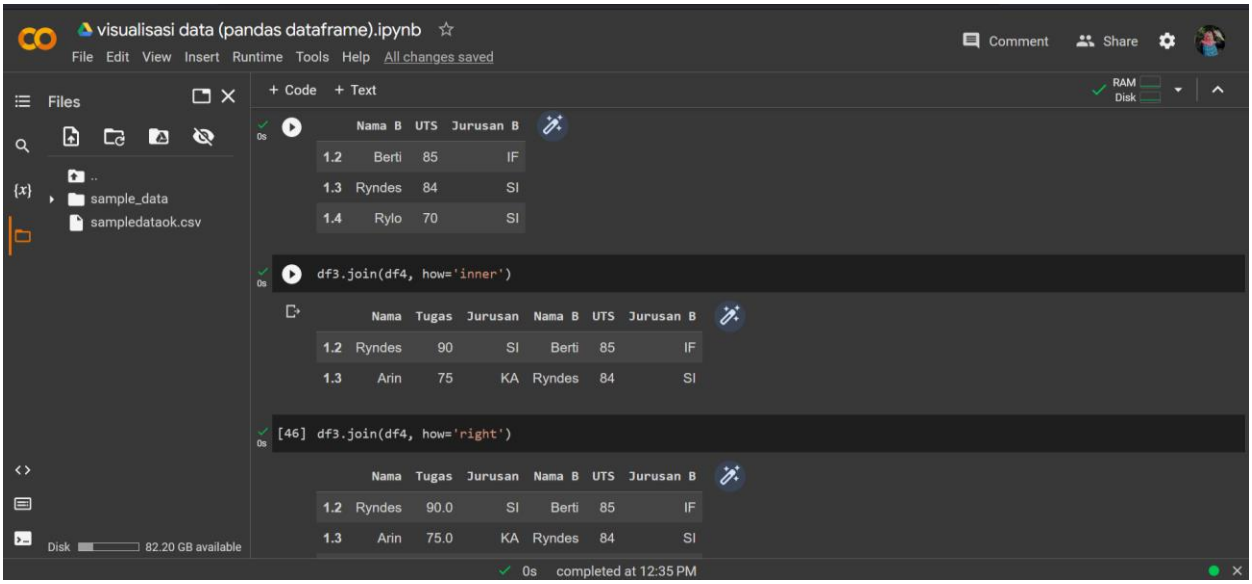

visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
sample_data
sampledataok.csv
Disk 82.20 GB available
uts = [85,84,70]
jurusan = ['IF','SI','SI']
df4 = pd.DataFrame({'Nama':nama,'UTS':uts,'Jurusan':jurusan})
df4
Nama UTS Jurusan
0 Berti 85 IF
1 Ryndes 84 SI
2 Rylo 70 SI
[33] df3.merge(df4)
Nama Tugas Jurusan UTS
0 Berti 95 IF 85
1 Ryndes 90 SI 84
[34] df3.merge(df4, on='Nama', how='left')
0s completed at 12:35 PM

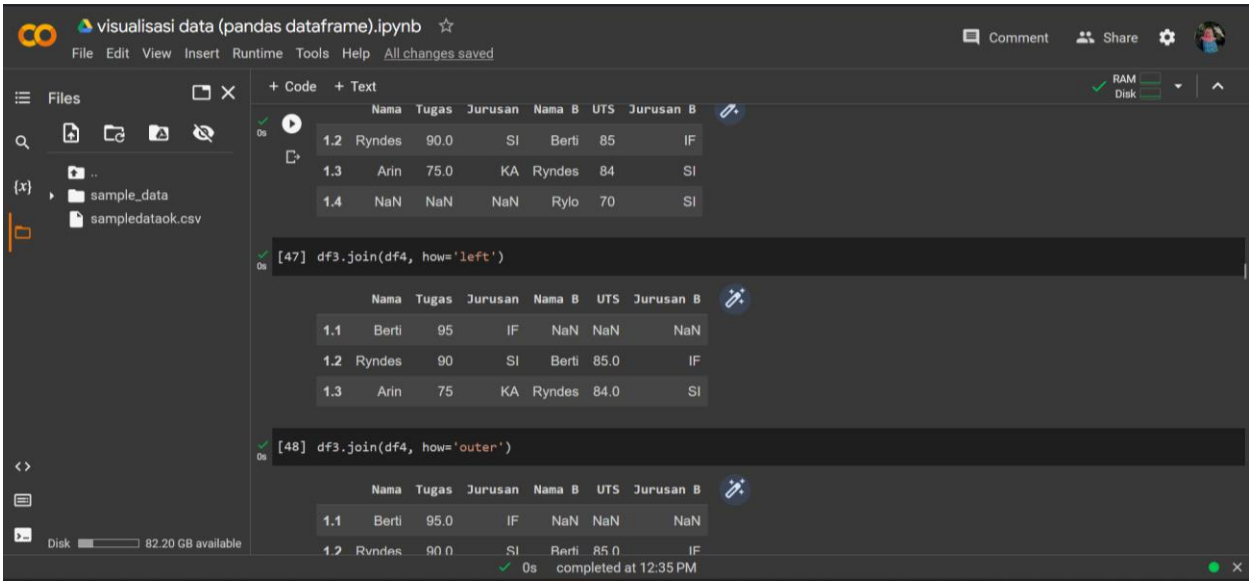

visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
RAM
Disk
Files
+ Code
+ Text
sample_data
sampledataok.csv
[34] df3.merge(df4, on='Nama', how='left')
Nama Tugas Jurusan_x UTS Jurusan_y
0 Berti 95 IF 85.0 IF
1 Ryndes 90 SI 84.0 SI
2 Arin 75 KA NaN NaN
df3.merge(df4, on='Nama', how='right')
Nama Tugas Jurusan_x UTS Jurusan_y
0 Berti 95.0 IF 85 IF
1 Ryndes 90.0 SI 84 SI
2 Rylo NaN NaN 70 SI
[37] df3.merge(df4, on='Nama', how='outer')
Disk 82.20 GB available
0s completed at 12:35 PM


visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
RAM
Disk
Files
+ Code
+ Text
sample_data
sampledataok.csv
df3.merge(df4, on='Nama', how='right')
Nama Tugas Jurusan_x UTS Jurusan_y
0 Berti 95.0 IF 85 IF
1 Ryndes 90.0 SI 84 SI
2 Rylo NaN NaN 70 SI
+ Code
+ Text
[37] df3.merge(df4, on='Nama', how='outer')
Nama Tugas Jurusan_x UTS Jurusan_y
0 Berti 95.0 IF 85.0 IF
1 Ryndes 90.0 SI 84.0 SI
2 Arin 75.0 KA NaN NaN
3 Rylo NaN NaN 70.0 SI
Disk 82.20 GB available
0s completed at 12:35 PM

