02_PANDAS

Monday, 14 June, 2021
09:14 PM
Deepankar Sharma

Pandas

Pandas is a very popular Python library built on top of NumPy.
Pandas share many functions and data structures offered within NumPy.
Pandas is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals.

```
# importing the pandas module
import pandas as pd
print(pd.__version__)
```

Series and DataFrame

The two primary components of pandas are the **Series** and the **DataFrame**.

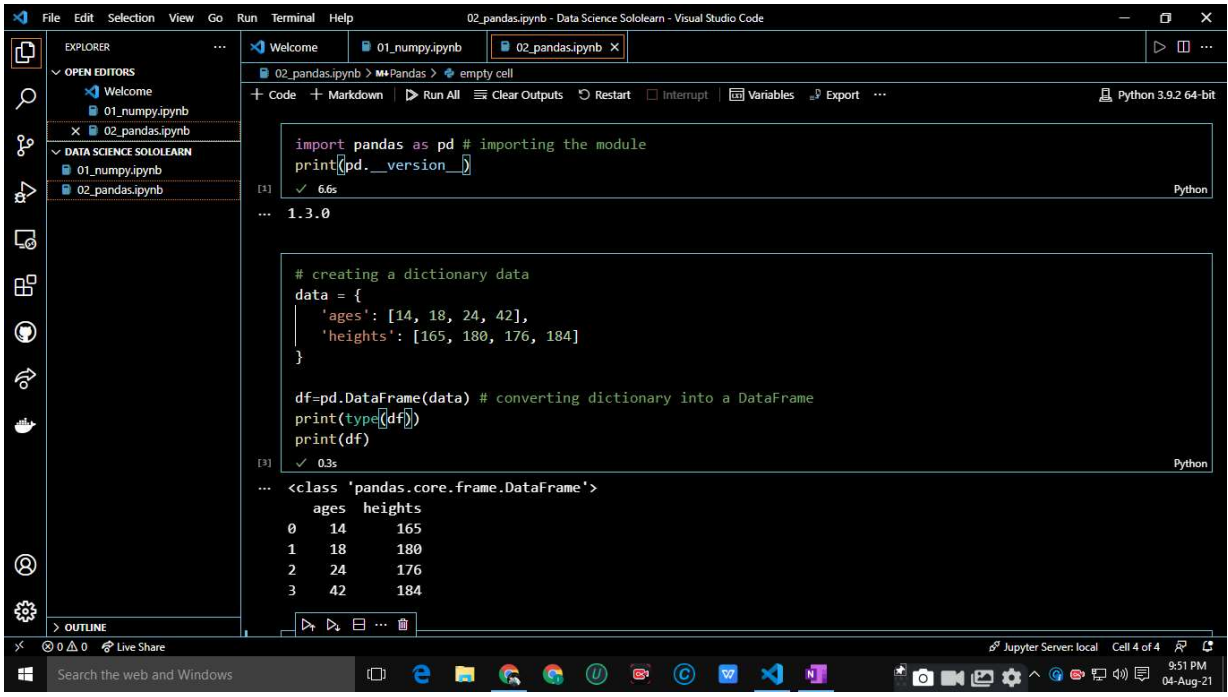
A **Series** is essentially a column, and a **DataFrame** is a multi-dimensional table made up of a collection of Series.

For example, the following DataFrame is made of two Series, **ages** and **heights**

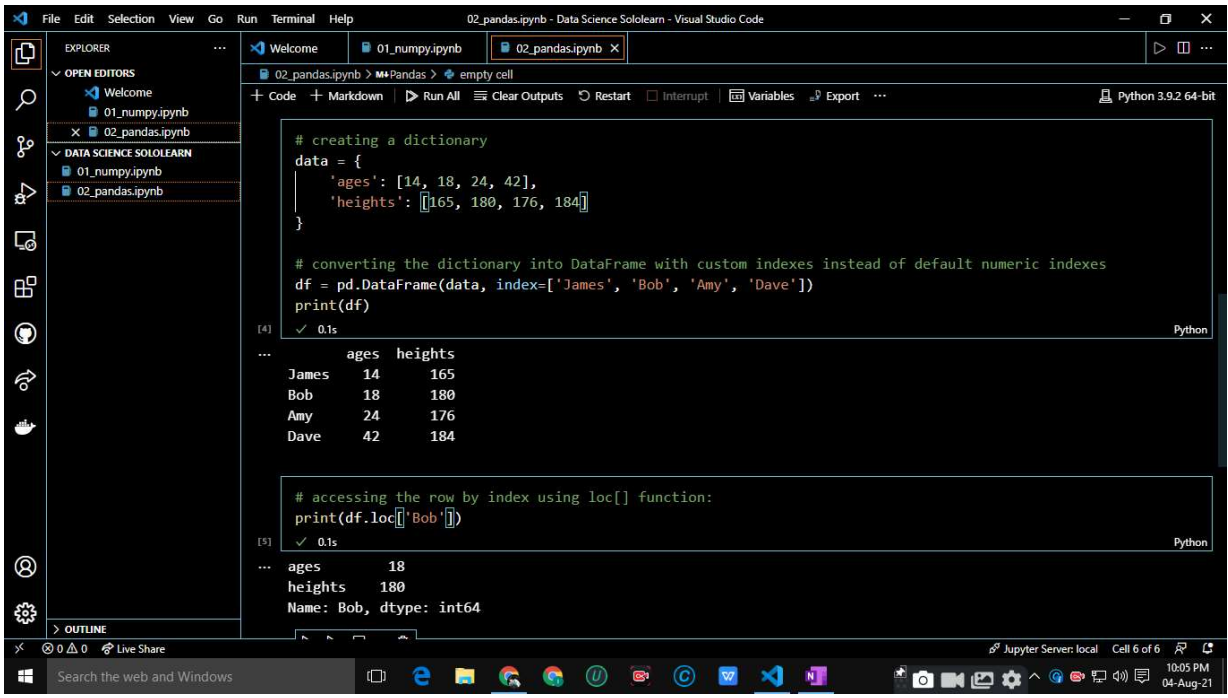
ages	heights
14	165
18	180
24	176
42	184

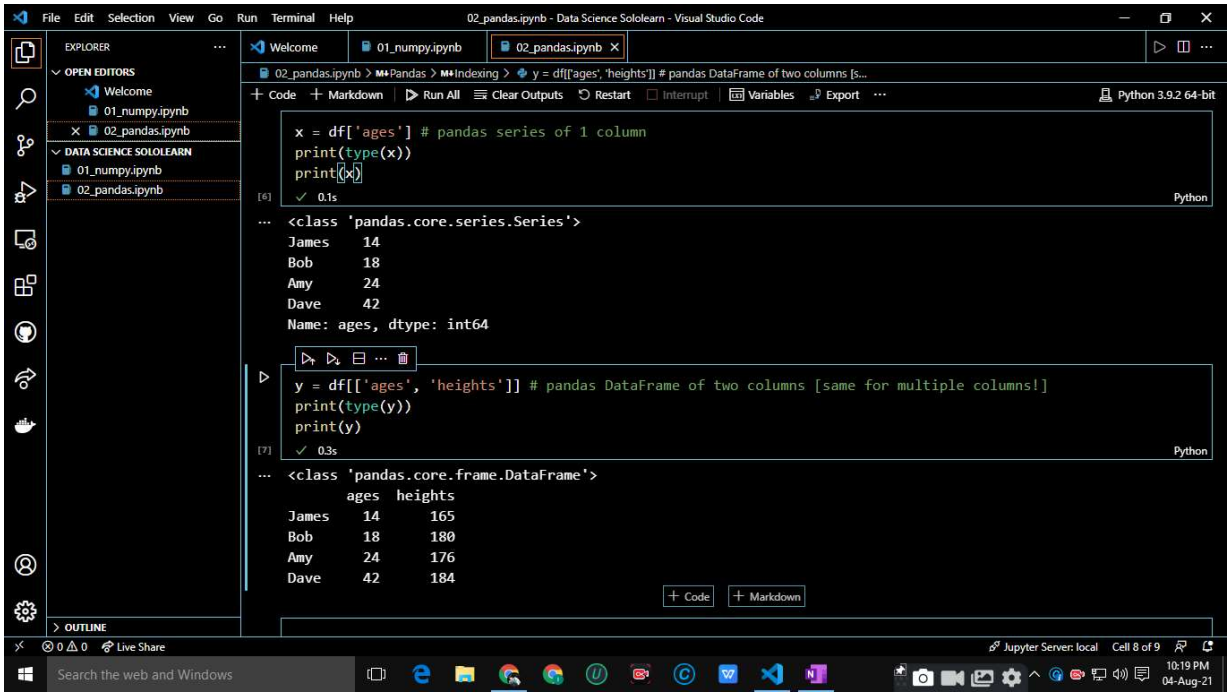
Series is a one-dimensional array

DataFrame is a multi-dimensional array.

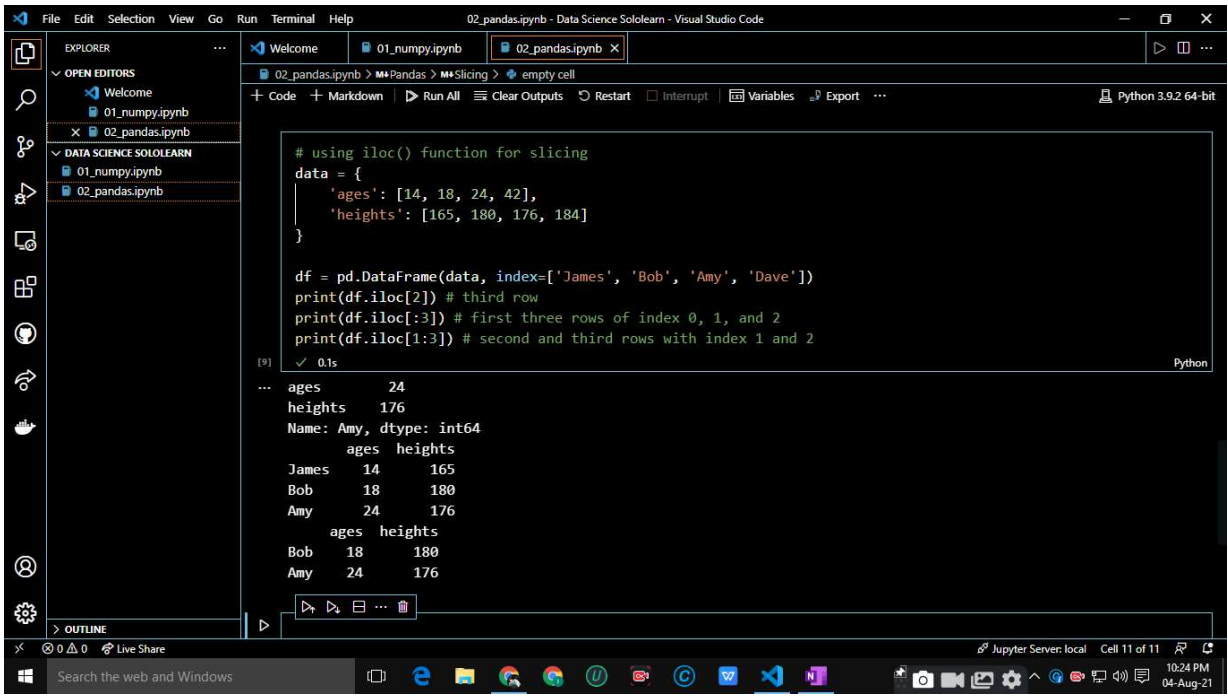


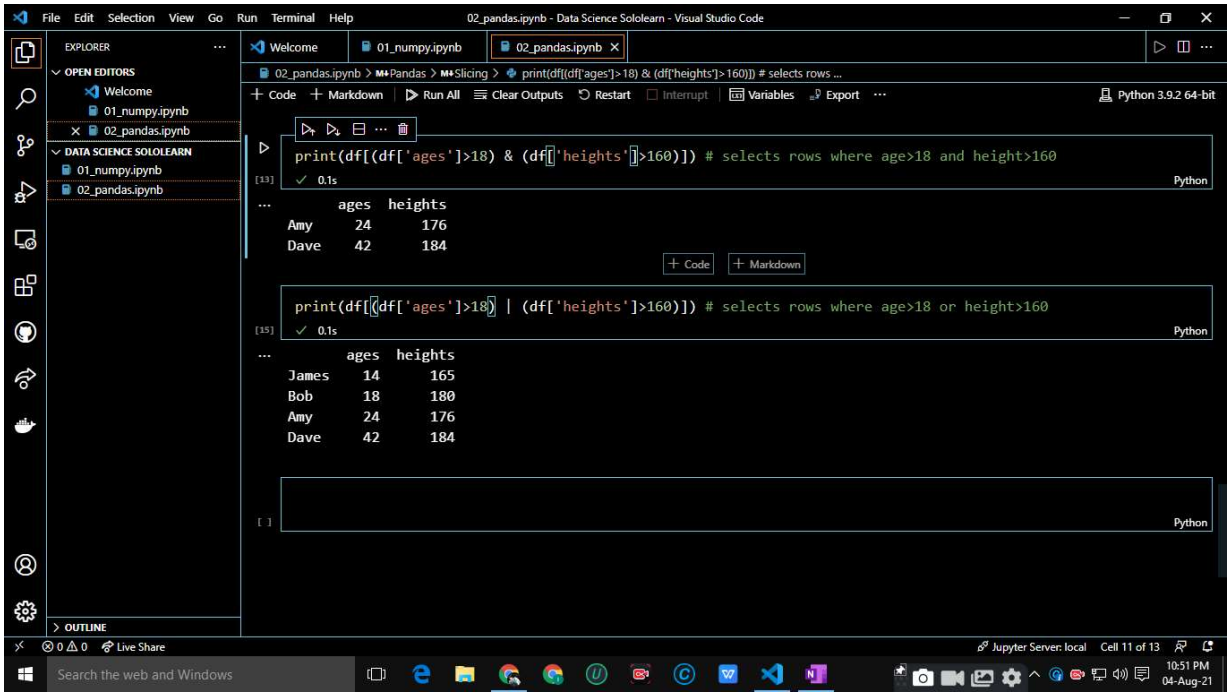
The DataFrame automatically creates a numeric **index** for each row.
We can specify a custom index, when creating the DataFrame and we can access a row using its index and the **loc[]** function:

