Torrodjae Somerville
CTEC 298-101: Symbolic Computation Using Big Data
Dr.Bemley
11/06/2025

Jupyter Pandas Tutorials 1-4 Last Checkpoint: 2 hours ago

File   Edit   View   Run   Kernel   Settings   Help

Code   ⌄   JupyterLab ⧉   ❄   Python [conda env:base] * ◯ ≡ ◯Anaconda Toolbox   Trusted

```python
import pandas as pd
df = pd.read_csv('C:\\Users\\TorrodjaeS\\Documents\\datasettt\\FINAL CTEC 128 DATASET.csv')
df
```

[3]:

nore	Baltimore_City	Calvert	Caroline	Carroll	Cecil	...	Prince_Georges	Queen_Annes	Somerset	St_Marys	Talbot	Washington	Wicomico	Worcester	Total Cases	Dai Ne Cas
3	1	0	0	1	0	...	9	0	0	0	0	0	0	0	31	
4	1	0	0	1	0	...	15	0	0	0	1	0	0	0	37	
6	1	0	0	1	0	...	14	0	0	0	1	0	0	0	57	
10	4	0	0	1	0	...	20	0	0	0	1	0	0	0	85	
12	8	1	0	2	0	...	23	0	0	0	1	0	0	0	107	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
8271	163906	16071	7881	30229	22441	...	250779	9497	6664	26713	7957	42743	27109	11624	1490853	
8274	163909	16071	7881	30231	22442	...	250783	9497	6664	26715	7957	42744	27109	11624	1490886	
8277	163912	16071	7881	30231	22442	...	250786	9497	6664	26717	7957	42744	27110	11624	1490908	
8283	163913	16071	7881	30233	22442	...	250789	9497	6665	26717	7957	42744	27111	11625	1490931	
8297	163926	16073	7881	30234	22446	...	250808	9497	6665	26725	7957	42745	27112	11626	1491078	1

```python
[4]: df['Baltimore_City'].max()
```

[4]: 163926

File   Edit   View   Run   Kernel   Settings   Help

Trusted

Code

JupyterLab   Python [conda env:base] *   Anaconda Toolbox

```
[4]: 163926
```

```
[13]: df['DATE'][df['Total Cases']=='29']
```

```
[13]: Series([], Name: DATE, dtype: object)
```

```
[14]: df.fillna(0, inplace=True)
      df['Daily New Cases'].mean()
```

```
[14]: np.float64(864.3750724637681)
```

```
[15]: rows, columns = df.shape
```

```
[16]: rows
```

```
[16]: 1725
```

```
[18]: columns
```

```
[18]: 28
```

```
[31]: df.head()
```

[31]:

	OBJECTID	DATE	Allegany	Anne_Arundel	Baltimore	Baltimore_City	Calvert	Caroline	Carroll	Cecil	...	Prince_Georges	Queen_Annes	Somerset	St_Marys
0	1	03/15/2020 6:00:00 AM	0	2	3	1	0	0	1	0	...	9	0	0	0
1	2	03/16/2020 6:00:00 AM	0	1	4	1	0	0	1	0	...	15	0	0	0
2	3	03/17/2020 6:00:00 AM	0	3	6	1	0	0	1	0	...	14	0	0	0
3	4	03/18/2020 6:00:00 AM	0	4	10	4	0	0	1	0	...	20	0	0	0
4	5	03/19/2020 6:00:00 AM	0	5	12	8	1	0	2	0	...	23	0	0	0

5 rows × 28 columns

```
[20]: df.head(2)
```

Code

JupyterLab ⬈ ⚙ Python [conda env:base] * ◯ ☰ ⬤Anaconda Toolbox ⊞

[20]:

	OBJECTID	DATE	Allegany	Anne_Arundel	Baltimore	Baltimore_City	Calvert	Caroline	Carroll	Cecil	...	Prince_Georges	Queen_Annes	Somerset	St_Marys
0	1	03/15/2020 6:00:00 AM	0	2	3	1	0	0	1	0	...	9	0	0	0
1	2	03/16/2020 6:00:00 AM	0	1	4	1	0	0	1	0	...	15	0	0	0

2 rows × 28 columns

◀ ━━━━━━━━━━━━━━━━━━━━━━━━ ▶

[21]: `df.tail()`

[21]:

	OBJECTID	DATE	Allegany	Anne_Arundel	Baltimore	Baltimore_City	Calvert	Caroline	Carroll	Cecil	...	Prince_Georges	Queen_Annes	Somerset	St_Mar
1720	1721	11/29/2024 5:00:00 AM	24676	129860	188271	163906	16071	7881	30229	22441	...	250779	9497	6664	267
1721	1722	11/30/2024 5:00:00 AM	24678	129862	188274	163909	16071	7881	30231	22442	...	250783	9497	6664	267
1722	1723	12/01/2024 5:00:00 AM	24678	129863	188277	163912	16071	7881	30231	22442	...	250786	9497	6664	267
1723	1724	12/02/2024 5:00:00 AM	24679	129863	188283	163913	16071	7881	30233	22442	...	250789	9497	6665	267
1724	1725	12/03/2024 5:00:00 AM	24690	129879	188297	163926	16073	7881	30234	22446	...	250808	9497	6665	267

5 rows × 28 columns

◀ ━━━━━━━━━━━━━━━━━━━━━━━━ ▶

[23]: `df.tail(1)`

[23]:

	OBJECTID	DATE	Allegany	Anne_Arundel	Baltimore	Baltimore_City	Calvert	Caroline	Carroll	Cecil	...	Prince_Georges	Queen_Annes	Somerset	St_Mar
1724	1725	12/03/2024 5:00:00 AM	24690	129879	188297	163926	16073	7881	30234	22446	...	250808	9497	6665	267

1 rows × 28 columns

File   Edit   View   Run   Kernel   Settings   Help

Code

JupyterLab ⬈   ⚙   Python [conda env:base] * ◯ ≡   ◯ Anaconda Toolbox ⊞

```
[24]: df[2:5]
```

[24]:

	OBJECTID	DATE	Allegany	Anne_Arundel	Baltimore	Baltimore_City	Calvert	Caroline	Carroll	Cecil	...	Prince_Georges	Queen_Annes	Somerset	St_Marys
2	3	03/17/2020 6:00:00 AM	0	3	6	1	0	0	1	0	...	14	0	0	0
3	4	03/18/2020 6:00:00 AM	0	4	10	4	0	0	1	0	...	20	0	0	0
4	5	03/19/2020 6:00:00 AM	0	5	12	8	1	0	2	0	...	23	0	0	0

3 rows × 28 columns

```
[25]: df.columns
```

```
[25]: Index(['OBJECTID', 'DATE', 'Allegany', 'Anne_Arundel', 'Baltimore',
             'Baltimore_City', 'Calvert', 'Caroline', 'Carroll', 'Cecil', 'Charles',
             'Dorchester', 'Frederick', 'Garrett', 'Harford', 'Howard', 'Kent',
             'Montgomery', 'Prince_Georges', 'Queen_Annes', 'Somerset', 'St_Marys',
             'Talbot', 'Washington', 'Wicomico', 'Worcester', 'Total Cases',
             'Daily New Cases'],
            dtype='object')
```

```
[26]: df.DATE
```

```
[26]: 0       03/15/2020 6:00:00 AM
      1       03/16/2020 6:00:00 AM
      2       03/17/2020 6:00:00 AM
      3       03/18/2020 6:00:00 AM
      4       03/19/2020 6:00:00 AM
                     ...
      1720    11/29/2024 5:00:00 AM
      1721    11/30/2024 5:00:00 AM
      1722    12/01/2024 5:00:00 AM
      1723    12/02/2024 5:00:00 AM
      1724    12/03/2024 5:00:00 AM
      Name: DATE, Length: 1725, dtype: object
```

```
[27]: df[['DATE','Baltimore','Anne_Arundel']]
```

[27]:

	DATE	Baltimore	Anne_Arundel
0	03/15/2020 6:00:00 AM	2	2

[27]:

	DATE	Baltimore	Anne_Arundel
0	03/15/2020 6:00:00 AM	3	2
1	03/16/2020 6:00:00 AM	4	1
2	03/17/2020 6:00:00 AM	6	3
3	03/18/2020 6:00:00 AM	10	4
4	03/19/2020 6:00:00 AM	12	5
...	...	...	...
1720	11/29/2024 5:00:00 AM	188271	129860
1721	11/30/2024 5:00:00 AM	188274	129862
1722	12/01/2024 5:00:00 AM	188277	129863
1723	12/02/2024 5:00:00 AM	188283	129863
1724	12/03/2024 5:00:00 AM	188297	129879