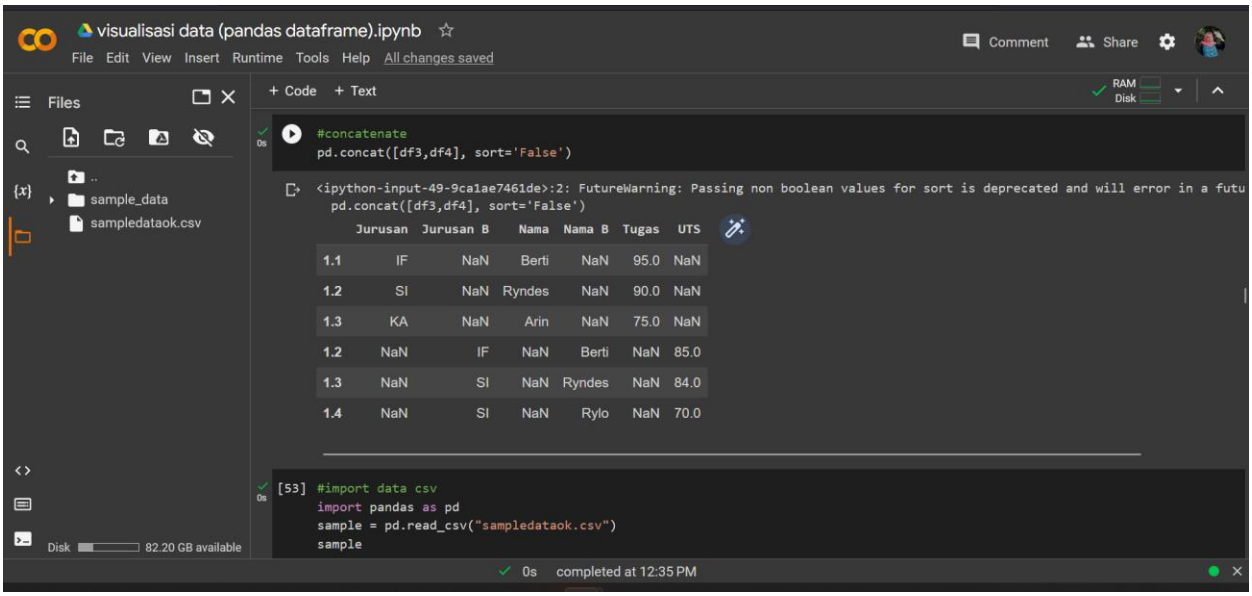
4. Join

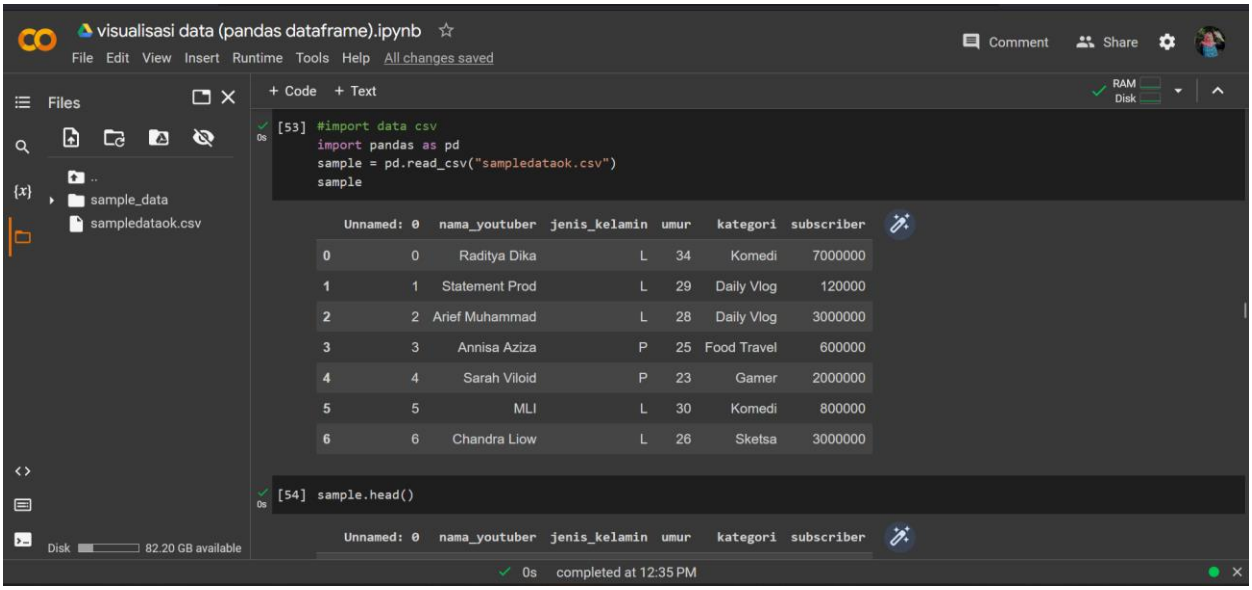

visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code
+ Text
RAM
Disk
sample_data
sampledataok.csv
Disk 82.20 GB available
#join
#data pertama
nama = ['Berti','Ryndes','Arin']
tugas = [95,90,75]
jurusan = ['IF','SI','KA']
df3 = pd.DataFrame({'Nama':nama,'Tugas':tugas,'Jurusan':jurusan},index=['1.1','1.2','1.3'])
df3
Nama Tugas Jurusan
1.1 Berti 95 IF
1.2 Ryndes 90 SI
1.3 Arin 75 KA
[44] #data kedua
nama = ['Berti','Ryndes','Rylo']
uts = [85,84,70]
jurusan = ['IF','SI','SI']
df4 = pd.DataFrame({'Nama B':nama,'UTS':uts,'Jurusan B':jurusan}, index=['1.2','1.3','1.4'])
df4
0s completed at 12:35 PM