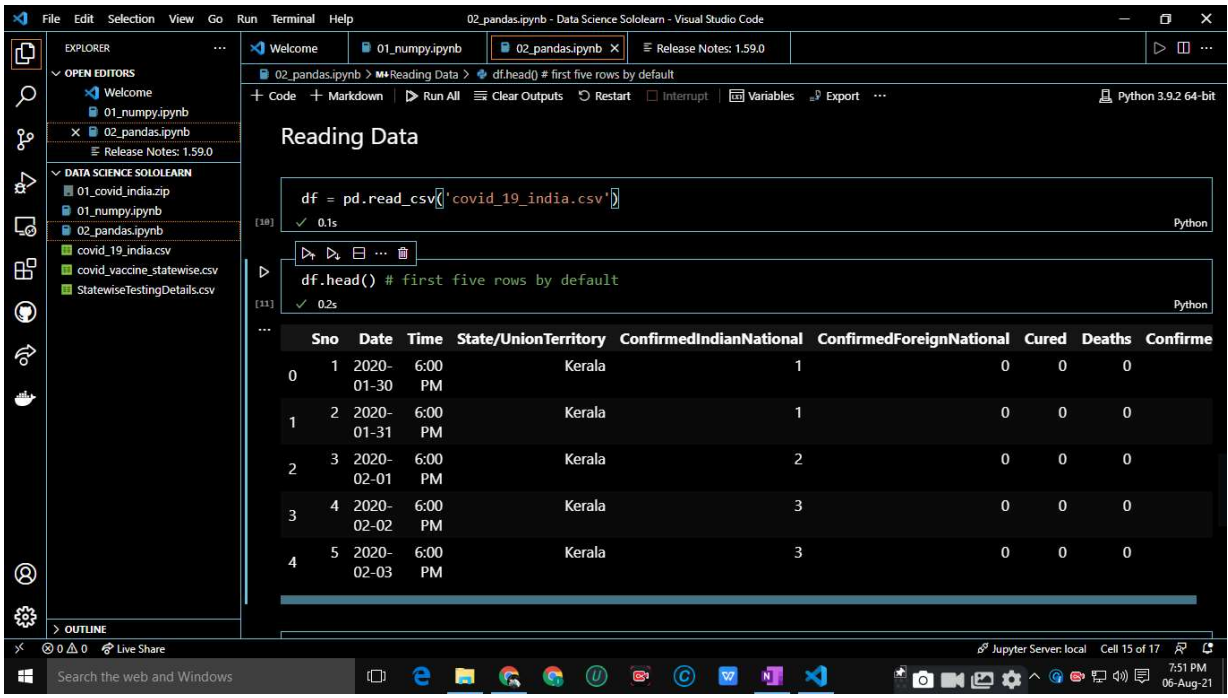


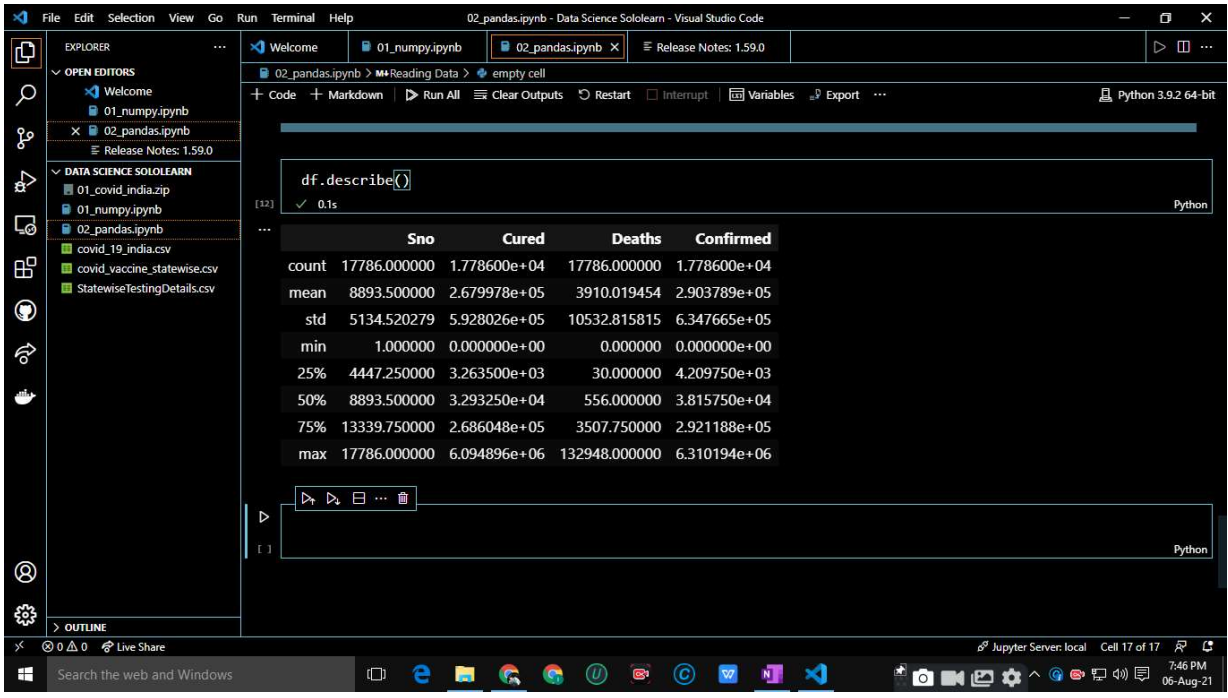


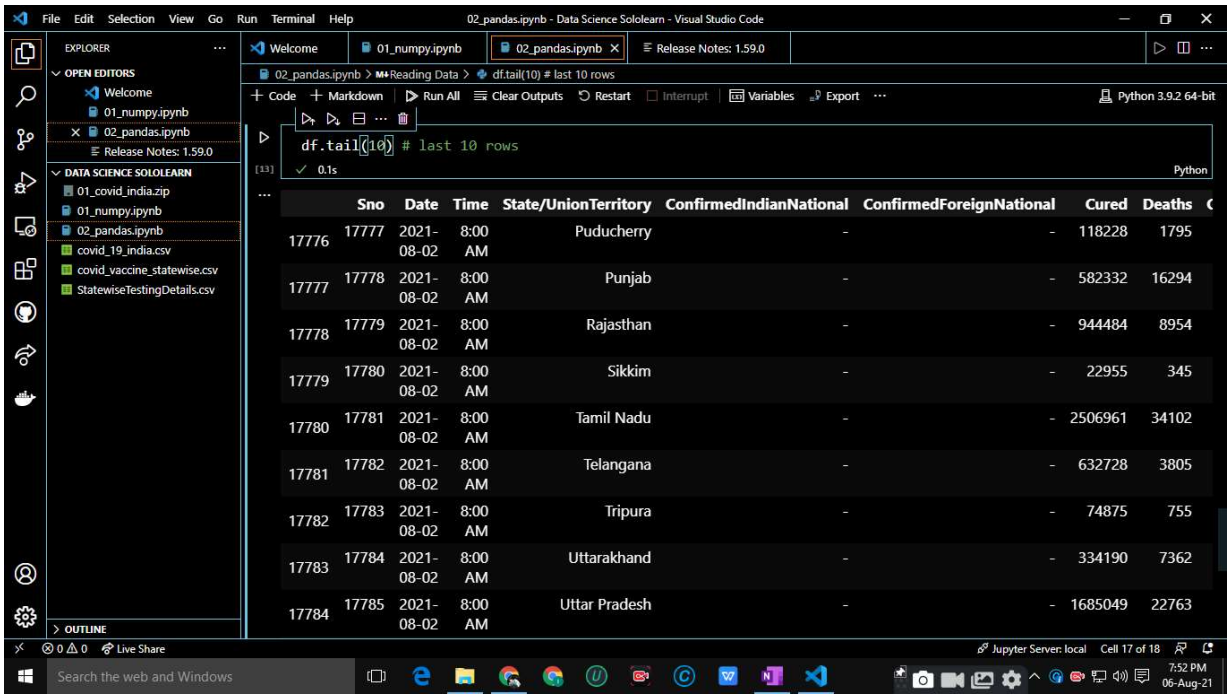
Pandas uses the **iloc** function to select data based on its numeric index.
It works the same way indexing lists does in Python.

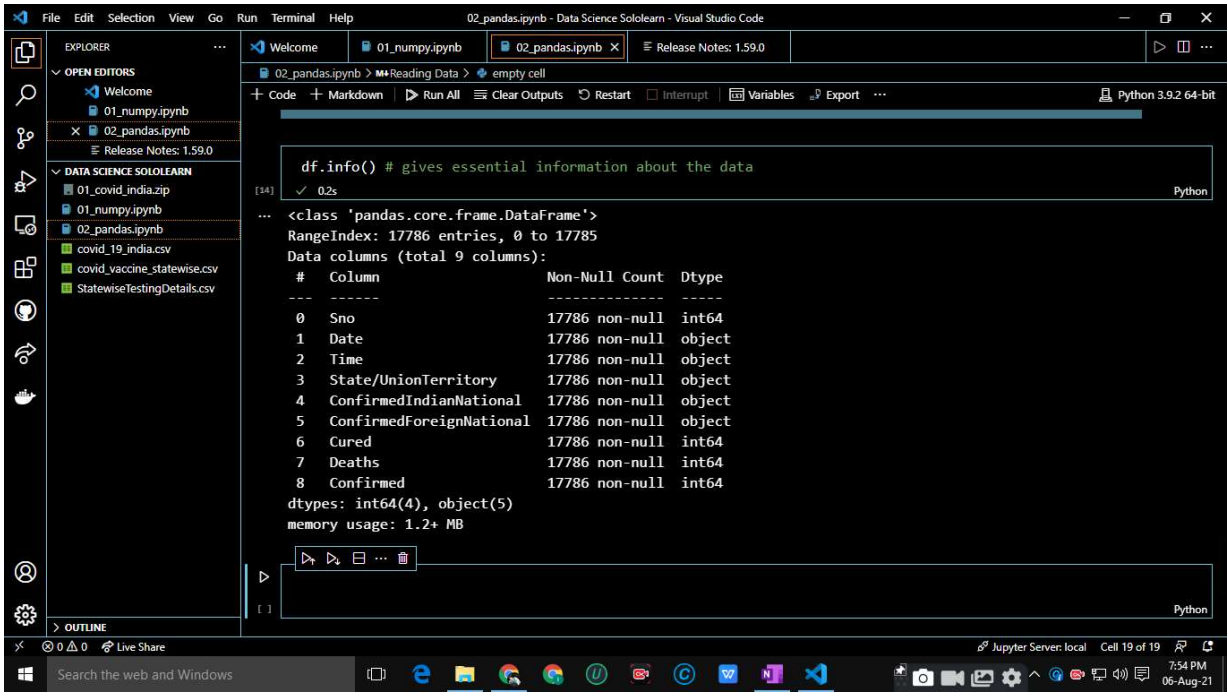

File Edit Selection View Go Run Terminal Help
02_pandas.ipynb - Data Science Sololearn - Visual Studio Code
EXPLORER
OPEN EDITORS
Welcome
01_numpy.ipynb
02_pandas.ipynb
DATA SCIENCE SOLOLEARN
01_numpy.ipynb
02_pandas.ipynb
OUTLINE
02_pandas.ipynb > Pandas > Slicing > empty cell
+ Code + Markdown Run All Clear Outputs Restart Interrupt Variables Export
Python 3.9.2 64-bit
# using iloc() function for slicing
data = {
'ages': [14, 18, 24, 42],
'heights': [165, 180, 176, 184]
}
df = pd.DataFrame(data, index=['James', 'Bob', 'Amy', 'Dave'])
print(df.iloc[2]) # third row
print(df.iloc[:3]) # first three rows of index 0, 1, and 2
print(df.iloc[1:3]) # second and third rows with index 1 and 2
[9] 0.1s
Python
ages 24
heights 176
Name: Amy, dtype: int64
ages heights
James 14 165
Bob 18 180
Amy 24 176
ages heights
Bob 18 180
Amy 24 176
Live Share
Jupyter Server: local
Cell 11 of 11
Search the web and Windows
10:24 PM
04-Aug-21


File Edit Selection View Go Run Terminal Help
02_pandas.ipynb - Data Science Sololearn - Visual Studio Code
EXPLORER
OPEN EDITORS
Welcome
01_numpy.ipynb
02_pandas.ipynb
DATA SCIENCE SOLOLEARN
01_numpy.ipynb
02_pandas.ipynb
OUTLINE
02_pandas.ipynb > Pandas > Slicing > print(df[(df['ages']>18) & (df['heights']>160)]) # selects rows ...
+ Code + Markdown Run All Clear Outputs Restart Interrupt Variables Export
Python 3.9.2 64-bit
print(df[(df['ages']>18) & (df['heights']>160)]) # selects rows where age>18 and height>160
[13] 0.1s
Python
ages heights
Amy 24 176
Dave 42 184
+ Code + Markdown
print(df[(df['ages']>18) | (df['heights']>160)]) # selects rows where age>18 or height>160
[15] 0.1s
Python
ages heights
James 14 165
Bob 18 180
Amy 24 176
Dave 42 184
[ ]
Python
Live Share
Jupyter Server: local
Cell 11 of 13
Search the web and Windows
10:51 PM
04-Aug-21

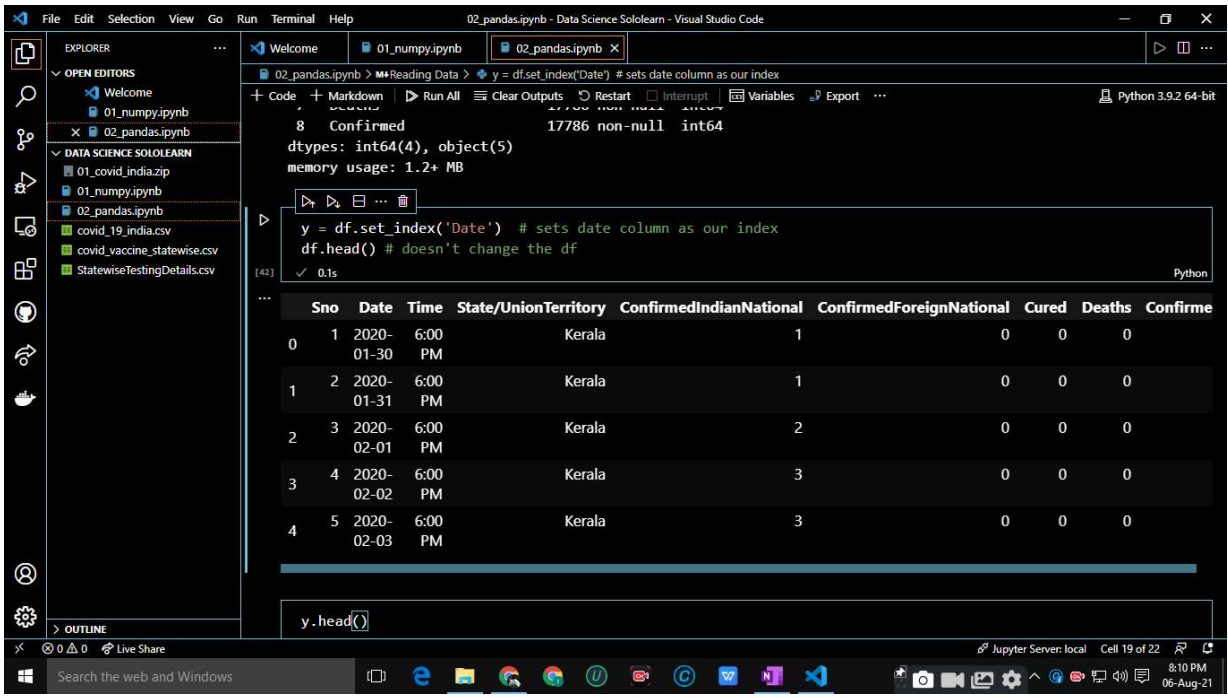

File Edit Selection View Go Run Terminal Help
02_pandas.ipynb - Data Science Sololearn - Visual Studio Code
EXPLORER
OPEN EDITORS
Welcome
01_numpy.ipynb
02_pandas.ipynb
Release Notes: 1.59.0
DATA SCIENCE SOLOLEARN
01_covid_india.zip
01_numpy.ipynb
02_pandas.ipynb
covid_19_india.csv
covid_vaccine_statewise.csv
StatewiseTestingDetails.csv
OUTLINE
02_pandas.ipynb > Reading Data > df.head() # first five rows by default
+ Code + Markdown Run All Clear Outputs Restart Interrupt Variables Export
Python 3.9.2 64-bit
Reading Data
df = pd.read_csv('covid_19_india.csv')
[10] 0.1s Python
df.head() # first five rows by default
[11] 0.2s Python
Sno Date Time State/UnionTerritory ConfirmedIndianNational ConfirmedForeignNational Cured Deaths Confirme
0 1 2020-01-30 6:00 PM Kerala 1 0 0 0
1 2 2020-01-31 6:00 PM Kerala 1 0 0 0
2 3 2020-02-01 6:00 PM Kerala 2 0 0 0
3 4 2020-02-02 6:00 PM Kerala 3 0 0 0
4 5 2020-02-03 6:00 PM Kerala 3 0 0 0
Jupyter Server: local Cell 15 of 17
Search the web and Windows
7:51 PM 06-Aug-21