1725 rows × 3 columns

```
[28]: df['Baltimore']
```

```
[28]: 0          3
      1          4
      2          6
      3         10
      4         12
              ...
      1720    188271
      1721    188274
      1722    188277
      1723    188283
      1724    188297
      Name: Baltimore, Length: 1725, dtype: int64
```

```
[30]: df['Baltimore'].max()
```

```
[30]: 188297
```

```
[32]: df['Baltimore'].min()
```

```
[32]: 3
```

```
[33]: df['Baltimore'].std()
```

```
[33]: 64914.57868153678
```

```
[34]: df.describe()
```

[34]:

	Cecil	Charles	...	Prince_Georges	Queen_Annes	Somerset	St_Marys	Talbot	Washington	Wicomico	Worcester	Total Cases	Daily New Cases
	5.000000	1725.000000	...	1725.000000	1725.000000	1725.000000	1725.000000	1725.000000	1725.000000	1725.000000	1725.000000	1.725000e+03	1725.000000
	1.844638	25873.725217	...	159564.041159	6091.724058	4400.736232	16347.215072	4918.933913	28400.811594	17103.313623	7519.427826	9.345538e+05	864.375072
	7.864015	15369.193326	...	84951.600961	3413.130472	2320.531903	9911.608130	2876.246148	15767.499989	9680.363721	4161.513457	5.244203e+05	1527.766272
	0.000000	1.000000	...	9.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	3.100000e+01	0.000000
	2.000000	10748.000000	...	84608.000000	2976.000000	2582.000000	5990.000000	2146.000000	14483.000000	7604.000000	3616.000000	4.570840e+05	189.000000
	3.000000	32720.000000	...	194512.000000	7808.000000	5534.000000	21180.000000	6193.000000	37226.000000	21604.000000	9563.000000	1.171947e+06	497.000000
	2.000000	39697.000000	...	235723.000000	9035.000000	6309.000000	25102.000000	7468.000000	41639.000000	25548.000000	11179.000000	1.401952e+06	989.000000
	5.000000	42728.000000	...	250808.000000	9497.000000	6665.000000	26725.000000	7957.000000	42745.000000	27112.000000	11626.000000	1.491078e+06	17872.000000

```
[37]: df[['Baltimore','DATE']][df.DATE==df['DATE'].max()]
```

[37]:

	Baltimore	DATE
1386	179933	12/31/2023 5:00:00 AM

```
[38]: df
```

[38]:

	OBJECTID	DATE	Allegany	Anne_Arundel	Baltimore	Baltimore_City	Calvert	Caroline	Carroll	Cecil	...	Prince_Georges	Queen_Annes	Somerset	St_Mar
0	1	03/15/2020 6:00:00 AM	0	2	3	1	0	0	1	0	...	9	0	0	
1	2	03/16/2020 6:00:00 AM	0	1	4	1	0	0	1	0	...	15	0	0	
2	3	03/17/2020 6:00:00 AM	0	3	6	1	0	0	1	0	...	14	0	0	

3	4	03/18/2020 6:00:00 AM	0	4	10	4	0	0	1	0	...	20	0	0	
4	5	03/19/2020 6:00:00 AM	0	5	12	8	1	0	2	0	...	23	0	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
1720	1721	11/29/2024 5:00:00 AM	24676	129860	188271	163906	16071	7881	30229	22441	...	250779	9497	6664	267
1721	1722	11/30/2024 5:00:00 AM	24678	129862	188274	163909	16071	7881	30231	22442	...	250783	9497	6664	267
1722	1723	12/01/2024 5:00:00 AM	24678	129863	188277	163912	16071	7881	30231	22442	...	250786	9497	6664	267
1723	1724	12/02/2024 5:00:00 AM	24679	129863	188283	163913	16071	7881	30233	22442	...	250789	9497	6665	267
1724	1725	12/03/2024 5:00:00 AM	24690	129879	188297	163926	16073	7881	30234	22446	...	250808	9497	6665	267

1725 rows × 28 columns

```python
[39]: df.index
```

```
[39]: RangeIndex(start=0, stop=1725, step=1)
```

```python
[40]: df.set_index('DATE')
```

[40]:

DATE	OBJECTID	Allegany	Anne_Arundel	Baltimore	Baltimore_City	Calvert	Caroline	Carroll	Cecil	Charles	...	Prince_Georges	Queen_Annes	Somerset	St_
03/15/2020 6:00:00 AM	1	0	2	3	1	0	0	1	0	1	...	9	0	0	
03/16/2020 6:00:00 AM	2	0	1	4	1	0	0	1	0	1	...	15	0	0	
03/17/2020 6:00:00 AM	3	0	3	6	1	0	0	1	0	1	...	14	0	0	
03/18/2020 6:00:00 AM	4	0	4	10	4	0	0	1	0	2	...	20	0	0	

											...			
11/29/2024 5:00:00 AM	1721	24676	129860	188271	163906	16071	7881	30229	22441	42722	...	250779	9497	6664
11/30/2024 5:00:00 AM	1722	24678	129862	188274	163909	16071	7881	30231	22442	42722	...	250783	9497	6664
12/01/2024 5:00:00 AM	1723	24678	129863	188277	163912	16071	7881	30231	22442	42723	...	250786	9497	6664
12/02/2024 5:00:00 AM	1724	24679	129863	188283	163913	16071	7881	30233	22442	42723	...	250789	9497	6665
12/03/2024 5:00:00 AM	1725	24690	129879	188297	163926	16073	7881	30234	22446	42728	...	250808	9497	6665

1725 rows × 27 columns

```
[41]: df.set_index('Baltimore',inplace=True)
```

```
[42]: df
```

[42]:

DATE	Allegany	Anne_Arundel	Baltimore_City	Calvert	Caroline	Carroll	Cecil	Charles	...	Prince_Georges	Queen_Annes	Somerset	St_Marys	Talbot	Washingto
3/15/2020 5:00:00 AM	0	2	1	0	0	1	0	1	...	9	0	0	0	0	
3/16/2020 5:00:00 AM	0	1	1	0	0	1	0	1	...	15	0	0	0	1	
3/17/2020 5:00:00 AM	0	3	1	0	0	1	0	1	...	14	0	0	0	1	
3/18/2020 5:00:00 AM	0	4	4	0	0	1	0	2	...	20	0	0	0	1	
3/19/2020 5:00:00 AM	0	5	8	1	0	2	0	2	...	23	0	0	0	1	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
1/29/2024 5:00:00 AM	24676	129860	163906	16071	7881	30229	22441	42722	...	250779	9497	6664	26713	7957	4274
1/30/2024 5:00:00 AM	24678	129862	163909	16071	7881	30231	22442	42722	...	250783	9497	6664	26715	7957	4274

File   Edit   View   Run   Kernel   Settings   Help

Trusted

Code

JupyterLab   Python [conda env:base] *   Anaconda Toolbox

```python
[51]: weather_data = {
          'day': ['1/1/2017','1/2/2017','1/3/2017'],
          'temperature':[32,35,28],
          'windspeed': [6,7,2],
          'event': ['Rain', 'Sunny', 'Snow']
      }
      df = pd.DataFrame(weather_data)
      df
```

[51]:

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/2/2017	35	7	Sunny
2	1/3/2017	28	2	Snow

```python
[52]: weather_data = [
          ('1/1/2017',32,6,'Rain'),
          ('1/2/2017',35,7,'Sunny'),
          ('1/3/2017',28,2,'Snow')
      ]
      df = pd.DataFrame(weather_data, columns=["day","temperature","windspeed","event"])
      df
```

[52]:

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/2/2017	35	7	Sunny
2	1/3/2017	28	2	Snow

```python
[53]: df
```

[53]:

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/2/2017	35	7	Sunny
2	1/3/2017	28	2	Snow

```python
[55]: import pandas as pd
      df = pd.read_csv('C:\\Users\\Torrodjaes\\Documents\\datasettt\\FINAL CTEC 128 DATASET.csv',nrows=3)
      df
```

	OBJECTID	DATE	Allegany	Anne_Arundel	Baltimore	Baltimore_City	Calvert	Caroline	Carroll	Cecil	...	Prince_Georges	Queen_Annes	Somerset	St_Marys
0	1	03/15/2020 6:00:00 AM	0	2	3	1	0	0	1	0	...	9	0	0	0
1	2	03/16/2020 6:00:00 AM	0	1	4	1	0	0	1	0	...	15	0	0	0
2	3	03/17/2020 6:00:00 AM	0	3	6	1	0	0	1	0	...	14	0	0	0