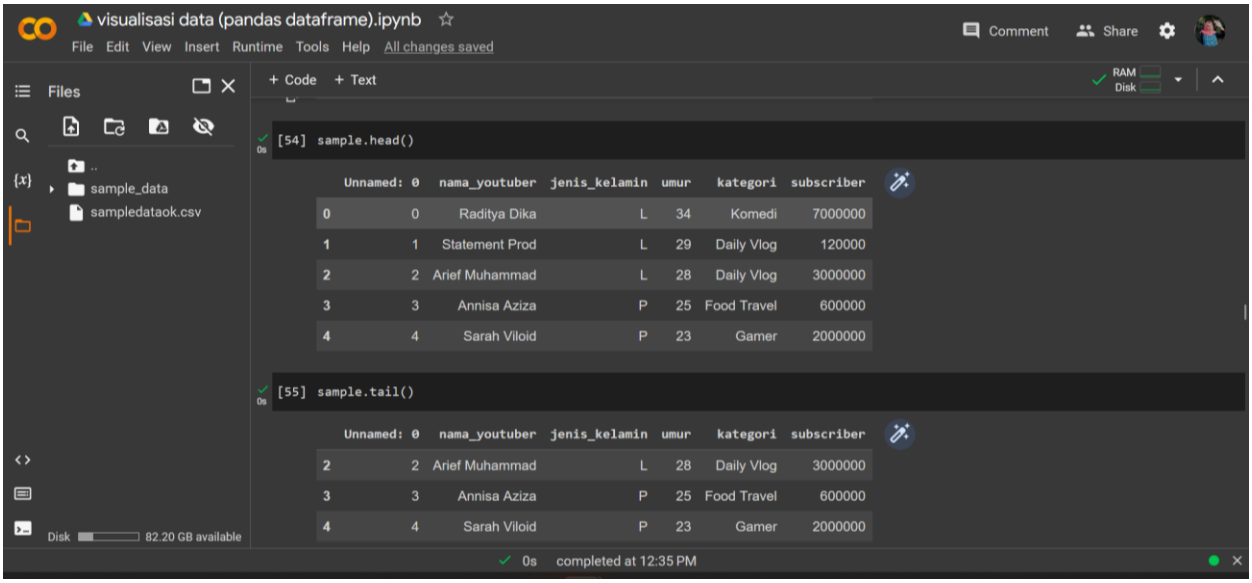

visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code
+ Text
RAM
Disk
sample_data
sampledataok.csv
Disk 82.20 GB available
Nama B UTS Jurusan B
1.2 Berti 85 IF
1.3 Ryndes 84 SI
1.4 Rylo 70 SI
df3.join(df4, how='inner')
Nama Tugas Jurusan Nama B UTS Jurusan B
1.2 Ryndes 90 SI Berti 85 IF
1.3 Arin 75 KA Ryndes 84 SI
[46] df3.join(df4, how='right')
Nama Tugas Jurusan Nama B UTS Jurusan B
1.2 Ryndes 90.0 SI Berti 85 IF
1.3 Arin 75.0 KA Ryndes 84 SI
0s completed at 12:35 PM