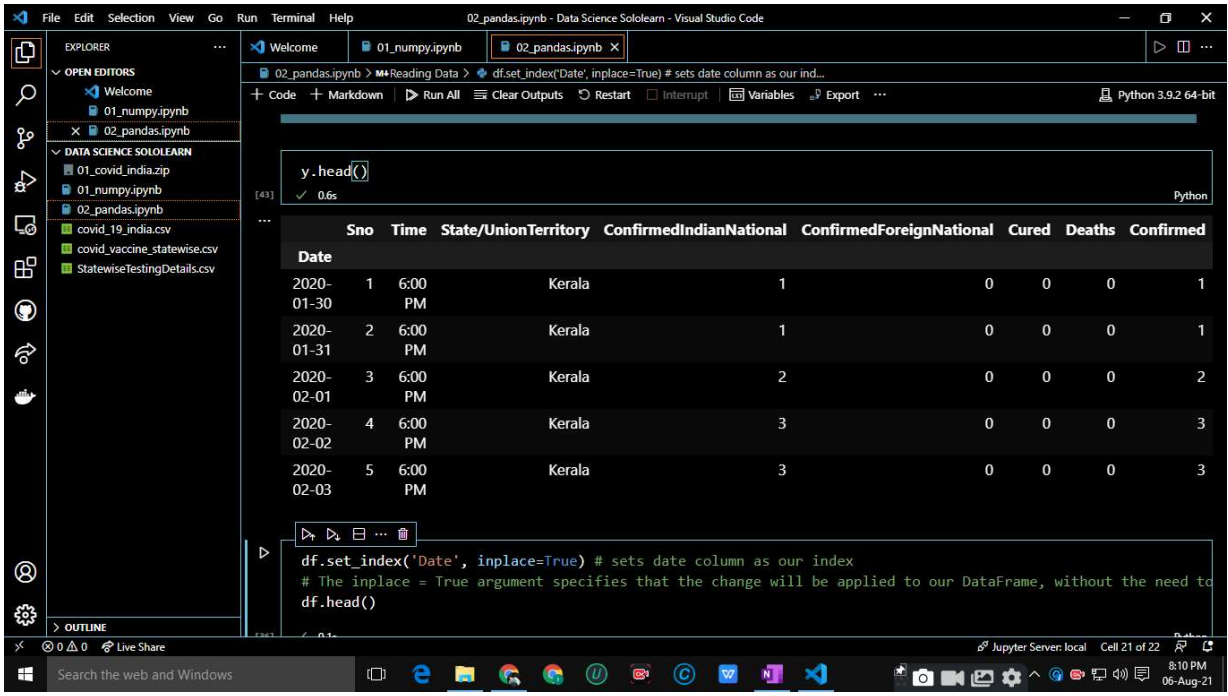

File Edit Selection View Go Run Terminal Help
02_pandas.ipynb - Data Science Sololearn - Visual Studio Code
EXPLORER
OPEN EDITORS
Welcome
01_numpy.ipynb
02_pandas.ipynb
Release Notes: 1.59.0
DATA SCIENCE SOLOLEARN
01_covid_india.zip
01_numpy.ipynb
02_pandas.ipynb
covid_19_india.csv
covid_vaccine_statewise.csv
StatewiseTestingDetails.csv
OUTLINE
02_pandas.ipynb > Reading Data > empty cell
+ Code + Markdown Run All Clear Outputs Restart Interrupt Variables Export
Python 3.9.2 64-bit
df.describe()
[12] 0.1s Python
Sno Cured Deaths Confirmed
count 17786.000000 1.778600e+04 17786.000000 1.778600e+04
mean 8893.500000 2.679978e+05 3910.019454 2.903789e+05
std 5134.520279 5.928026e+05 10532.815815 6.347665e+05
min 1.000000 0.000000e+00 0.000000 0.000000e+00
25% 4447.250000 3.263500e+03 30.000000 4.209750e+03
50% 8893.500000 3.293250e+04 556.000000 3.815750e+04
75% 13339.750000 2.686048e+05 3507.750000 2.921188e+05
max 17786.000000 6.094896e+06 132948.000000 6.310194e+06
[ ] Python
Jupyter Server: local Cell 17 of 17
Search the web and Windows
7:46 PM 06-Aug-21

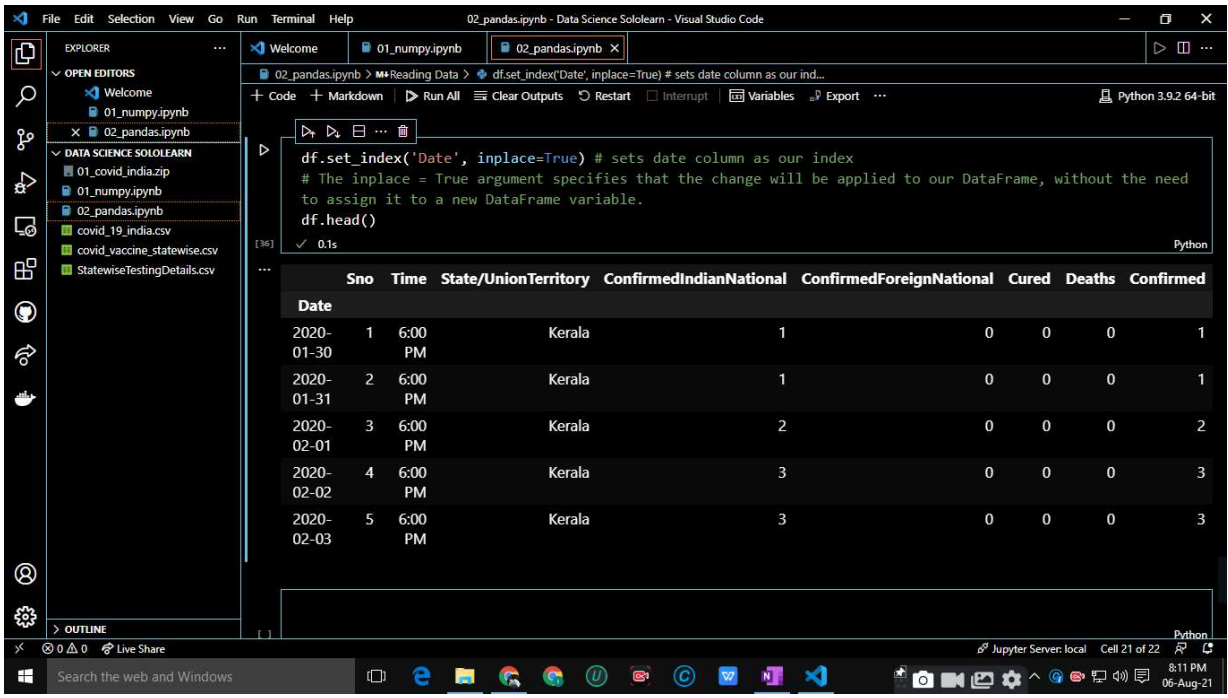




We can set our own index column by using the **set_index**() function:

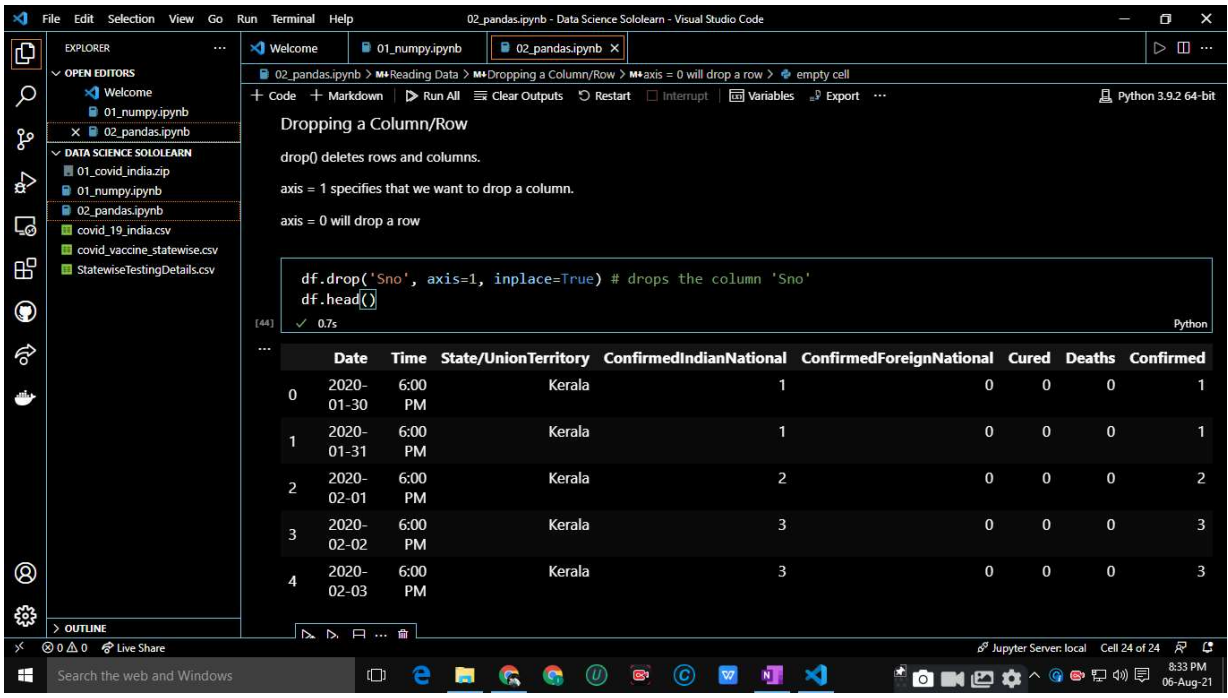

File Edit Selection View Go Run Terminal Help
02_pandas.ipynb - Data Science Sololearn - Visual Studio Code
EXPLORER
OPEN EDITORS
Welcome
01_numpy.ipynb
02_pandas.ipynb
DATA SCIENCE SOLOLEARN
01_covid_india.zip
01_numpy.ipynb
02_pandas.ipynb
covid_19_india.csv
covid_vaccine_statewise.csv
StatewiseTestingDetails.csv
OUTLINE
02_pandas.ipynb > Reading Data > y = df.set_index('Date') # sets date column as our index
+ Code + Markdown Run All Clear Outputs Restart Interrupt Variables Export
Python 3.9.2 64-bit
8 Confirmed 17786 non-null int64
dtypes: int64(4), object(5)
memory usage: 1.2+ MB
y = df.set_index('Date') # sets date column as our index
df.head() # doesn't change the df
[42] 0.1s
Python
Sno Date Time State/UnionTerritory ConfirmedIndianNational ConfirmedForeignNational Cured Deaths Confirme
0 1 2020-01-30 6:00 PM Kerala 1 0 0 0
1 2 2020-01-31 6:00 PM Kerala 1 0 0 0
2 3 2020-02-01 6:00 PM Kerala 2 0 0 0
3 4 2020-02-02 6:00 PM Kerala 3 0 0 0
4 5 2020-02-03 6:00 PM Kerala 3 0 0 0
y.head()
Live Share
Jupyter Server: local
Cell 19 of 22
Search the web and Windows
8:10 PM
06-Aug-21


File Edit Selection View Go Run Terminal Help
02_pandas.ipynb - Data Science Sololearn - Visual Studio Code
EXPLORER
OPEN EDITORS
Welcome
01_numpy.ipynb
02_pandas.ipynb
DATA SCIENCE SOLOLEARN
01_covid_india.zip
01_numpy.ipynb
02_pandas.ipynb
covid_19_india.csv
covid_vaccine_statewise.csv
StatewiseTestingDetails.csv
OUTLINE
02_pandas.ipynb > Reading Data > df.set_index('Date', inplace=True) # sets date column as our ind...
+ Code + Markdown Run All Clear Outputs Restart Interrupt Variables Export
Python 3.9.2 64-bit
y.head()
[43] 0.6s
Python
Date Sno Time State/UnionTerritory ConfirmedIndianNational ConfirmedForeignNational Cured Deaths Confirmed
2020-01-30 1 6:00 PM Kerala 1 0 0 0 1
2020-01-31 2 6:00 PM Kerala 1 0 0 0 1
2020-02-01 3 6:00 PM Kerala 2 0 0 0 2
2020-02-02 4 6:00 PM Kerala 3 0 0 0 3
2020-02-03 5 6:00 PM Kerala 3 0 0 0 3
df.set_index('Date', inplace=True) # sets date column as our index
# The inplace = True argument specifies that the change will be applied to our DataFrame, without the need to
df.head()
Live Share
Jupyter Server: local
Cell 21 of 22
Search the web and Windows
8:10 PM
06-Aug-21



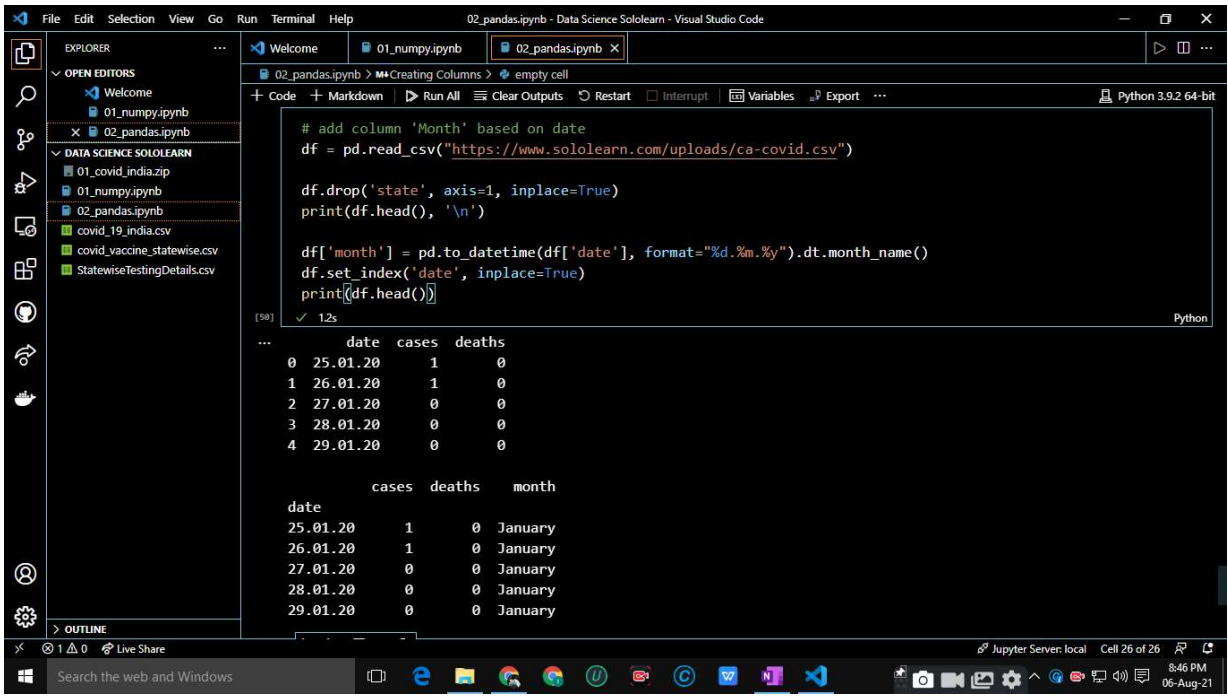
Date	Sno	Time	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Deaths	Confirmed
2020-01-30	1	6:00 PM	Kerala	1	0	0	0	1
2020-01-31	2	6:00 PM	Kerala	1	0	0	0	1
2020-02-01	3	6:00 PM	Kerala	2	0	0	0	2
2020-02-02	4	6:00 PM	Kerala	3	0	0	0	3
2020-02-03	5	6:00 PM	Kerala	3	0	0	0	3