3 rows × 28 columns

```
[56]: df.to_csv('new.csv')
```

```
[57]: df.to_csv('new.csv', index=False)
```

```
[58]: df.columns
```

```
[58]: Index(['OBJECTID', 'DATE', 'Allegany', 'Anne_Arundel', 'Baltimore',
       'Baltimore_City', 'Calvert', 'Caroline', 'Carroll', 'Cecil', 'Charles',
       'Dorchester', 'Frederick', 'Garrett', 'Harford', 'Howard', 'Kent',
       'Montgomery', 'Prince_Georges', 'Queen_Annes', 'Somerset', 'St_Marys',
       'Talbot', 'Washington', 'Wicomico', 'Worcester', 'Total Cases',
       'Daily New Cases'],
      dtype='object')
```

```
[59]: df.to_csv("new.csv" ,columns=['DATE', 'Baltimore'])
```

```
[60]: df.to_excel("new.xlsx",sheet_name="stocks", startrow=1, startcol=2)
```

```
[61]: df_stocks = pd.DataFrame({
          'tickers':['GOOGL', 'WMT', 'MSFT'],
          'price':[845,65,64],
          'pe':[30.37,14.26,30.97],
          'eps':[27.82,4.61,2.12]
      })

      df_weather = pd.DataFrame({
          'day': ['1/1/2017','1/2/2017','1/3/2017'],
          'temperature': [32,35,28],
          'event': ['Rain', 'Sunny', 'Snow']
      })
```

File   Edit   View   Run   Kernel   Settings   Help

Code

JupyterLab   Python [conda env:base] *   Anaconda Toolbox

```python
[56]: df.to_csv('new.csv')
```

```python
[57]: df.to_csv('new.csv', index=False)
```

```python
[58]: df.columns
```

```
[58]: Index(['OBJECTID', 'DATE', 'Allegany', 'Anne_Arundel', 'Baltimore',
             'Baltimore_City', 'Calvert', 'Caroline', 'Carroll', 'Cecil', 'Charles',
             'Dorchester', 'Frederick', 'Garrett', 'Harford', 'Howard', 'Kent',
             'Montgomery', 'Prince_Georges', 'Queen_Annes', 'Somerset', 'St_Marys',
             'Talbot', 'Washington', 'Wicomico', 'Worcester', 'Total Cases',
             'Daily New Cases'],
           dtype='object')
```

```python
[59]: df.to_csv("new.csv" ,columns=['DATE', 'Baltimore'])
```

```python
[60]: df.to_excel("new.xlsx",sheet_name="stocks", startrow=1, startcol=2)
```

```python
[61]: df_stocks = pd.DataFrame({
          'tickers':['GOOGL', 'WMT', 'MSFT'],
          'price':[845,65,64],
          'pe':[30.37,14.26,30.97],
          'eps':[27.82,4.61,2.12]
      })

      df_weather = pd.DataFrame({
          'day': ['1/1/2017','1/2/2017','1/3/2017'],
          'temperature': [32,35,28],
          'event': ['Rain', 'Sunny', 'Snow']
      })
```

```python
[62]: with pd.ExcelWriter('stocks_weather.xlsx')as writer:
          df_stocks.to_excel(writer, sheet_name="stocks")
          df_weather.to_excel(writer, sheet_name="weather")
```

```python
[ ]:
```

	OBJECTID	DATE	Allegany	nne_Arund	Baltimore	ltimore_Ci	Calvert	Caroline	Carroll	Cecil	Charles	Dorchester	Frederick	Garrett
0	1	03/15/202	0	2	3	1	0	0	1	0	1	0	0	0
1	2	03/16/202	0	1	4	1	0	0	1	0	1	0	0	0
2	3	03/17/202	0	3	6	1	0	0	1	0	1	0	1	0

stocks sheet

	A	B	C	D	E
1		tickers	price	pe	eps
2	0	GOOGL	845	30.37	27.82
3	1	WMT	65	14.26	4.61
4	2	MSFT	64	30.97	2.12

Tabs: stocks | weather

weather sheet

	A	B	C	D
1		day	temperature	event
2	0	1/1/2017	32	Rain
3	1	1/2/2017	35	Sunny
4	2	1/3/2017	28	Snow

Tabs: stocks | weather