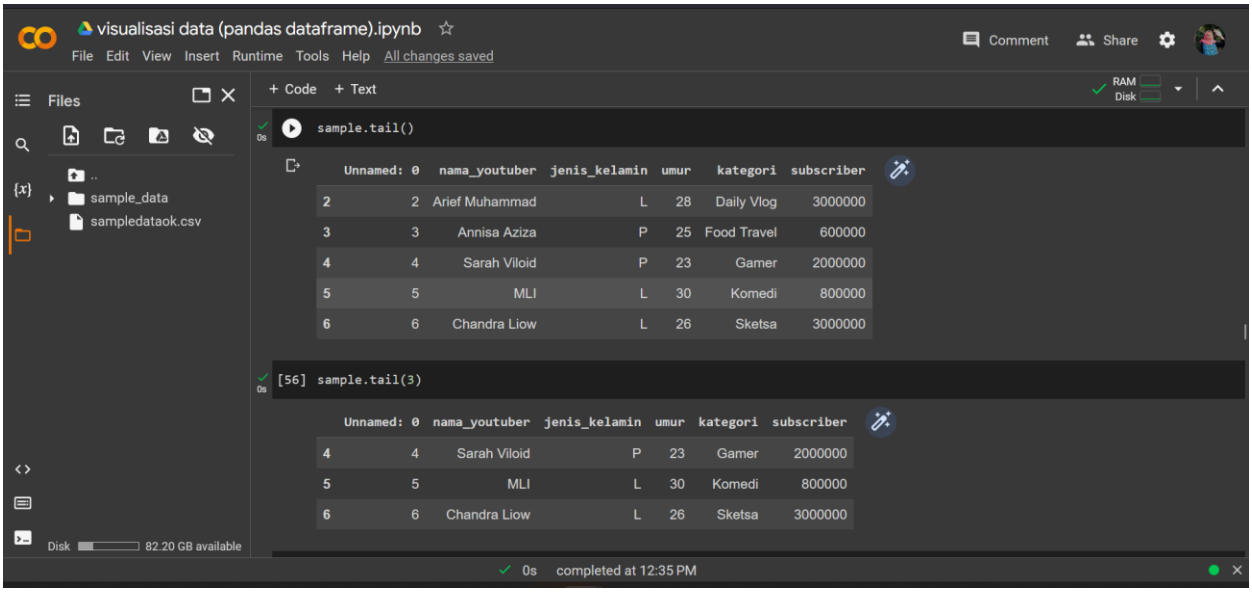


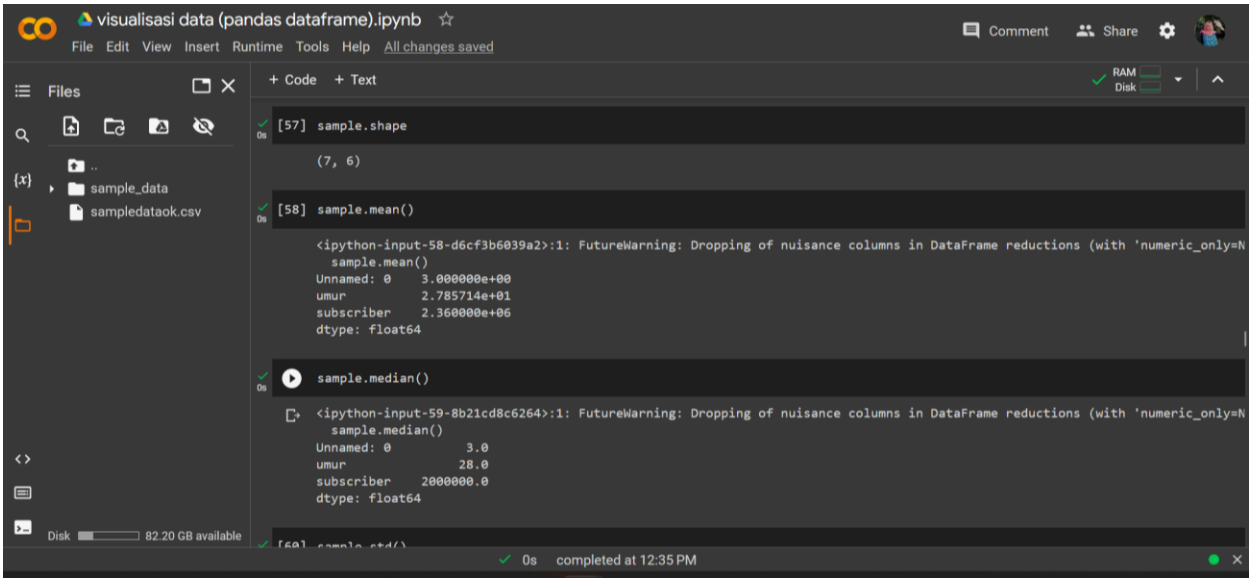

**5.** Concatenate

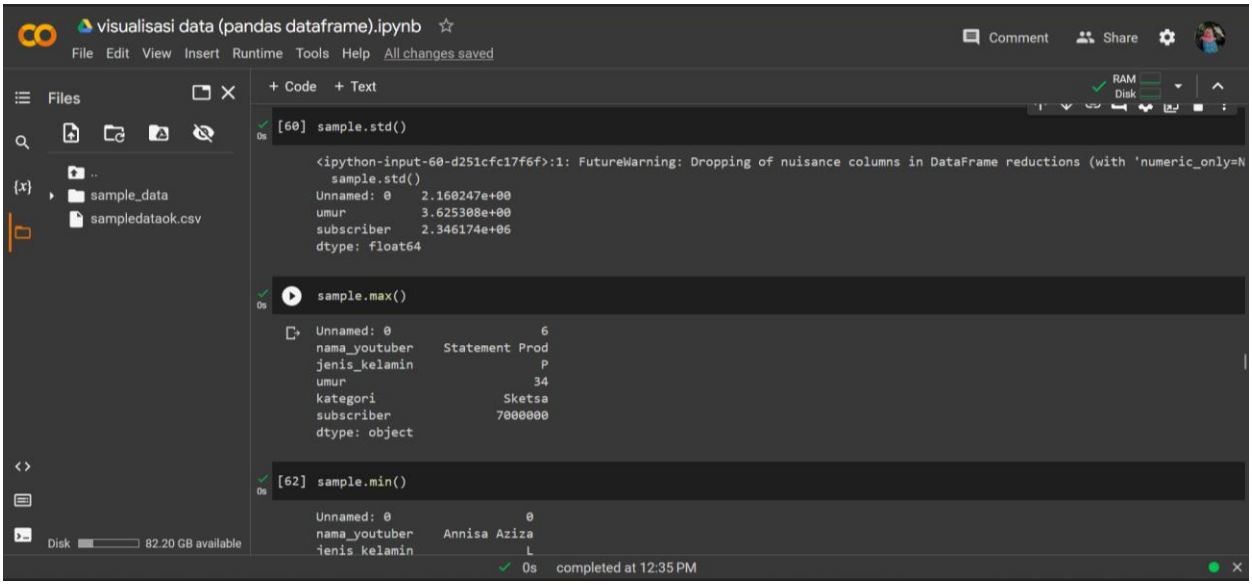


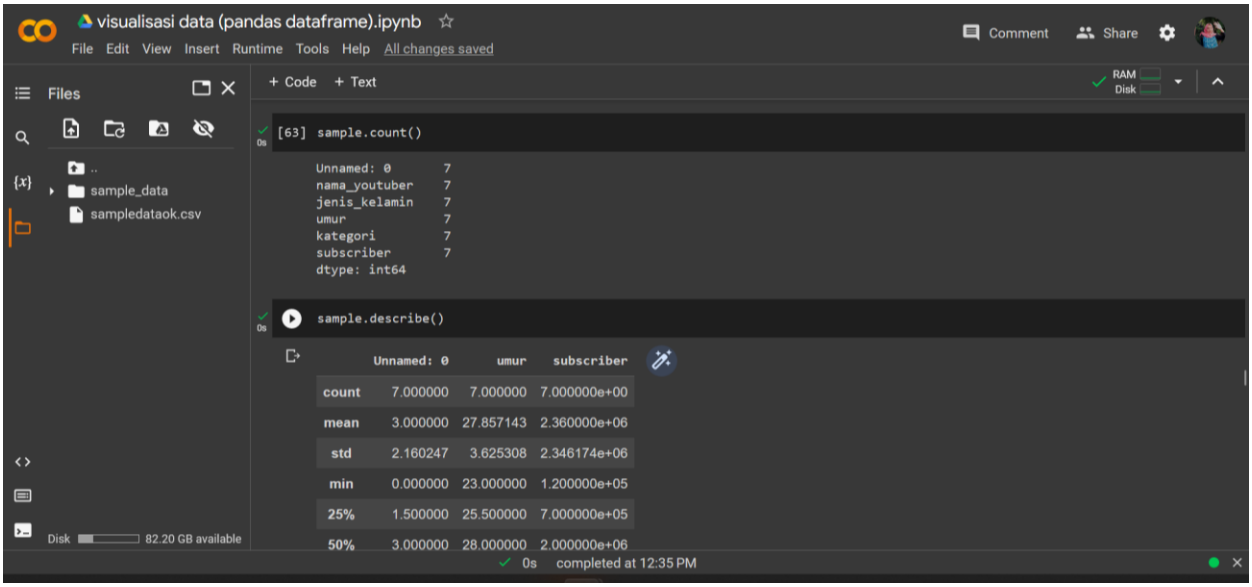

visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
..
sample_data
sampledataok.csv
[53] #import data csv
import pandas as pd
sample = pd.read_csv("sampledataok.csv")
sample
Unnamed: 0 nama_youtuber jenis_kelamin umur kategori subscriber
0 0 Raditya Dika L 34 Komedi 7000000
1 1 Statement Prod L 29 Daily Vlog 120000
2 2 Arief Muhammad L 28 Daily Vlog 3000000
3 3 Annisa Aziza P 25 Food Travel 600000
4 4 Sarah Viloid P 23 Gamer 2000000
5 5 MLI L 30 Komedi 800000
6 6 Chandra Liow L 26 Sketsa 3000000
[54] sample.head()
Unnamed: 0 nama_youtuber jenis_kelamin umur kategori subscriber
Disk 82.20 GB available
0s completed at 12:35 PM

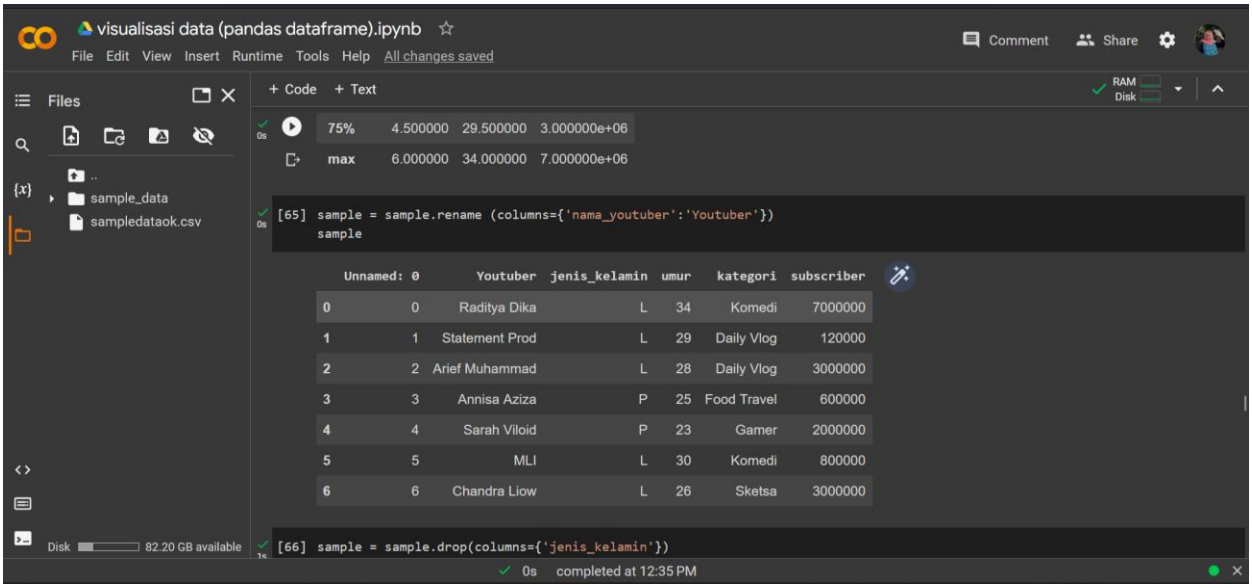

visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
..
sample_data
sampledataok.csv
[54] sample.head()
Unnamed: 0 nama_youtuber jenis_kelamin umur kategori subscriber
0 0 Raditya Dika L 34 Komedi 7000000
1 1 Statement Prod L 29 Daily Vlog 120000
2 2 Arief Muhammad L 28 Daily Vlog 3000000
3 3 Annisa Aziza P 25 Food Travel 600000
4 4 Sarah Viloid P 23 Gamer 2000000
[55] sample.tail()
Unnamed: 0 nama_youtuber jenis_kelamin umur kategori subscriber
2 2 Arief Muhammad L 28 Daily Vlog 3000000
3 3 Annisa Aziza P 25 Food Travel 600000
4 4 Sarah Viloid P 23 Gamer 2000000
Disk 82.20 GB available
0s completed at 12:35 PM