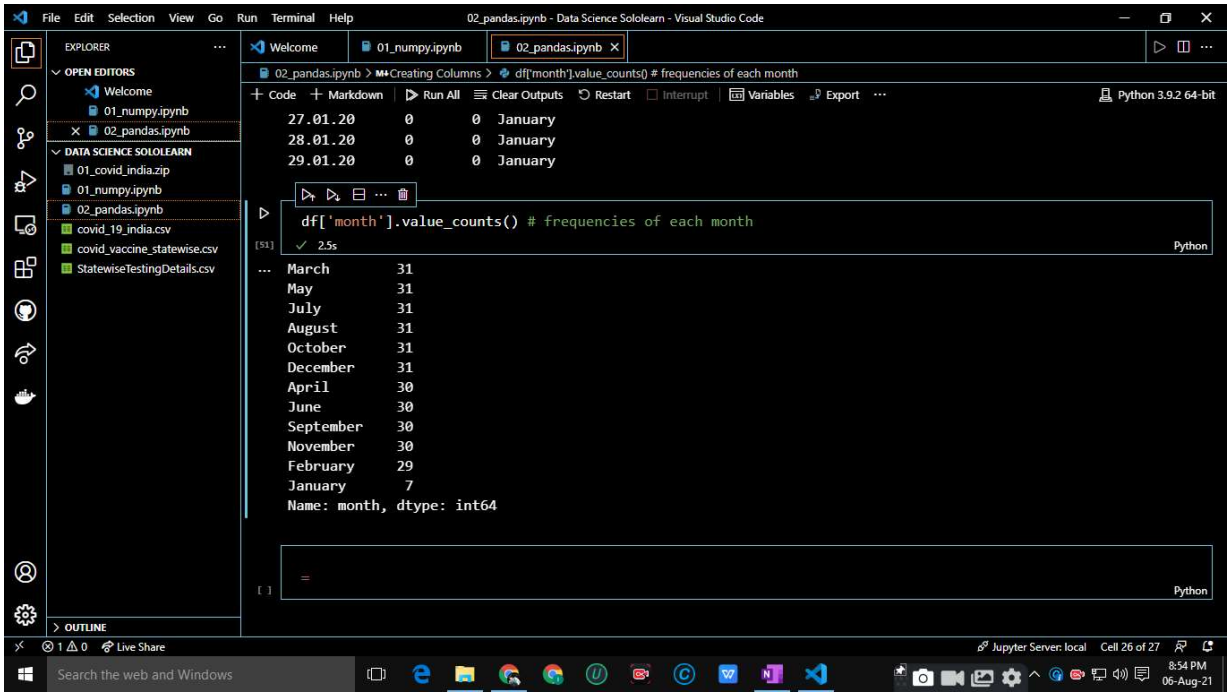
```
## Dropping a Column/Row
## drop() deletes rows and columns.
## axis = 1 specifies that we want to drop a column.
## axis = 0 will drop a row
```

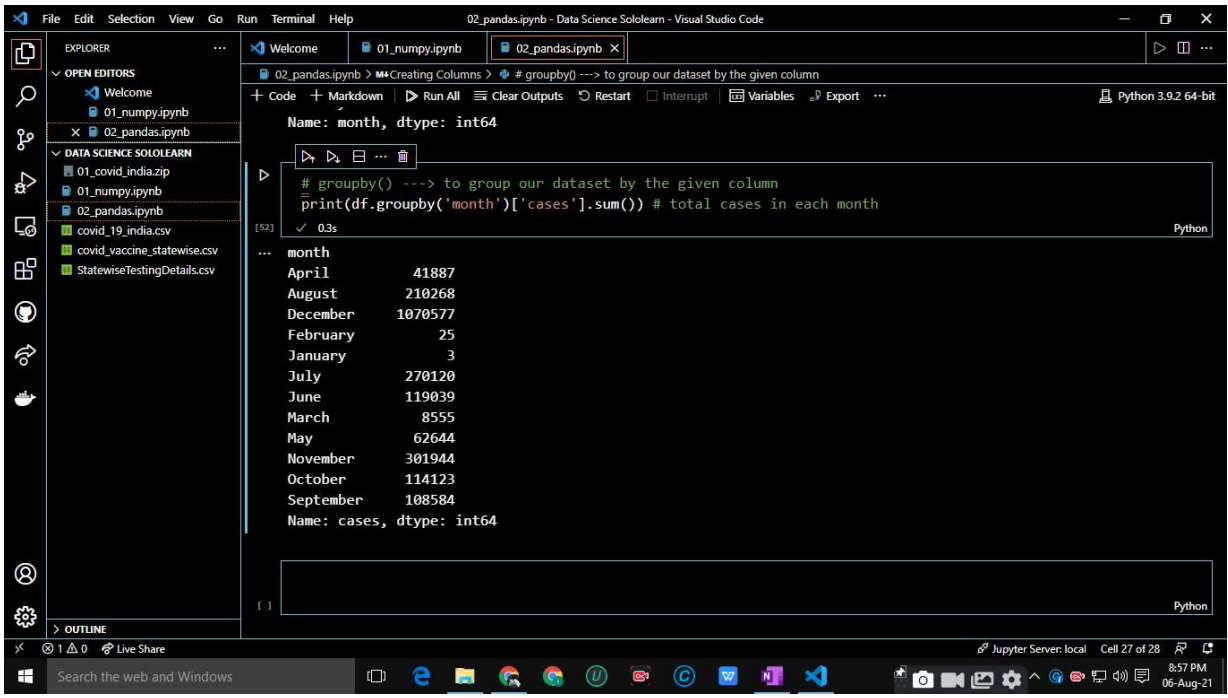


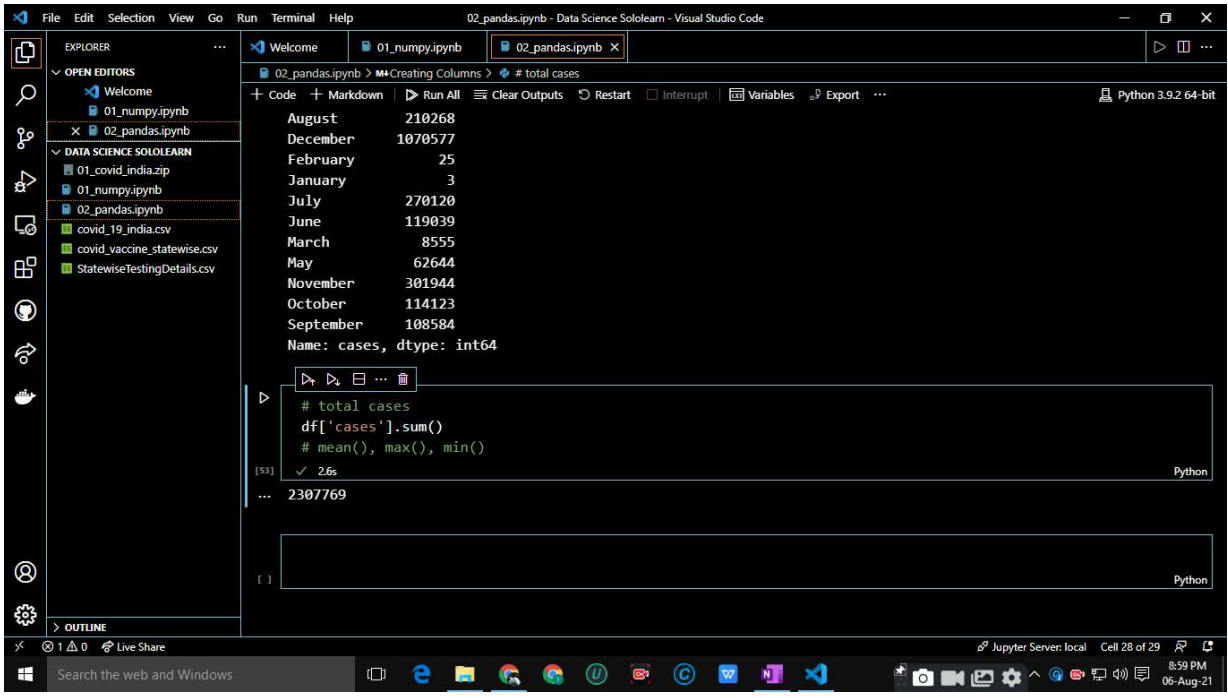
	Date	Time	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Deaths	Confirmed
0	2020-01-30	6:00 PM	Kerala	1	0	0	0	1
1	2020-01-31	6:00 PM	Kerala	1	0	0	0	1
2	2020-02-01	6:00 PM	Kerala	2	0	0	0	2
3	2020-02-02	6:00 PM	Kerala	3	0	0	0	3
4	2020-02-03	6:00 PM	Kerala	3	0	0	0	3

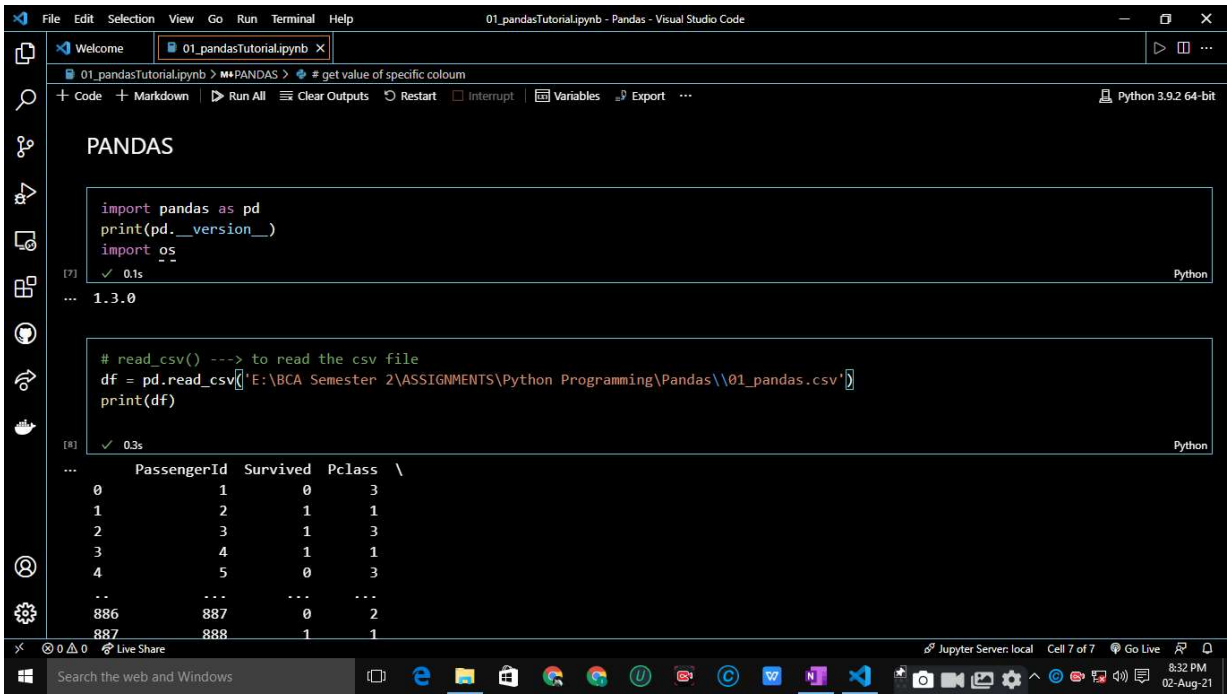
Pandas allows us to create our own columns.
For example, we can add a month column based on the date column:

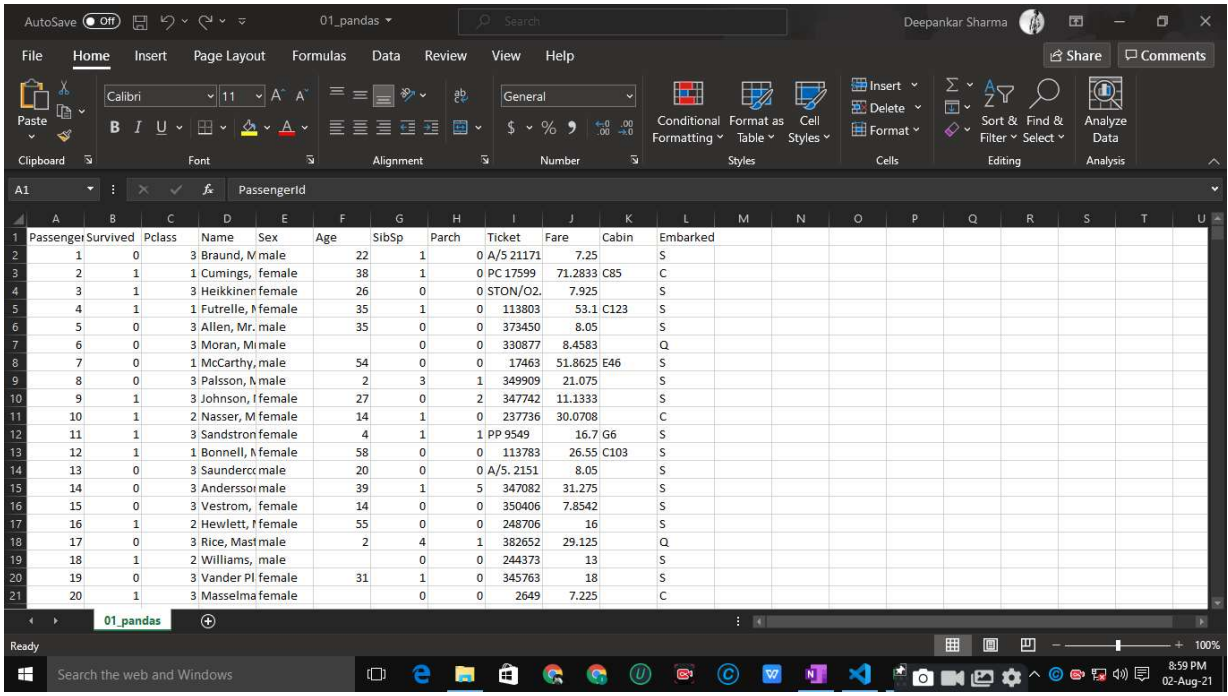

02_pandas.ipynb - Data Science Sololearn - Visual Studio Code
# add column 'Month' based on date
df = pd.read_csv("https://www.sololearn.com/uploads/ca-covid.csv")
df.drop('state', axis=1, inplace=True)
print(df.head(), '\n')
df['month'] = pd.to_datetime(df['date'], format="%d.%m.%y").dt.month_name()
df.set_index('date', inplace=True)
print(df.head())
date cases deaths
0 25.01.20 1 0
1 26.01.20 1 0
2 27.01.20 0 0
3 28.01.20 0 0
4 29.01.20 0 0
cases deaths month
date
25.01.20 1 0 January
26.01.20 1 0 January
27.01.20 0 0 January
28.01.20 0 0 January
29.01.20 0 0 January


02_pandas.ipynb - Data Science Sololearn - Visual Studio Code
27.01.20 0 0 January
28.01.20 0 0 January
29.01.20 0 0 January
df['month'].value_counts() # frequencies of each month
March 31
May 31
July 31
August 31
October 31
December 31
April 30
June 30
September 30
November 30
February 29
January 7
Name: month, dtype: int64