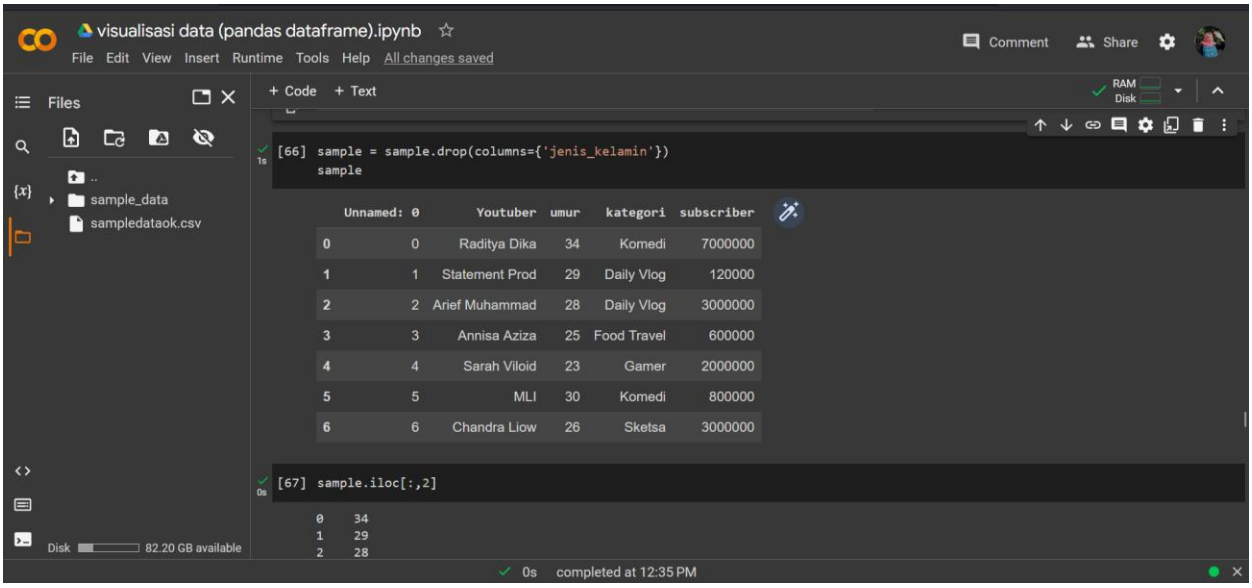

visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
sample.tail()
Unnamed: 0 nama_youtuber jenis_kelamin umur kategori subscriber
2 2 Arief Muhammad L 28 Daily Vlog 3000000
3 3 Annisa Aziza P 25 Food Travel 600000
4 4 Sarah Viloid P 23 Gamer 2000000
5 5 MLI L 30 Komedi 800000
6 6 Chandra Liow L 26 Sketsa 3000000
[56] sample.tail(3)
Unnamed: 0 nama_youtuber jenis_kelamin umur kategori subscriber
4 4 Sarah Viloid P 23 Gamer 2000000
5 5 MLI L 30 Komedi 800000
6 6 Chandra Liow L 26 Sketsa 3000000
sample_data
sampledataok.csv
Disk 82.20 GB available
0s completed at 12:35 PM

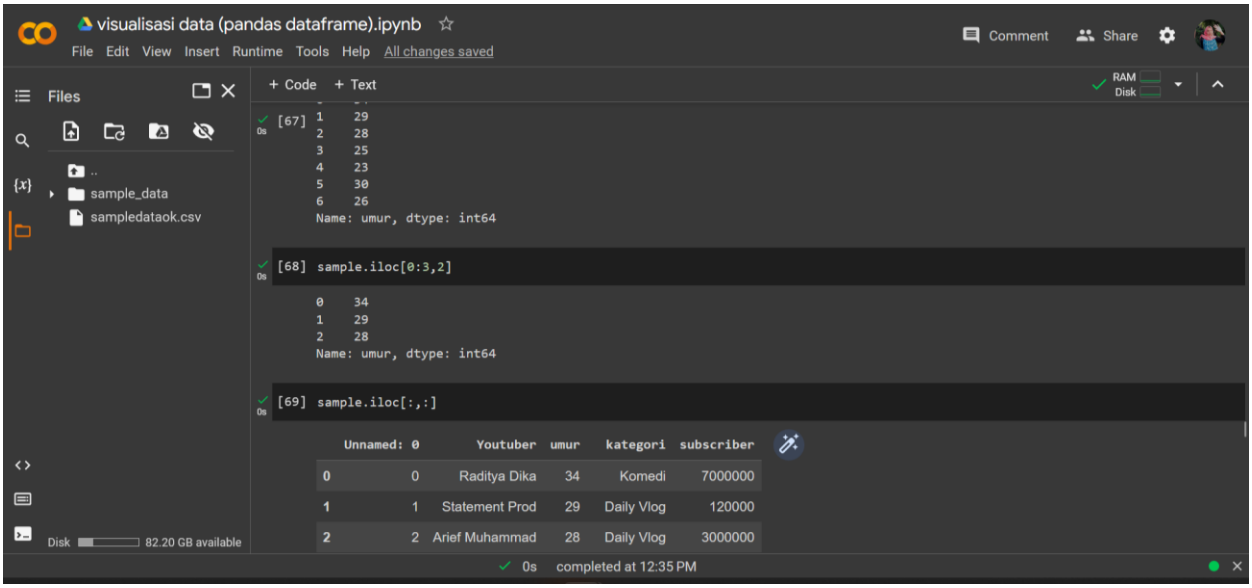

visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
[57] sample.shape
(7, 6)
[58] sample.mean()
<ipython-input-58-d6cf3b6039a2>:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=N
sample.mean()
Unnamed: 0 3.000000e+00
umur 2.785714e+01
subscriber 2.360000e+06
dtype: float64
sample.median()
<ipython-input-59-8b21cd8c6264>:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=N
sample.median()
Unnamed: 0 3.0
umur 28.0
subscriber 2000000.0
dtype: float64
sample_data
sampledataok.csv
Disk 82.20 GB available
0s completed at 12:35 PM