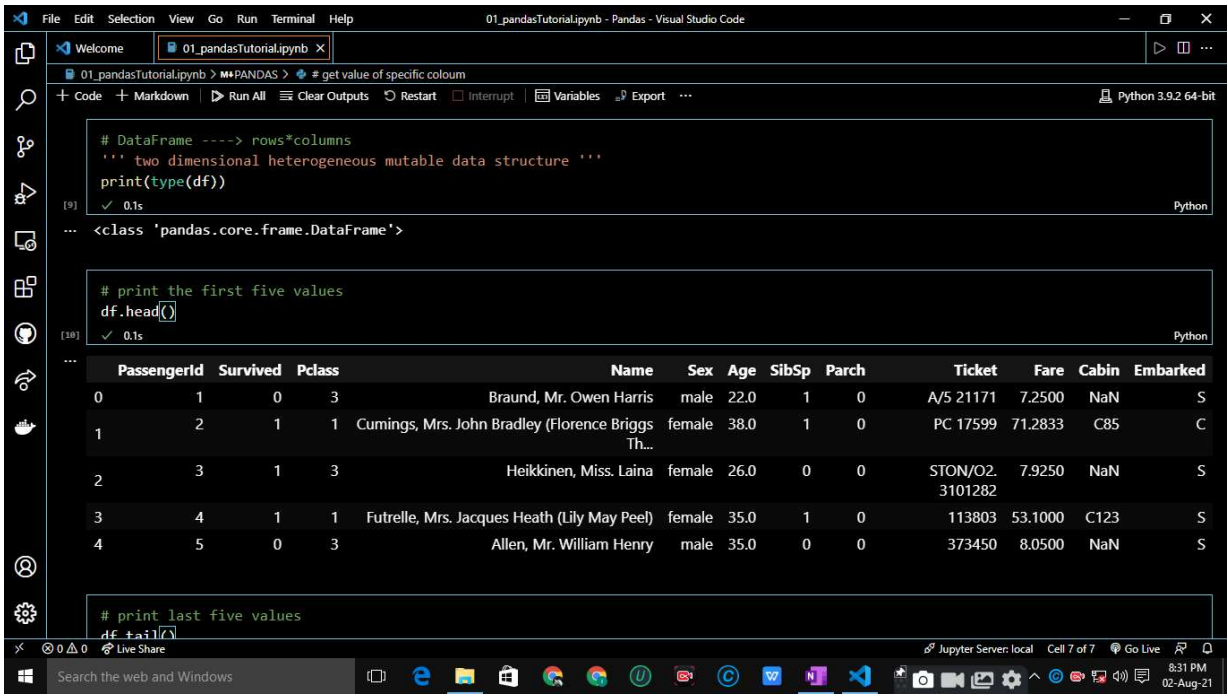

File Edit Selection View Go Run Terminal Help
02_pandas.ipynb - Data Science Sololearn - Visual Studio Code
EXPLORER
OPEN EDITORS
Welcome
01_numpy.ipynb
02_pandas.ipynb
DATA SCIENCE SOLOLEARN
01_covid_india.zip
01_numpy.ipynb
02_pandas.ipynb
covid_19_india.csv
covid_vaccine_statewise.csv
StatewiseTestingDetails.csv
OUTLINE
02_pandas.ipynb > Creating Columns > # groupby() ---> to group our dataset by the given column
+ Code + Markdown Run All Clear Outputs Restart Interrupt Variables Export
Python 3.9.2 64-bit
Name: month, dtype: int64
groupby() ---> to group our dataset by the given column
print(df.groupby('month')['cases'].sum()) # total cases in each month
[52] 0.3s
Python
month
April 41887
August 210268
December 1070577
February 25
January 3
July 270120
June 119039
March 8555
May 62644
November 301944
October 114123
September 108584
Name: cases, dtype: int64
[ ]
Python
Live Share
Jupyter Server: local
Cell 27 of 28
Search the web and Windows
8:57 PM
06-Aug-21

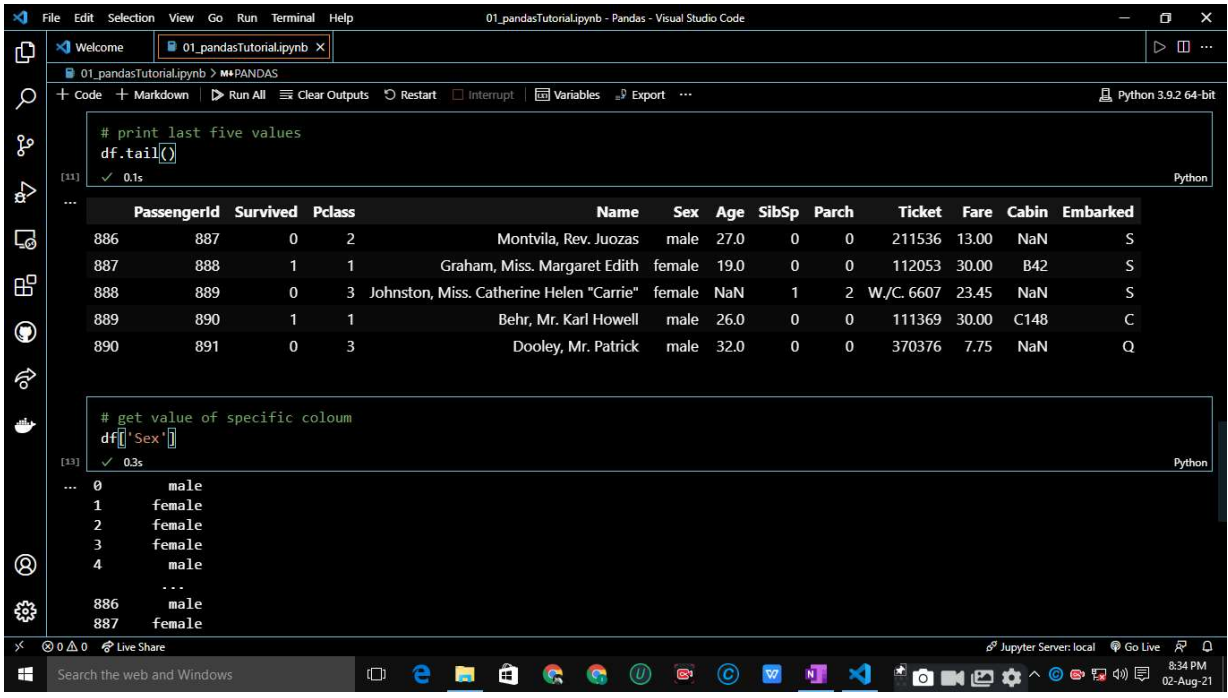

File Edit Selection View Go Run Terminal Help
02_pandas.ipynb - Data Science Sololearn - Visual Studio Code
EXPLORER
OPEN EDITORS
Welcome
01_numpy.ipynb
02_pandas.ipynb
DATA SCIENCE SOLOLEARN
01_covid_india.zip
01_numpy.ipynb
02_pandas.ipynb
covid_19_india.csv
covid_vaccine_statewise.csv
StatewiseTestingDetails.csv
OUTLINE
02_pandas.ipynb > Creating Columns > # total cases
+ Code + Markdown Run All Clear Outputs Restart Interrupt Variables Export
Python 3.9.2 64-bit
August 210268
December 1070577
February 25
January 3
July 270120
June 119039
March 8555
May 62644
November 301944
October 114123
September 108584
Name: cases, dtype: int64
total cases
df['cases'].sum()
mean(), max(), min()
[53] 2.6s
Python
2307769
[ ]
Python
Live Share
Jupyter Server: local
Cell 28 of 29
Search the web and Windows
8:59 PM
06-Aug-21

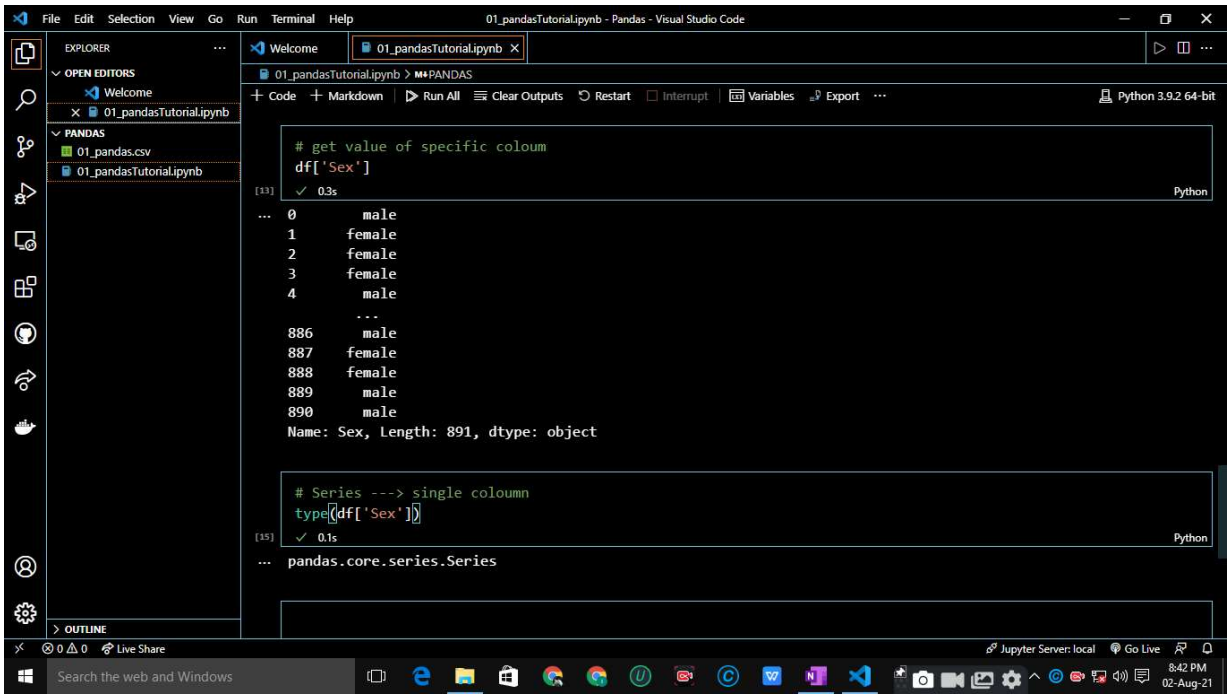


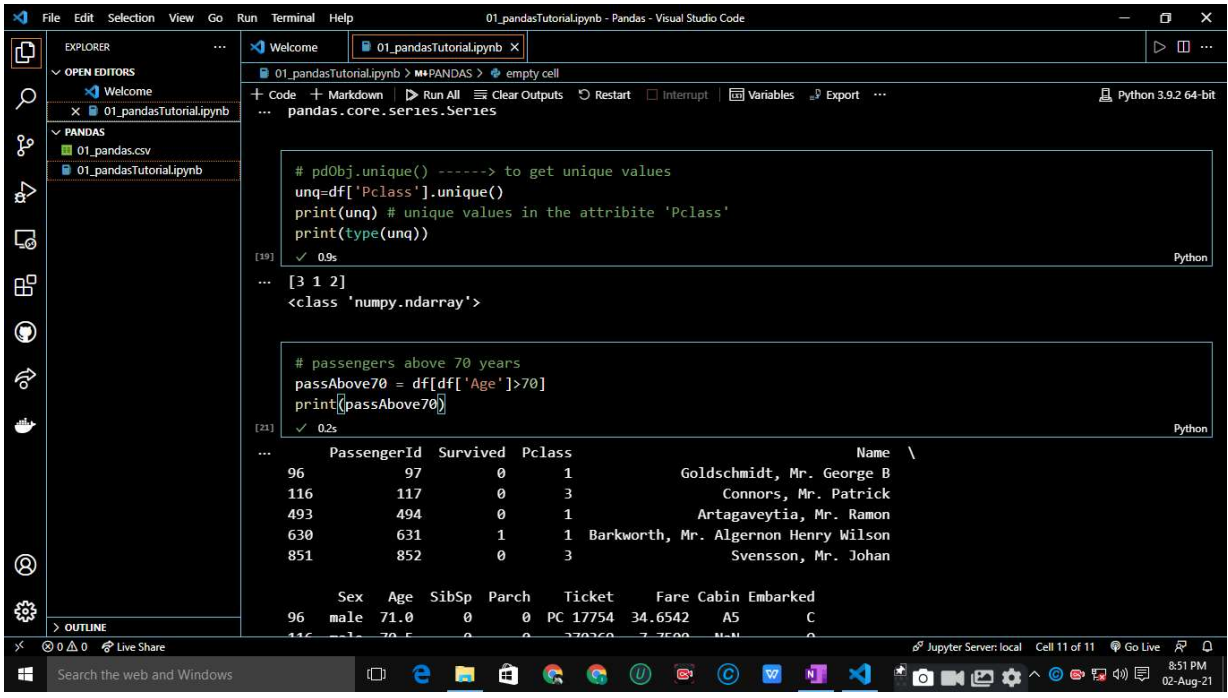


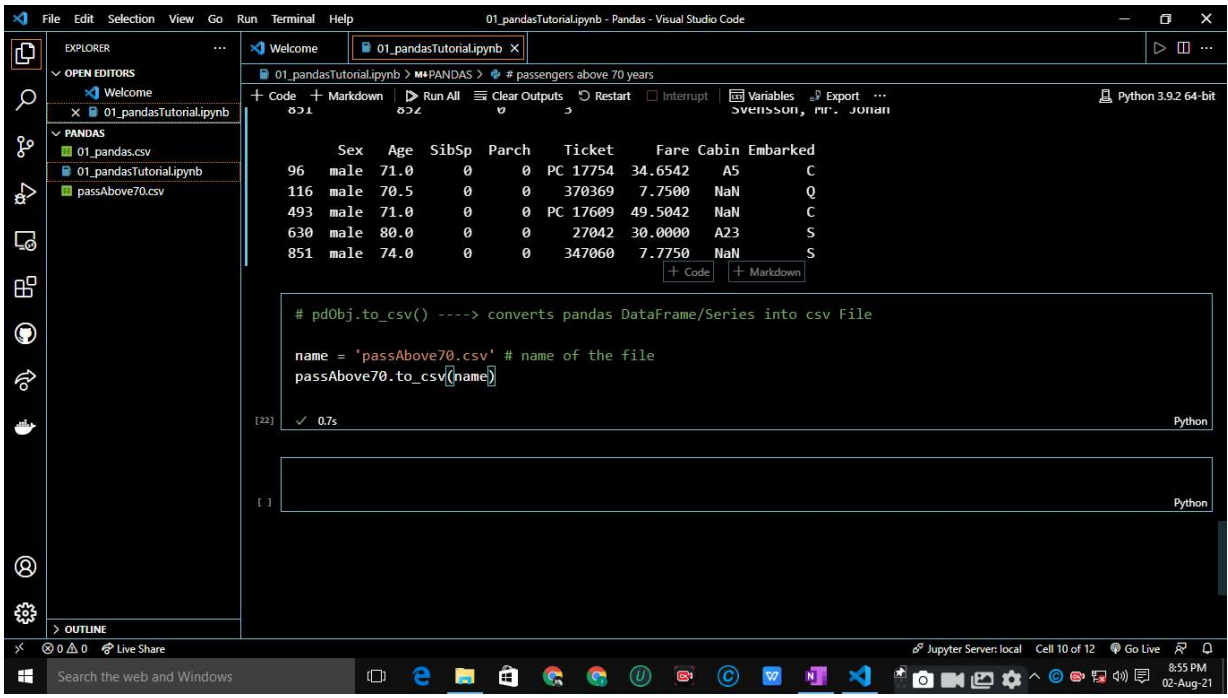
	A	B	C	D	E	F	G	H	I	J	K	L
1	Passenger	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
2	1	0	3	Braund, M	male	22	1	0	A/5 21171	7.25		S
3	2	1	1	Cumings,	female	38	1	0	PC 17599	71.2833	C85	C
4	3	1	3	Heikkinen	female	26	0	0	STON/O2.	7.925		S
5	4	1	1	Futrelle, M	female	35	1	0	113803	53.1	C123	S
6	5	0	3	Allen, Mr.	male	35	0	0	373450	8.05		S
7	6	0	3	Moran, Mi	male		0	0	330877	8.4583		Q
8	7	0	1	McCarthy,	male	54	0	0	17463	51.8625	E46	S
9	8	0	3	Palsson, M	male	2	3	1	349909	21.075		S
10	9	1	3	Johnson, I	female	27	0	2	347742	11.1333		S
11	10	1	2	Nasser, M	female	14	1	0	237736	30.0708		C
12	11	1	3	Sandstron	female	4	1	1	PP 9549	16.7	G6	S
13	12	1	1	Bonnell, M	female	58	0	0	113783	26.55	C103	S
14	13	0	3	Saunderco	male	20	0	0	A/5. 2151	8.05		S
15	14	0	3	Andersson	male	39	1	5	347082	31.275		S
16	15	0	3	Vestrom,	female	14	0	0	350406	7.8542		S
17	16	1	2	Hewlett, M	female	55	0	0	248706	16		S
18	17	0	3	Rice, Mast	male	2	4	1	382652	29.125		Q
19	18	1	2	Williams,	male		0	0	244373	13		S
20	19	0	3	Vander Pl	female	31	1	0	345763	18		S
21	20	1	3	Masselma	female		0	0	2649	7.225		C