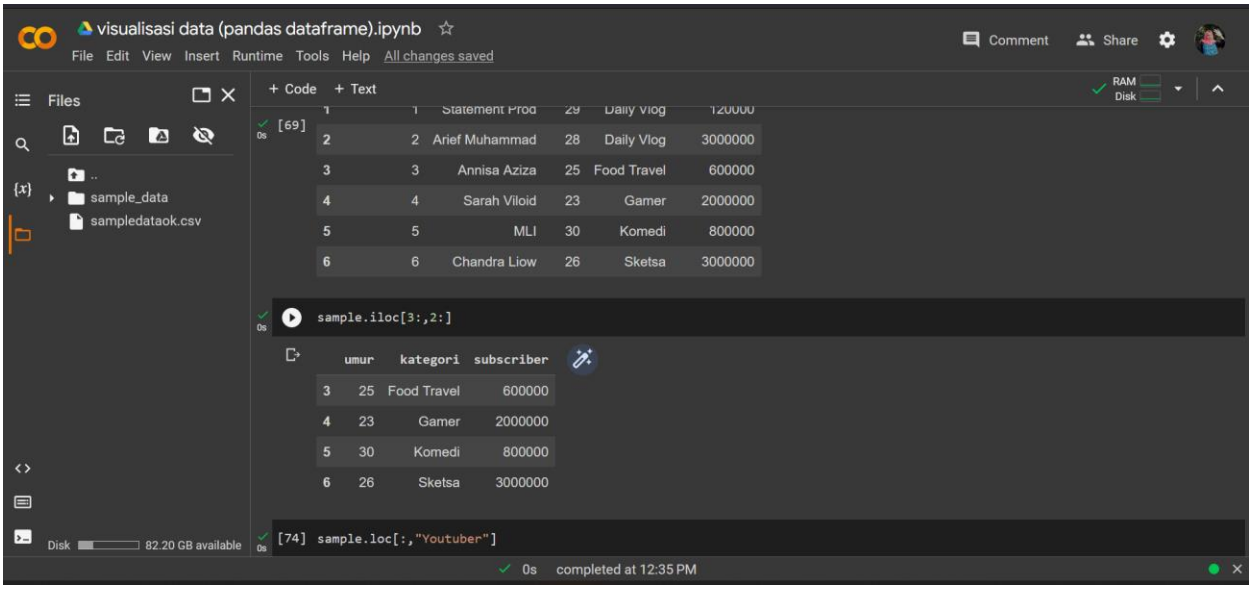

visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
[60] sample.std()
<ipython-input-60-d251cfc17f6f>:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=N
sample.std()
Unnamed: 0 2.160247e+00
umur 3.625308e+00
subscriber 2.346174e+06
dtype: float64
sample.max()
Unnamed: 0 6
nama_youtuber Statement Prod
jenis_kelamin P
umur 34
kategori Sketsa
subscriber 7000000
dtype: object
[62] sample.min()
Unnamed: 0 0
nama_youtuber Annisa Aziza
jenis_kelamin L
sample_data
sampledataok.csv
Disk 82.20 GB available
0s completed at 12:35 PM

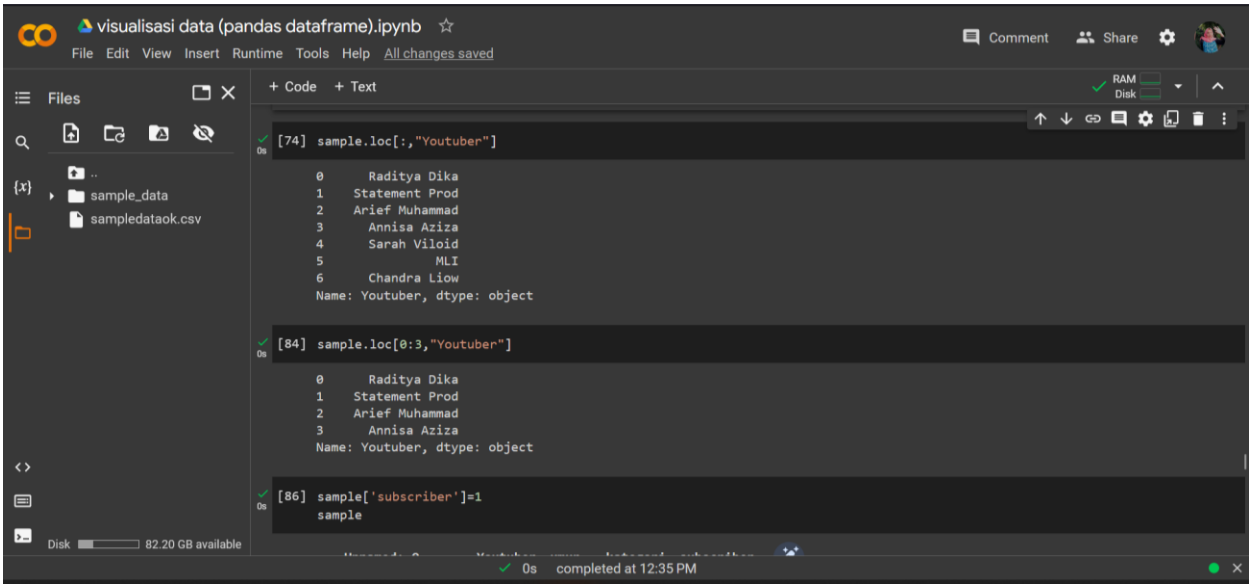

visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
[63] sample.count()
Unnamed: 0 7
nama_youtuber 7
jenis_kelamin 7
umur 7
kategori 7
subscriber 7
dtype: int64
sample.describe()
Unnamed: 0 umur subscriber
count 7.000000 7.000000 7.000000e+00
mean 3.000000 27.857143 2.360000e+06
std 2.160247 3.625308 2.346174e+06
min 0.000000 23.000000 1.200000e+05
25% 1.500000 25.500000 7.000000e+05
50% 3.000000 28.000000 2.000000e+06
sample_data
sampledataok.csv
Disk 82.20 GB available
0s completed at 12:35 PM

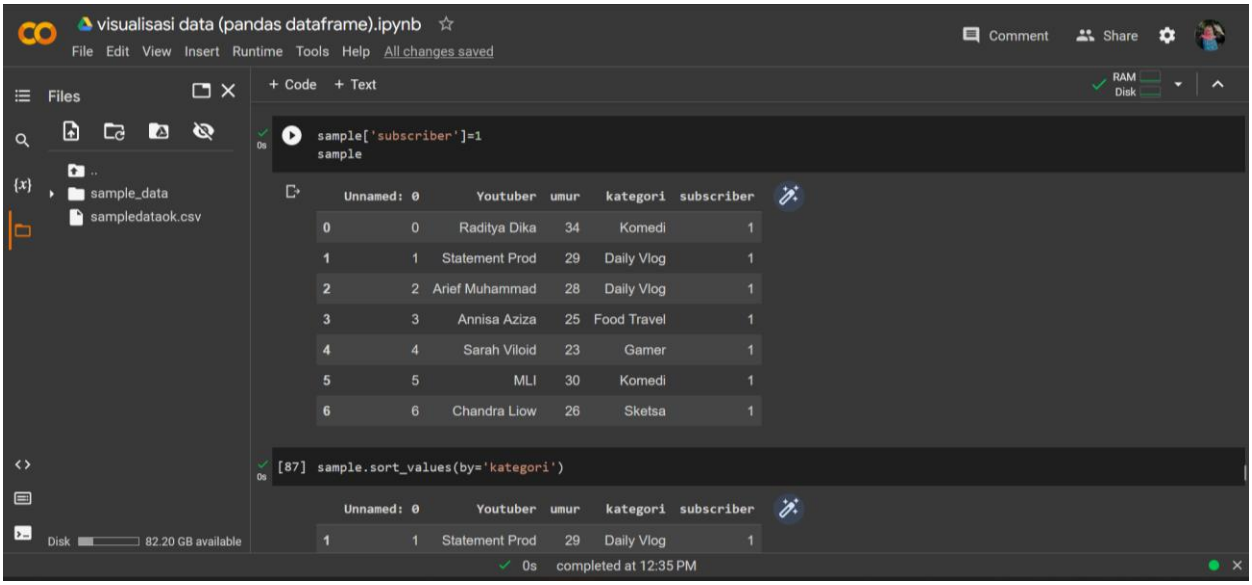

visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code
+ Text
RAM
Disk
75% 4.500000 29.500000 3.000000e+06
max 6.000000 34.000000 7.000000e+06
[65] sample = sample.rename (columns={'nama_youtuber':'Youtuber'})
sample
Unnamed: 0 Youtuber jenis_kelamin umur kategori subscriber
0 0 Raditya Dika L 34 Komedi 7000000
1 1 Statement Prod L 29 Daily Vlog 120000
2 2 Arief Muhammad L 28 Daily Vlog 3000000
3 3 Annisa Aziza P 25 Food Travel 600000
4 4 Sarah Viloid P 23 Gamer 2000000
5 5 MLI L 30 Komedi 800000
6 6 Chandra Liow L 26 Sketsa 3000000
[66] sample = sample.drop(columns={'jenis_kelamin'})
sample_data
sampledataok.csv
Disk 82.20 GB available
0s completed at 12:35 PM

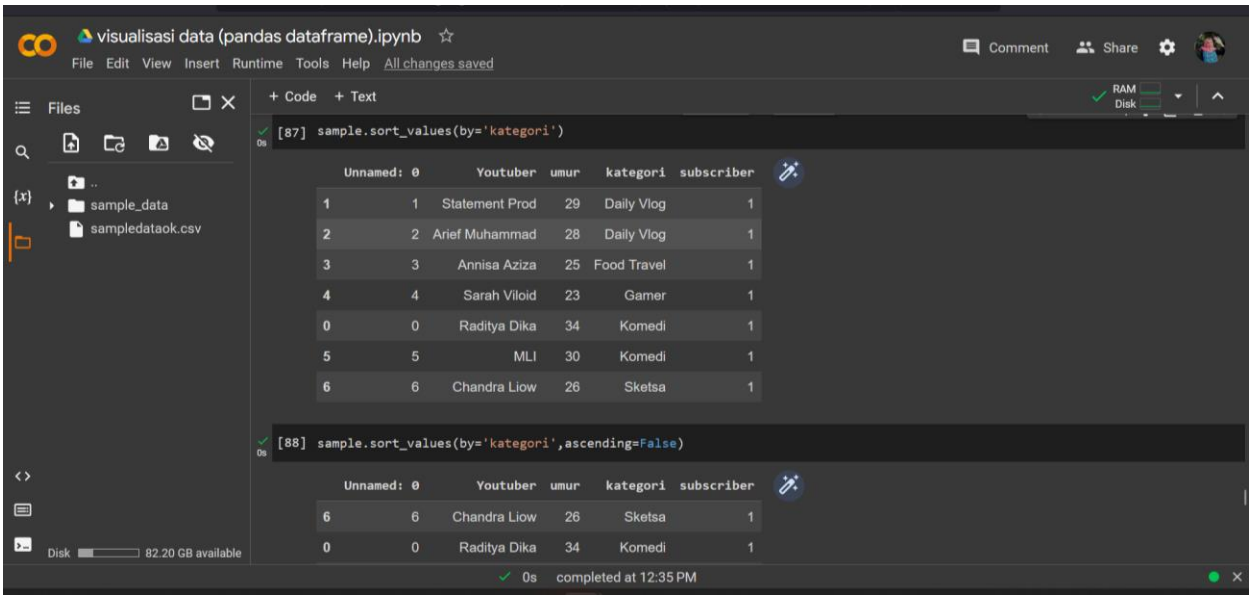

visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code
+ Text
RAM
Disk
[66] sample = sample.drop(columns={'jenis_kelamin'})
sample
Unnamed: 0 Youtuber umur kategori subscriber
0 0 Raditya Dika 34 Komedi 7000000
1 1 Statement Prod 29 Daily Vlog 120000
2 2 Arief Muhammad 28 Daily Vlog 3000000
3 3 Annisa Aziza 25 Food Travel 600000
4 4 Sarah Viloid 23 Gamer 2000000
5 5 MLI 30 Komedi 800000
6 6 Chandra Liow 26 Sketsa 3000000
[67] sample.iloc[:,2]
0 34
1 29
2 28
sample_data
sampledataok.csv
Disk 82.20 GB available
0s completed at 12:35 PM

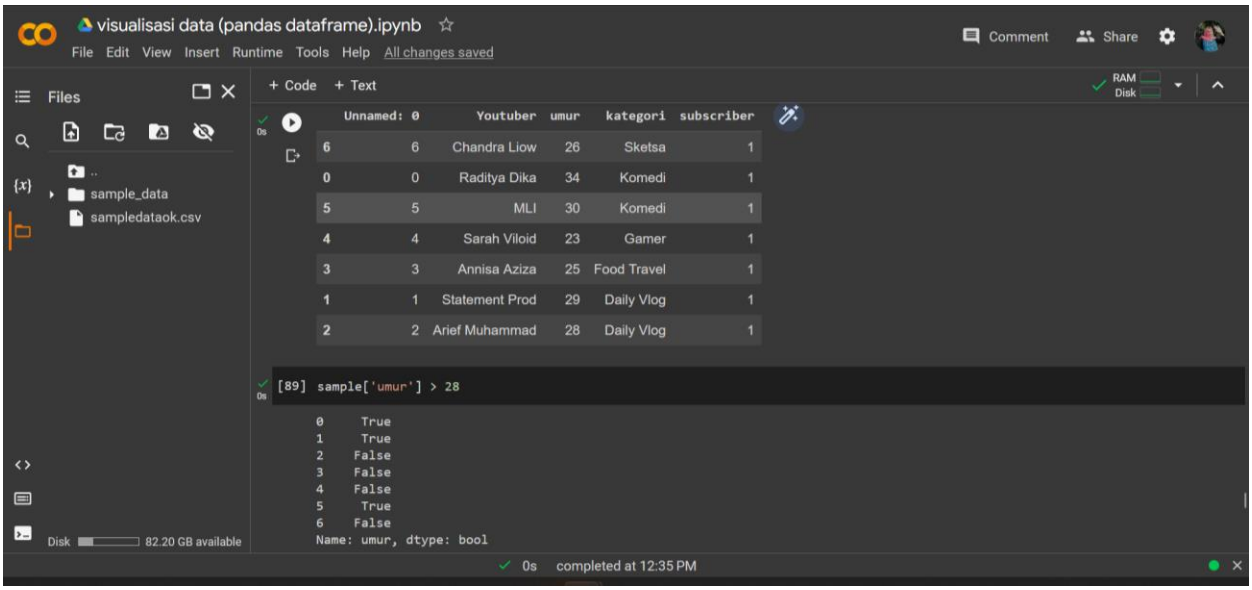

visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code
+ Text
RAM
Disk
sample_data
sampledataok.csv
[67] 1 29
2 28
3 25
4 23
5 30
6 26
Name: umur, dtype: int64
[68] sample.iloc[0:3,2]
0 34
1 29
2 28
Name: umur, dtype: int64
[69] sample.iloc[:,:]
Unnamed: 0 Youtuber umur kategori subscriber
0 0 Raditya Dika 34 Komedi 7000000
1 1 Statement Prod 29 Daily Vlog 120000
2 2 Arief Muhammad 28 Daily Vlog 3000000
Disk 82.20 GB available
0s completed at 12:35 PM