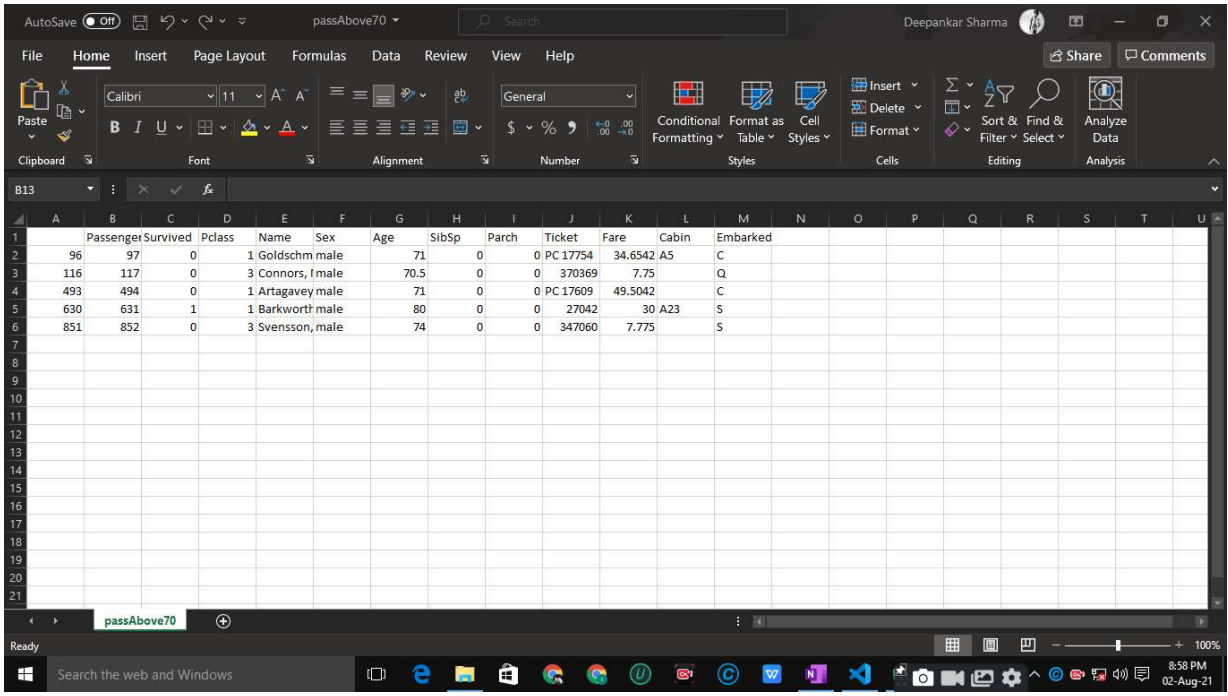

File Edit Selection View Go Run Terminal Help
01_pandasTutorial.ipynb - Pandas - Visual Studio Code
Welcome
01_pandasTutorial.ipynb
01_pandasTutorial.ipynb > MD PANDAS > # get value of specific coloum
+ Code + Markdown Run All Clear Outputs Restart Interrupt Variables Export
Python 3.9.2 64-bit
DataFrame ----> rows*columns
''' two dimensional heterogeneous mutable data structure '''
print(type(df))
[9] 0.1s
Python
<class 'pandas.core.frame.DataFrame'>
# print the first five values
df.head()
[10] 0.1s
Python
PassengerId Survived Pclass Name Sex Age SibSp Parch Ticket Fare Cabin Embarked
0 1 0 3 Braund, Mr. Owen Harris male 22.0 1 0 A/5 21171 7.2500 NaN S
1 2 1 1 Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.0 1 0 PC 17599 71.2833 C85 C
2 3 1 3 Heikkinen, Miss. Laina female 26.0 0 0 STON/O2. 3101282 7.9250 NaN S
3 4 1 1 Futrelle, Mrs. Jacques Heath (Lily May Peel) female 35.0 1 0 113803 53.1000 C123 S
4 5 0 3 Allen, Mr. William Henry male 35.0 0 0 373450 8.0500 NaN S
# print last five values
df.tail()
Live Share
Jupyter Server: local
Cell 7 of 7
Go Live
Search the web and Windows
8:31 PM
02-Aug-21

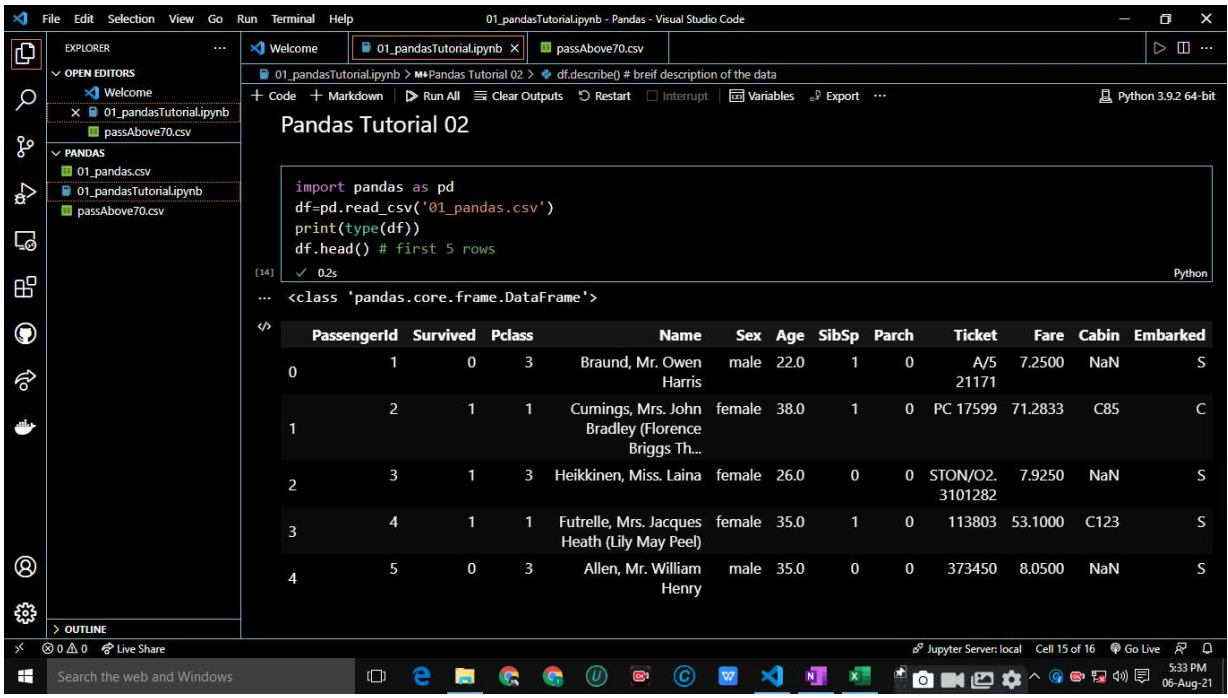

File Edit Selection View Go Run Terminal Help
01_pandasTutorial.ipynb - Pandas - Visual Studio Code
Welcome
01_pandasTutorial.ipynb
01_pandasTutorial.ipynb > MD PANDAS
+ Code + Markdown Run All Clear Outputs Restart Interrupt Variables Export
Python 3.9.2 64-bit
# print last five values
df.tail()
[11] 0.1s
Python
PassengerId Survived Pclass Name Sex Age SibSp Parch Ticket Fare Cabin Embarked
886 887 0 2 Montvila, Rev. Juozas male 27.0 0 0 211536 13.00 NaN S
887 888 1 1 Graham, Miss. Margaret Edith female 19.0 0 0 112053 30.00 B42 S
888 889 0 3 Johnston, Miss. Catherine Helen "Carrie" female NaN 1 2 W./C. 6607 23.45 NaN S
889 890 1 1 Behr, Mr. Karl Howell male 26.0 0 0 111369 30.00 C148 C
890 891 0 3 Dooley, Mr. Patrick male 32.0 0 0 370376 7.75 NaN Q
get value of specific coloum
df['Sex']
[13] 0.3s
Python
0 male
1 female
2 female
3 female
4 male
...
886 male
887 female
Live Share
Jupyter Server: local
Go Live
Search the web and Windows
8:34 PM
02-Aug-21


File Edit Selection View Go Run Terminal Help
01_pandasTutorial.ipynb - Pandas - Visual Studio Code
EXPLORER
OPEN EDITORS
Welcome
01_pandasTutorial.ipynb
PANDAS
01_pandas.csv
01_pandasTutorial.ipynb
OUTLINE
01_pandasTutorial.ipynb > PANDAS
+ Code + Markdown Run All Clear Outputs Restart Interrupt Variables Export
Python 3.9.2 64-bit
get value of specific coloum
df['Sex']
[13] ✓ 0.3s Python
0 male
1 female
2 female
3 female
4 male
...
886 male
887 female
888 female
889 male
890 male
Name: Sex, Length: 891, dtype: object
# Series ---> single coloumn
type(df['Sex'])
[15] ✓ 0.1s Python
pandas.core.series.Series
Live Share
Jupyter Server: local
Go Live
Search the web and Windows
8:42 PM
02-Aug-21

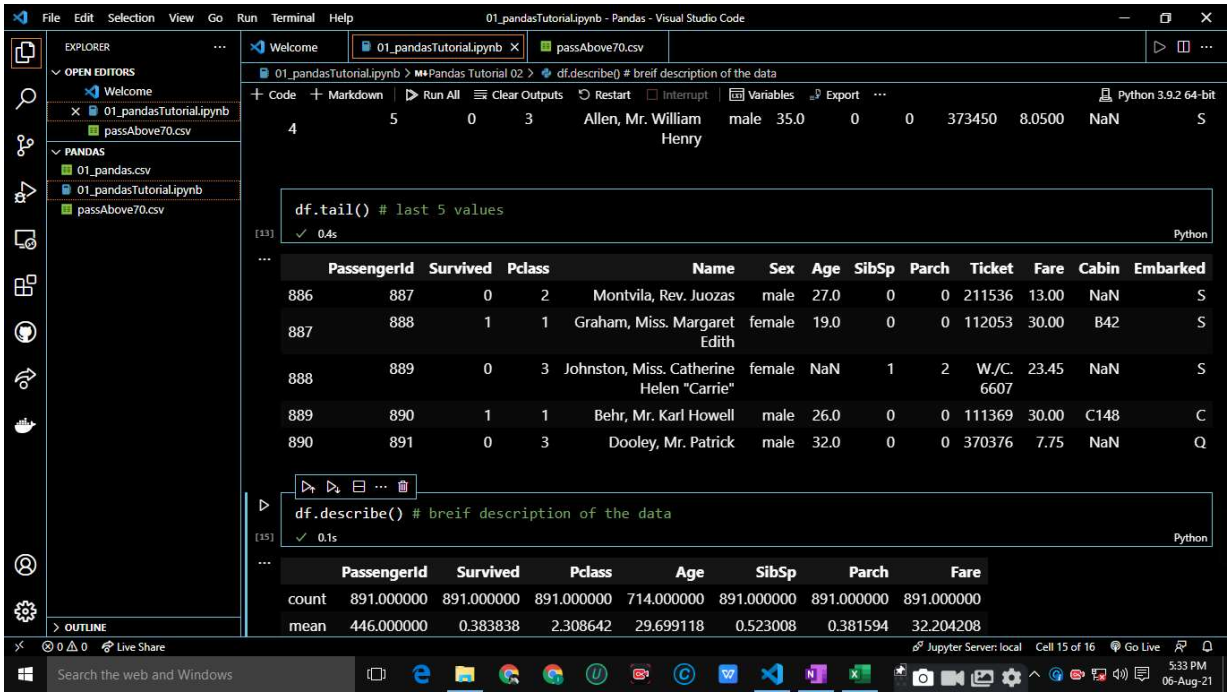

File Edit Selection View Go Run Terminal Help
01_pandasTutorial.ipynb - Pandas - Visual Studio Code
EXPLORER
OPEN EDITORS
Welcome
01_pandasTutorial.ipynb
PANDAS
01_pandas.csv
01_pandasTutorial.ipynb
OUTLINE
01_pandasTutorial.ipynb > PANDAS > empty cell
+ Code + Markdown Run All Clear Outputs Restart Interrupt Variables Export
Python 3.9.2 64-bit
pandas.core.series.Series
# pdObj.unique() ------> to get unique values
unq=df['Pclass'].unique()
print(unq) # unique values in the attribite 'Pclass'
print(type(unq))
[19] ✓ 0.9s Python
[3 1 2]
<class 'numpy.ndarray'>
# passengers above 70 years
passAbove70 = df[df['Age']>70]
print(passAbove70)
[21] ✓ 0.2s Python
PassengerId Survived Pclass Name \
96 97 0 1 Goldschmidt, Mr. George B
116 117 0 3 Connors, Mr. Patrick
493 494 0 1 Artagaveytia, Mr. Ramon
630 631 1 1 Barkworth, Mr. Algernon Henry Wilson
851 852 0 3 Svensson, Mr. Johan
Sex Age SibSp Parch Ticket Fare Cabin Embarked
96 male 71.0 0 0 PC 17754 34.6542 A5 C
Live Share
Jupyter Server: local
Cell 11 of 11
Go Live
Search the web and Windows
8:51 PM
02-Aug-21


01_pandasTutorial.ipynb - Pandas - Visual Studio Code
EXPLORER
OPEN EDITORS
Welcome
01_pandasTutorial.ipynb
PANDAS
01_pandas.csv
01_pandasTutorial.ipynb
passAbove70.csv
OUTLINE
01_pandasTutorial.ipynb > PANDAS > # passengers above 70 years
Code
Markdown
Run All
Clear Outputs
Restart
Interrupt
Variables
Export
Python 3.9.2 64-bit
Sex Age SibSp Parch Ticket Fare Cabin Embarked
96 male 71.0 0 0 PC 17754 34.6542 A5 C
116 male 70.5 0 0 370369 7.7500 NaN Q
493 male 71.0 0 0 PC 17609 49.5042 NaN C
630 male 80.0 0 0 27042 30.0000 A23 S
851 male 74.0 0 0 347060 7.7750 NaN S
# pdObj.to_csv() ----> converts pandas DataFrame/Series into csv File
name = 'passAbove70.csv' # name of the file
passAbove70.to_csv(name)
0.7s
Python
Jupyter Server: local
Cell 10 of 12
Go Live
Search the web and Windows
8:55 PM
02-Aug-21


AutoSave
passAbove70
Deepankar Sharma
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Help
Share
Comments
Paste
Calibri
11
General
Conditional Formatting
Format as Table
Cell Styles
Insert
Delete
Format
Sort & Filter
Find & Select
Analyze Data
Clipboard
Font
Alignment
Number
Styles
Cells
Editing
Analysis
B13
Passenger Survived Pclass Name Sex Age SibSp Parch Ticket Fare Cabin Embarked
96 97 0 1 Goldschm male 71 0 0 PC 17754 34.6542 A5 C
116 117 0 3 Connors, I male 70.5 0 0 370369 7.75 Q
493 494 0 1 Artagavey male 71 0 0 PC 17609 49.5042 C
630 631 1 1 Barkworth male 80 0 0 27042 30 A23 S
851 852 0 3 Svensson, male 74 0 0 347060 7.775 S
passAbove70
Ready
100%
Search the web and Windows
8:58 PM
02-Aug-21

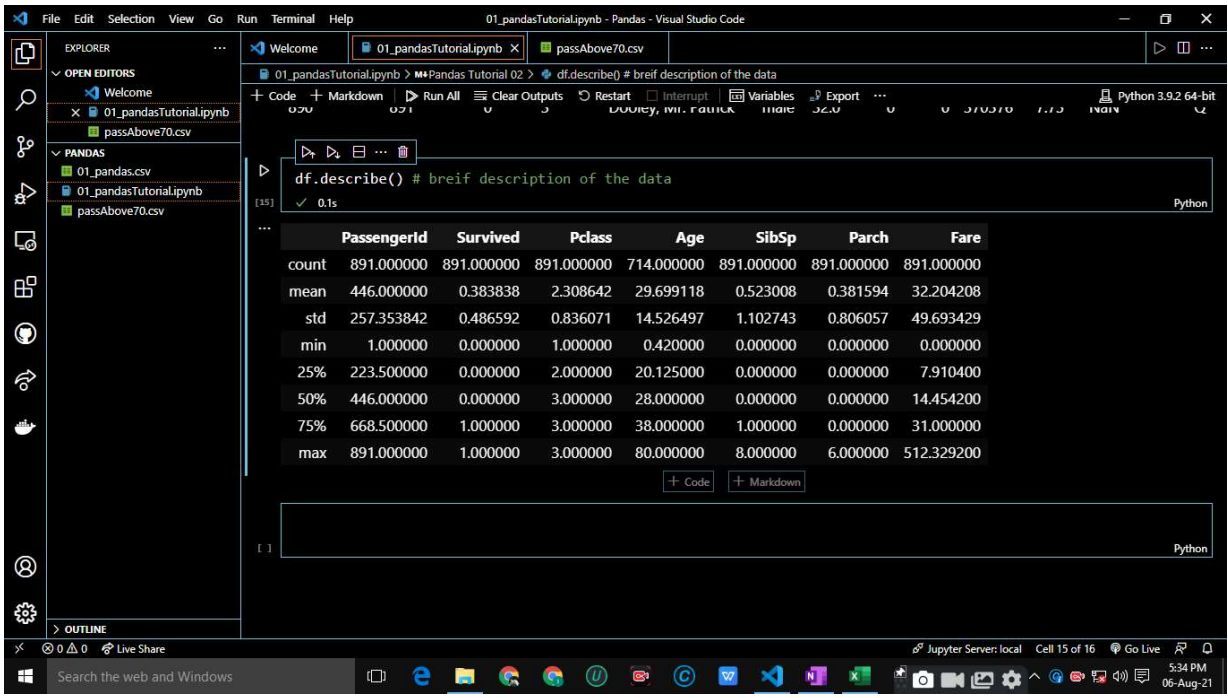


	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

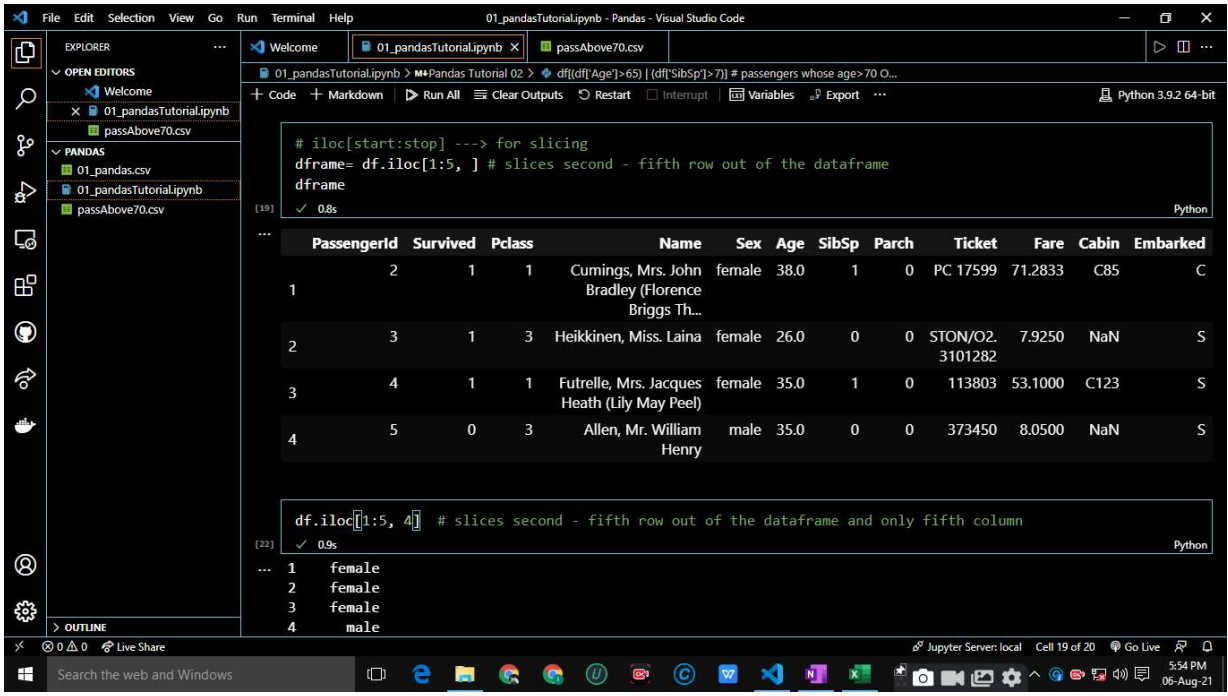


	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75	NaN	Q

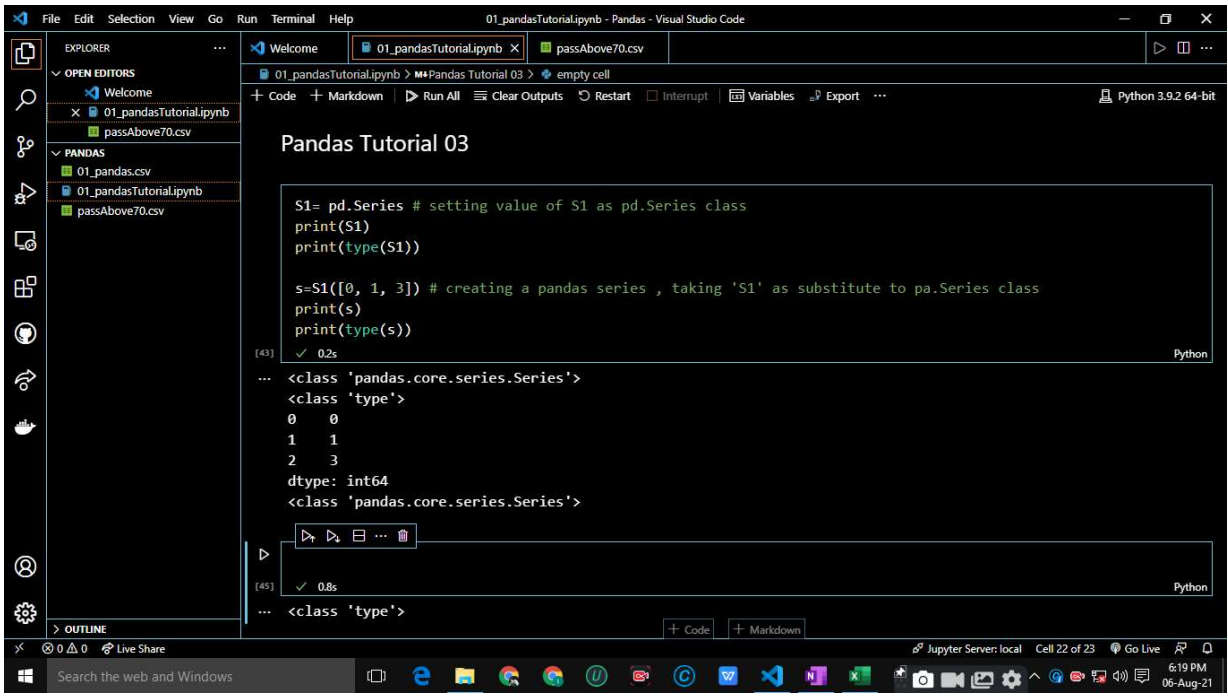
	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208

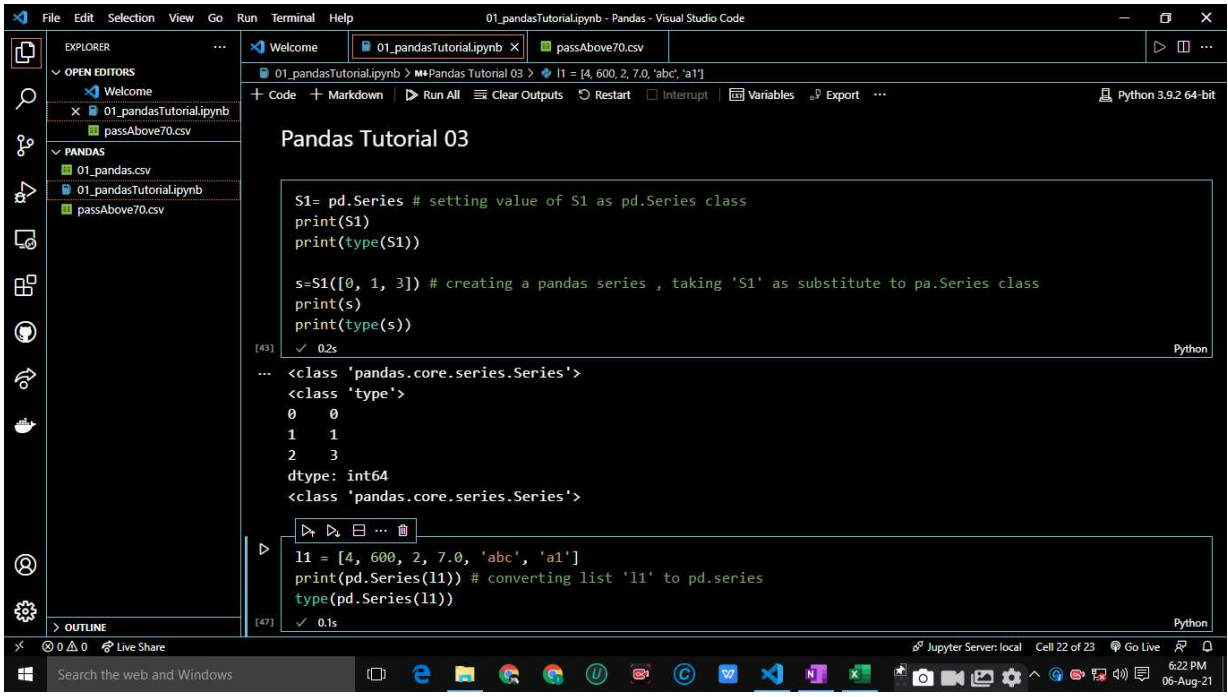


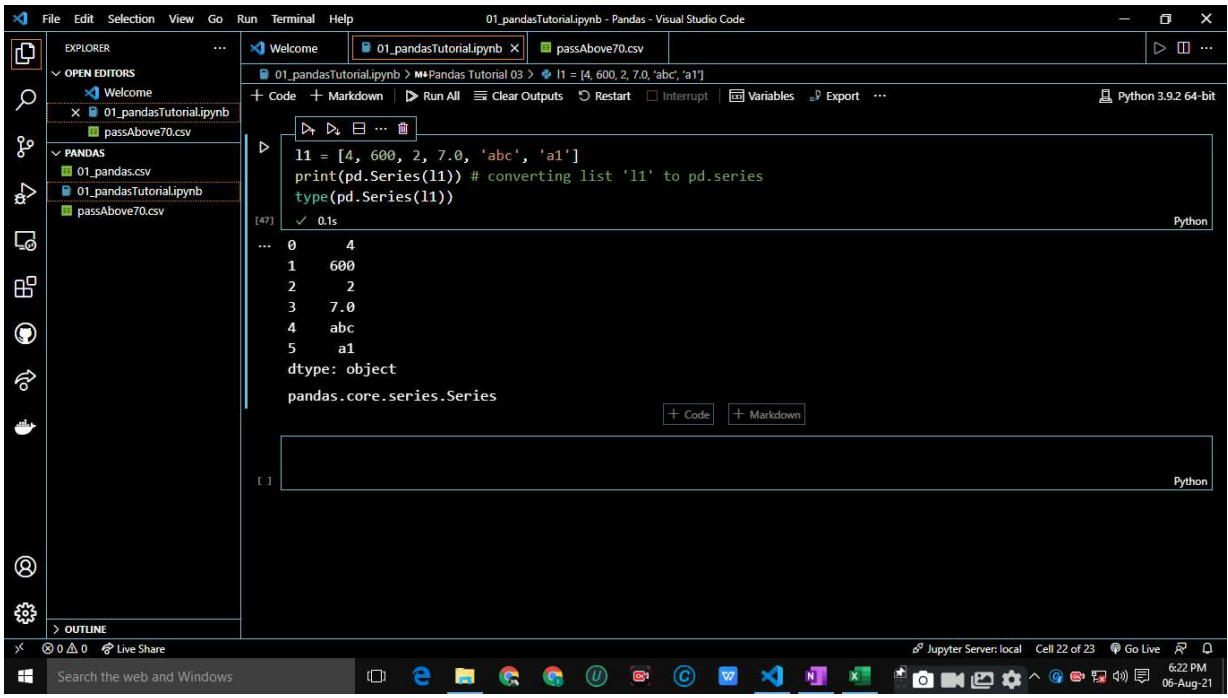
	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

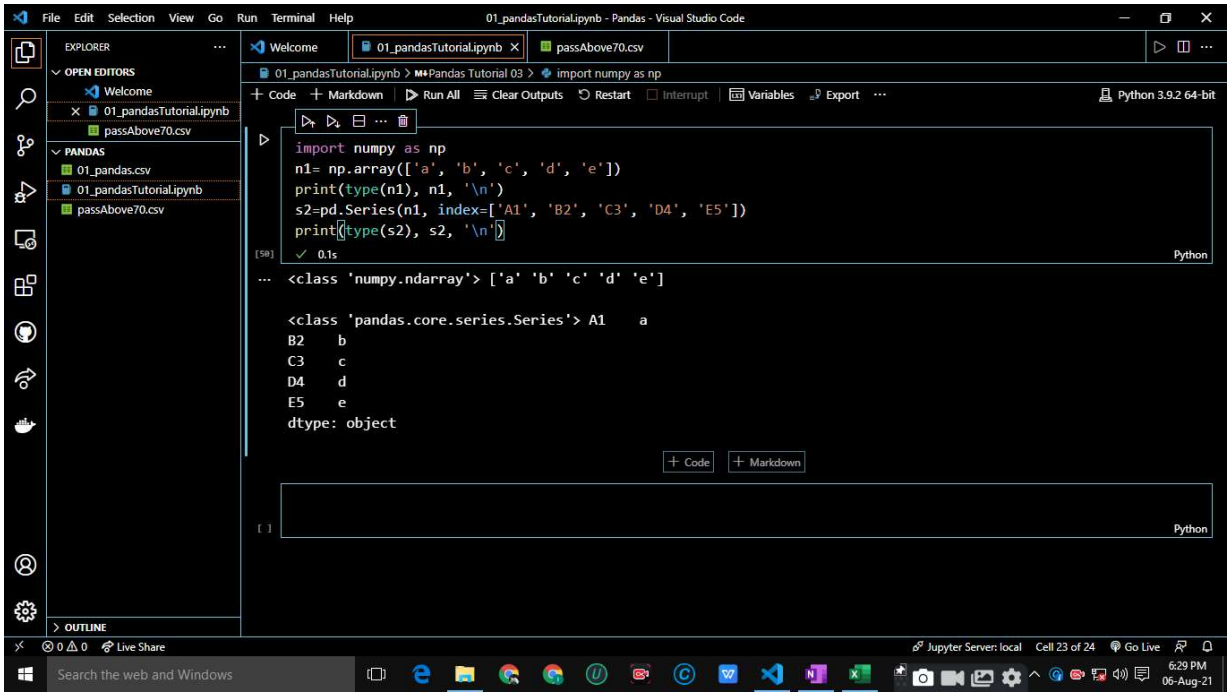


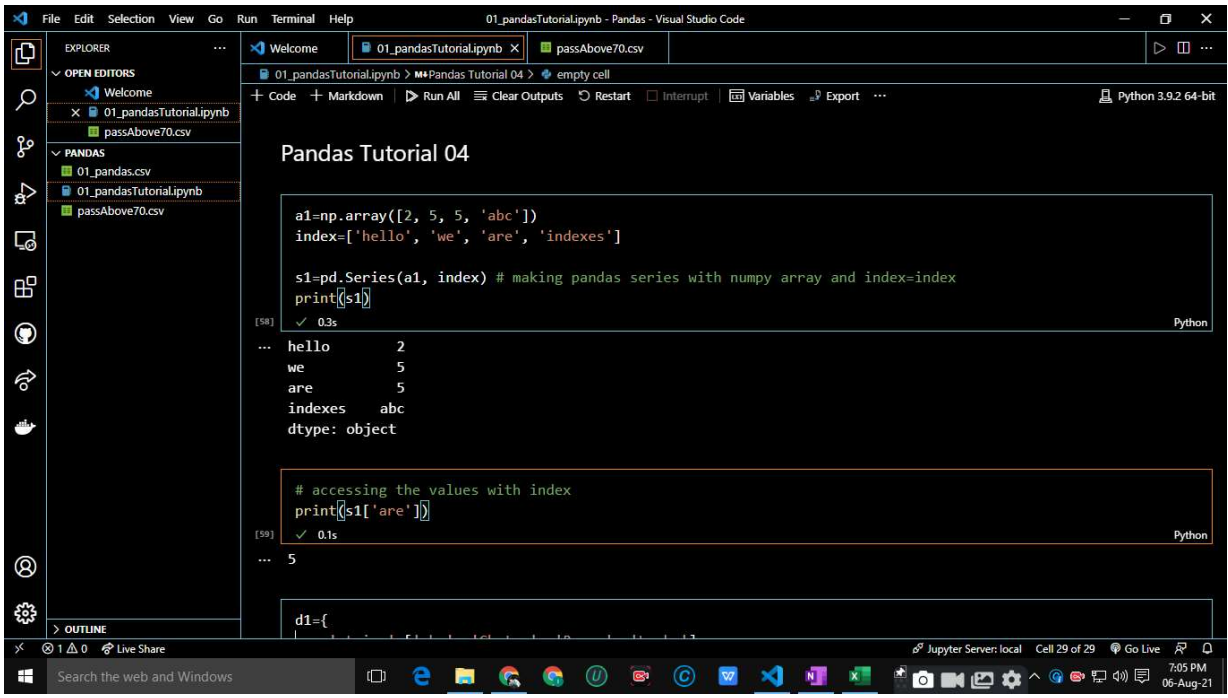
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