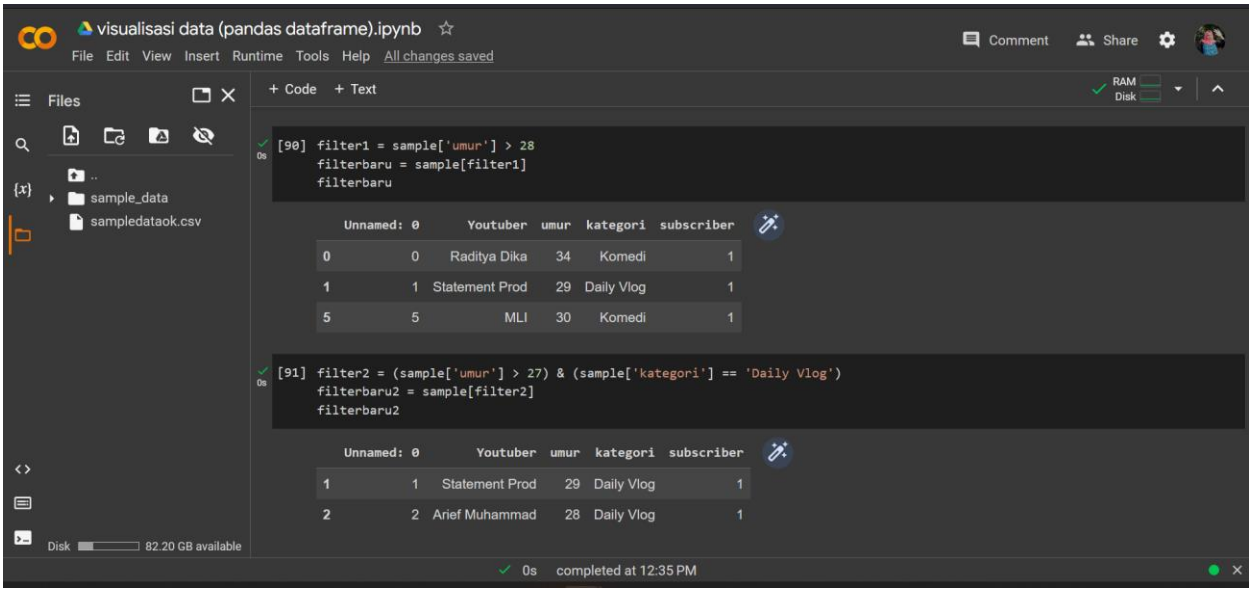

visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code
+ Text
RAM
Disk
sample_data
sampledataok.csv
[69]
2 2 Arief Muhammad 28 Daily Vlog 3000000
3 3 Annisa Aziza 25 Food Travel 600000
4 4 Sarah Viloid 23 Gamer 2000000
5 5 MLI 30 Komedi 800000
6 6 Chandra Liow 26 Sketsa 3000000
sample.iloc[3:,2:]
umur kategori subscriber
3 25 Food Travel 600000
4 23 Gamer 2000000
5 30 Komedi 800000
6 26 Sketsa 3000000
[74] sample.loc[:,"Youtuber"]
Disk 82.20 GB available
0s completed at 12:35 PM


visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
sample_data
sampledataok.csv
[74] sample.loc[:,"Youtuber"]
0 Raditya Dika
1 Statement Prod
2 Arief Muhammad
3 Annisa Aziza
4 Sarah Viloid
5 MLI
6 Chandra Liow
Name: Youtuber, dtype: object
[84] sample.loc[0:3,"Youtuber"]
0 Raditya Dika
1 Statement Prod
2 Arief Muhammad
3 Annisa Aziza
Name: Youtuber, dtype: object
[86] sample['subscriber']=1
sample
Disk 82.20 GB available
0s completed at 12:35 PM


visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
sample_data
sampledataok.csv
sample['subscriber']=1
sample
Unnamed: 0 Youtuber umur kategori subscriber
0 0 Raditya Dika 34 Komedi 1
1 1 Statement Prod 29 Daily Vlog 1
2 2 Arief Muhammad 28 Daily Vlog 1
3 3 Annisa Aziza 25 Food Travel 1
4 4 Sarah Viloid 23 Gamer 1
5 5 MLI 30 Komedi 1
6 6 Chandra Liow 26 Sketsa 1
[87] sample.sort_values(by='kategori')
Unnamed: 0 Youtuber umur kategori subscriber
1 1 Statement Prod 29 Daily Vlog 1
Disk 82.20 GB available
0s completed at 12:35 PM


visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
Files
+ Code + Text
RAM Disk
[87] sample.sort_values(by='kategori')
Unnamed: 0 Youtuber umur kategori subscriber
1 1 Statement Prod 29 Daily Vlog 1
2 2 Arief Muhammad 28 Daily Vlog 1
3 3 Annisa Aziza 25 Food Travel 1
4 4 Sarah Viloid 23 Gamer 1
0 0 Raditya Dika 34 Komedi 1
5 5 MLI 30 Komedi 1
6 6 Chandra Liow 26 Sketsa 1
[88] sample.sort_values(by='kategori',ascending=False)
Unnamed: 0 Youtuber umur kategori subscriber
6 6 Chandra Liow 26 Sketsa 1
0 0 Raditya Dika 34 Komedi 1
sample_data
sampledataok.csv
Disk 82.20 GB available
0s completed at 12:35 PM


visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
Files
+ Code + Text
RAM Disk
Unnamed: 0 Youtuber umur kategori subscriber
6 6 Chandra Liow 26 Sketsa 1
0 0 Raditya Dika 34 Komedi 1
5 5 MLI 30 Komedi 1
4 4 Sarah Viloid 23 Gamer 1
3 3 Annisa Aziza 25 Food Travel 1
1 1 Statement Prod 29 Daily Vlog 1
2 2 Arief Muhammad 28 Daily Vlog 1
[89] sample['umur'] > 28
0 True
1 True
2 False
3 False
4 False
5 True
6 False
Name: umur, dtype: bool
sample_data
sampledataok.csv
Disk 82.20 GB available
0s completed at 12:35 PM


visualisasi data (pandas dataframe).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code
+ Text
RAM
Disk
..
sample_data
sampledataok.csv
[90] filter1 = sample['umur'] > 28
filterbaru = sample[filter1]
filterbaru
Unnamed: 0 Youtuber umur kategori subscriber
0 0 Raditya Dika 34 Komedi 1
1 1 Statement Prod 29 Daily Vlog 1
5 5 MLI 30 Komedi 1
[91] filter2 = (sample['umur'] > 27) & (sample['kategori'] == 'Daily Vlog')
filterbaru2 = sample[filter2]
filterbaru2
Unnamed: 0 Youtuber umur kategori subscriber
1 1 Statement Prod 29 Daily Vlog 1
2 2 Arief Muhammad 28 Daily Vlog 1
Disk 82.20 GB available
0s completed at 12:35 PM