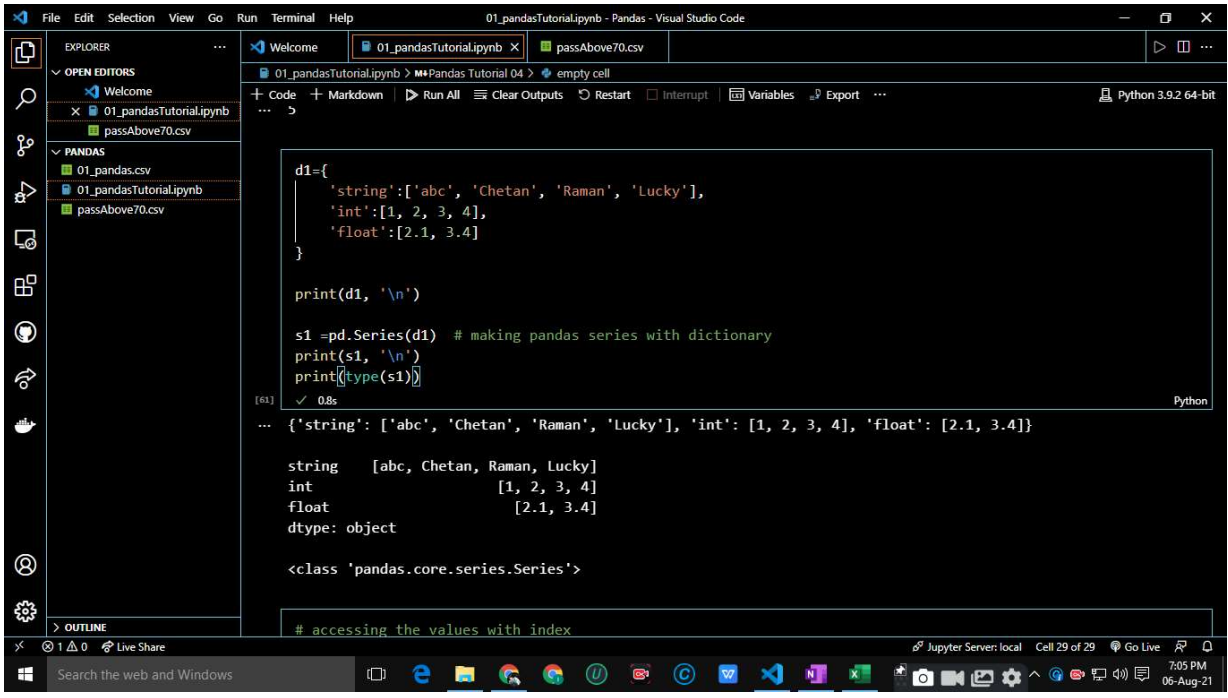

File Edit Selection View Go Run Terminal Help
01_pandasTutorial.ipynb - Pandas - Visual Studio Code
EXPLORER
OPEN EDITORS
Welcome
01_pandasTutorial.ipynb
passAbove70.csv
PANDAS
01_pandas.csv
01_pandasTutorial.ipynb
passAbove70.csv
OUTLINE
Welcome
01_pandasTutorial.ipynb
passAbove70.csv
01_pandasTutorial.ipynb > Pandas Tutorial 03 > empty cell
+ Code + Markdown Run All Clear Outputs Restart Interrupt Variables Export
Python 3.9.2 64-bit
Pandas Tutorial 03
S1= pd.Series # setting value of S1 as pd.Series class
print(S1)
print(type(S1))
s=S1([0, 1, 3]) # creating a pandas series , taking 'S1' as substitute to pa.Series class
print(s)
print(type(s))
[43] 0.2s
Python
<class 'pandas.core.series.Series'>
<class 'type'>
0 0
1 1
2 3
dtype: int64
<class 'pandas.core.series.Series'>
[45] 0.8s
Python
<class 'type'>
+ Code + Markdown
0 0 Live Share
Jupyter Server: local
Cell 22 of 23
Go Live
Search the web and Windows
6:19 PM
06-Aug-21

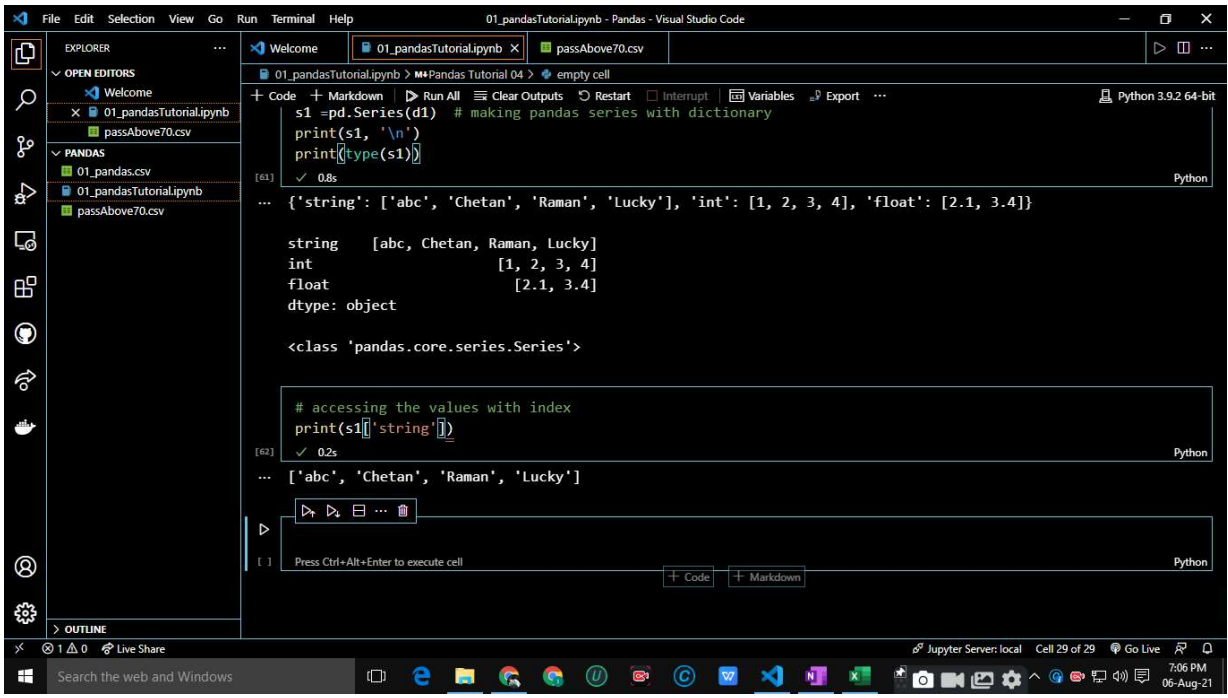

File Edit Selection View Go Run Terminal Help
01_pandasTutorial.ipynb - Pandas - Visual Studio Code
EXPLORER
OPEN EDITORS
Welcome
01_pandasTutorial.ipynb
passAbove70.csv
PANDAS
01_pandas.csv
01_pandasTutorial.ipynb
passAbove70.csv
OUTLINE
Welcome
01_pandasTutorial.ipynb
passAbove70.csv
01_pandasTutorial.ipynb > Pandas Tutorial 03 > l1 = [4, 600, 2, 7.0, 'abc', 'a1']
+ Code + Markdown Run All Clear Outputs Restart Interrupt Variables Export
Python 3.9.2 64-bit
Pandas Tutorial 03
S1= pd.Series # setting value of S1 as pd.Series class
print(S1)
print(type(S1))
s=S1([0, 1, 3]) # creating a pandas series , taking 'S1' as substitute to pa.Series class
print(s)
print(type(s))
[43] 0.2s
Python
<class 'pandas.core.series.Series'>
<class 'type'>
0 0
1 1
2 3
dtype: int64
<class 'pandas.core.series.Series'>
l1 = [4, 600, 2, 7.0, 'abc', 'a1']
print(pd.Series(l1)) # converting list 'l1' to pd.series
type(pd.Series(l1))
[47] 0.1s
Python
0 0 Live Share
Jupyter Server: local
Cell 22 of 23
Go Live
Search the web and Windows
6:22 PM
06-Aug-21


01_pandasTutorial.ipynb - Pandas - Visual Studio Code
EXPLORER
OPEN EDITORS
Welcome
01_pandasTutorial.ipynb
passAbove70.csv
PANDAS
01_pandas.csv
01_pandasTutorial.ipynb
passAbove70.csv
OUTLINE
Welcome
01_pandasTutorial.ipynb
passAbove70.csv
01_pandasTutorial.ipynb > Pandas Tutorial 03 > l1 = [4, 600, 2, 7.0, 'abc', 'a1']
Code
Markdown
Run All
Clear Outputs
Restart
Interrupt
Variables
Export
Python 3.9.2 64-bit
l1 = [4, 600, 2, 7.0, 'abc', 'a1']
print(pd.Series(l1)) # converting list 'l1' to pd.series
type(pd.Series(l1))
[47]
0.1s
Python
0 4
1 600
2 2
3 7.0
4 abc
5 a1
dtype: object
pandas.core.series.Series
Code
Markdown
Python
Live Share
Jupyter Server: local
Cell 22 of 23
Go Live
Search the web and Windows
6:22 PM
06-Aug-21


01_pandasTutorial.ipynb - Pandas - Visual Studio Code
EXPLORER
OPEN EDITORS
Welcome
01_pandasTutorial.ipynb
passAbove70.csv
PANDAS
01_pandas.csv
01_pandasTutorial.ipynb
passAbove70.csv
OUTLINE
Welcome
01_pandasTutorial.ipynb
passAbove70.csv
01_pandasTutorial.ipynb > Pandas Tutorial 03 > import numpy as np
Code
Markdown
Run All
Clear Outputs
Restart
Interrupt
Variables
Export
Python 3.9.2 64-bit
import numpy as np
n1= np.array(['a', 'b', 'c', 'd', 'e'])
print(type(n1), n1, '\n')
s2=pd.Series(n1, index=['A1', 'B2', 'C3', 'D4', 'E5'])
print(type(s2), s2, '\n')
[58]
0.1s
Python
<class 'numpy.ndarray'> ['a' 'b' 'c' 'd' 'e']
<class 'pandas.core.series.Series'> A1 a
B2 b
C3 c
D4 d
E5 e
dtype: object
Code
Markdown
Python
Live Share
Jupyter Server: local
Cell 23 of 24
Go Live
Search the web and Windows
6:29 PM
06-Aug-21


File Edit Selection View Go Run Terminal Help
01_pandasTutorial.ipynb - Pandas - Visual Studio Code
EXPLORER
OPEN EDITORS
Welcome
01_pandasTutorial.ipynb
passAbove70.csv
PANDAS
01_pandas.csv
01_pandasTutorial.ipynb
passAbove70.csv
OUTLINE
Welcome
01_pandasTutorial.ipynb
passAbove70.csv
01_pandasTutorial.ipynb > Pandas Tutorial 04 > empty cell
+ Code + Markdown Run All Clear Outputs Restart Interrupt Variables Export
Python 3.9.2 64-bit
Pandas Tutorial 04
a1=np.array([2, 5, 5, 'abc'])
index=['hello', 'we', 'are', 'indexes']
s1=pd.Series(a1, index) # making pandas series with numpy array and index=index
print(s1)
[58] 0.3s
Python
hello 2
we 5
are 5
indexes abc
dtype: object
# accessing the values with index
print(s1['are'])
[59] 0.1s
Python
5
d1={
1 ⚠ 0
Live Share
Jupyter Server: local
Cell 29 of 29
Go Live
Search the web and Windows
7:05 PM
06-Aug-21


File Edit Selection View Go Run Terminal Help
01_pandasTutorial.ipynb - Pandas - Visual Studio Code
EXPLORER
OPEN EDITORS
Welcome
01_pandasTutorial.ipynb
passAbove70.csv
PANDAS
01_pandas.csv
01_pandasTutorial.ipynb
passAbove70.csv
OUTLINE
01_pandasTutorial.ipynb > Pandas Tutorial 04 > empty cell
+ Code + Markdown Run All Clear Outputs Restart Interrupt Variables Export
Python 3.9.2 64-bit
5
d1={
'string':['abc', 'Chetan', 'Raman', 'Lucky'],
'int':[1, 2, 3, 4],
'float':[2.1, 3.4]
}
print(d1, '\n')
s1 =pd.Series(d1) # making pandas series with dictionary
print(s1, '\n')
print(type(s1))
[61] 0.8s
Python
{'string': ['abc', 'Chetan', 'Raman', 'Lucky'], 'int': [1, 2, 3, 4], 'float': [2.1, 3.4]}
string [abc, Chetan, Raman, Lucky]
int [1, 2, 3, 4]
float [2.1, 3.4]
dtype: object
<class 'pandas.core.series.Series'>
accessing the values with index
1 ⚠ 0
Live Share
Jupyter Server: local
Cell 29 of 29
Go Live
Search the web and Windows
7:05 PM
06-Aug-21


File Edit Selection View Go Run Terminal Help
01_pandasTutorial.ipynb - Pandas - Visual Studio Code
EXPLORER
OPEN EDITORS
Welcome
01_pandasTutorial.ipynb
passAbove70.csv
PANDAS
01_pandas.csv
01_pandasTutorial.ipynb
passAbove70.csv
OUTLINE
Welcome
01_pandasTutorial.ipynb
passAbove70.csv
01_pandasTutorial.ipynb > Pandas Tutorial 04 > empty cell
+ Code + Markdown Run All Clear Outputs Restart Interrupt Variables Export
Python 3.9.2 64-bit
s1 =pd.Series(d1) # making pandas series with dictionary
print(s1, '\n')
print(type(s1))
[61] 0.8s
Python
{'string': ['abc', 'Chetan', 'Raman', 'Lucky'], 'int': [1, 2, 3, 4], 'float': [2.1, 3.4]}
string [abc, Chetan, Raman, Lucky]
int [1, 2, 3, 4]
float [2.1, 3.4]
dtype: object
<class 'pandas.core.series.Series'>
# accessing the values with index
print(s1['string'])
[62] 0.2s
Python
['abc', 'Chetan', 'Raman', 'Lucky']
[ ]
Press Ctrl+Alt+Enter to execute cell
Python
+ Code + Markdown
1 0 Live Share
Jupyter Server: local Cell 29 of 29 Go Live
Search the web and Windows
7:06 PM
06-Aug-21