Data Input and Output with Pandas

1. CSV
2. Excle
3. HTML
4. SQL

> **Important library to input and output**
>
> 1. conda install sqlalchemy
> 2. conda install lxml
> 3. conda install html5lib
> 4. conda install BeautifulSoup4
> 5. conda install xlrd

In [1]:
```
pwd
```

Out[1]:
```
'/home/chandan/Data Science'
```

In [2]:
```
import pandas as pd
```

Read data from comma seprated file

In [3]:
```
pd.read_csv('./Weather Data.csv')
```

Out[3]:

	Unnamed: 0	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
0	0	1901	17.99	19.43	23.49	26.41	28.28	28.60	27.49	26.98	26.26	25.08	21.73	18.95
1	1	1902	19.00	20.39	24.10	26.54	28.68	28.44	27.29	27.05	25.95	24.37	21.33	18.78
2	2	1903	18.32	19.79	22.46	26.03	27.93	28.41	28.04	26.63	26.34	24.57	20.96	18.29
3	3	1904	17.77	19.39	22.95	26.73	27.83	27.85	26.84	26.73	25.84	24.36	21.07	18.84
4	4	1905	17.40	17.79	21.78	24.84	28.32	28.69	27.67	27.47	26.29	26.16	22.07	18.71
5	5	1906	17.50	19.14	22.21	26.53	29.06	28.02	27.46	26.82	26.23	24.75	21.93	19.55
6	6	1907	19.27	19.42	22.03	26.98	27.52	27.66	27.28	26.38	26.26	24.72	22.11	18.46
7	7	1908	18.35	19.73	22.93	27.06	28.07	28.49	27.16	26.41	25.74	24.25	21.06	18.15
8	8	1909	17.79	19.05	23.40	25.76	27.97	27.67	26.56	26.43	25.47	24.37	22.01	18.86
9	9	1910	18.14	19.72	22.90	25.96	28.36	27.72	26.93	26.61	25.98	24.04	20.72	18.05
10	10	1911	18.52	19.18	22.05	26.00	28.55	28.02	27.44	27.04	26.22	24.57	21.10	18.76
11	11	1912	18.60	20.84	22.93	26.21	28.30	28.53	27.49	26.68	25.81	24.44	21.00	18.44
12	12	1913	18.20	19.98	22.15	26.61	27.95	27.91	27.00	26.80	26.02	24.35	20.92	18.70
13	13	1914	18.96	19.66	22.63	25.73	28.24	28.46	27.29	26.49	26.38	23.97	21.87	18.73
14	14	1915	17.93	19.00	22.69	26.23	29.17	28.58	27.77	27.32	26.82	25.46	22.18	18.31
15	15	1916	18.52	19.68	24.24	26.84	28.16	27.81	27.08	26.77	26.00	24.60	21.03	18.17
16	16	1917	18.16	19.94	22.46	25.29	26.97	27.41	27.15	26.64	25.79	23.85	20.80	18.39
17	17	1918	17.25	19.58	22.64	25.26	28.27	27.58	27.74	26.97	26.00	24.54	21.71	18.37
18	18	1919	18.78	19.15	23.15	26.19	28.20	28.30	27.28	26.84	25.84	24.36	21.48	18.58
19	19	1920	18.27	19.16	23.02	25.73	27.06	27.76	27.48	26.61	26.34	24.85	21.84	18.78
20	20	1921	18.61	22.12	24.26	27.11	30.78	28.04	27.40	26.59	25.92	23.84	21.10	19.24
21	21	1922	18.30	20.50	23.65	26.55	29.80	28.03	27.26	26.95	26.04	23.94	21.28	18.39
22	22	1923	18.23	19.35	23.21	26.81	28.33	28.95	26.89	26.71	25.89	23.89	21.23	19.08
23	23	1924	18.06	19.97	24.12	26.87	27.43	28.60	27.50	26.91	25.72	24.50	21.04	18.97
24	24	1925	17.53	19.08	23.50	27.24	27.82	27.76	27.01	26.64	26.29	24.61	21.29	18.63
25	25	1926	18.65	20.70	22.87	25.30	27.69	28.97	27.65	27.08	25.95	24.41	20.59	18.80
26	26	1927	18.15	19.40	22.71	26.13	27.70	27.93	27.23	26.66	26.10	24.70	21.16	19.41
27	27	1928	18.66	20.13	23.23	26.44	28.88	28.22	27.28	26.90	26.32	24.92	21.77	18.73
28	28	1929	18.25	19.00	24.05	26.44	28.58	28.07	27.26	26.64	26.41	24.12	21.71	18.44
29	29	1930	17.75	18.72	23.36	25.94	28.11	27.84	27.31	26.85	26.17	24.60	21.41	19.11
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
87	87	1988	18.58	20.83	23.51	26.60	28.60	28.37	27.12	26.91	26.48	24.79	21.63	19.64
88	88	1989	17.68	19.85	23.00	25.95	28.13	27.72	26.96	26.86	26.38	24.84	21.80	19.23
89	89	1990	18.56	20.33	22.79	26.15	27.71	28.11	27.06	26.99	26.55	24.51	22.08	19.78
90	90	1991	17.99	20.43	23.58	26.13	28.35	28.17	27.59	26.88	26.65	24.56	21.60	19.44
91	91	1992	18.36	19.77	22.95	26.21	27.75	28.40	27.31	26.78	26.11	24.70	22.01	19.42
92	92	1993	18.37	20.84	22.71	26.50	28.73	28.50	27.24	27.29	26.34	24.88	22.20	19.55
93	93	1994	19.33	20.20	24.14	26.04	28.72	28.56	27.34	27.01	26.31	24.56	22.06	19.30
94	94	1995	19.18	21.48	24.26	27.23	29.34	29.88	27.80	27.40	27.15	26.00	22.99	20.65
95	95	1996	19.81	21.73	25.13	26.67	28.37	28.08	27.42	26.77	26.64	24.50	21.81	18.96
96	96	1997	17.86	19.88	23.64	25.55	27.86	28.33	28.01	27.27	26.81	24.48	22.05	19.31

In [4]:
```
df = pd.read_csv('./Weather Data.csv')
df
```

Out[4]:

	Unnamed: 0	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
0	0	1901	17.99	19.43	23.49	26.41	28.28	28.60	27.49	26.98	26.26	25.08	21.73	18.95
1	1	1902	19.00	20.39	24.10	26.54	28.68	28.44	27.29	27.05	25.95	24.37	21.33	18.78
2	2	1903	18.32	19.79	22.46	26.03	27.93	28.41	28.04	26.63	26.34	24.57	20.96	18.29
3	3	1904	17.77	19.39	22.95	26.73	27.83	27.85	26.84	26.73	25.84	24.36	21.07	18.84
4	4	1905	17.40	17.79	21.78	24.84	28.32	28.69	27.67	27.47	26.29	26.16	22.07	18.71
5	5	1906	17.50	19.14	22.21	26.53	29.06	28.02	27.46	26.82	26.23	24.75	21.93	19.55
6	6	1907	19.27	19.42	22.03	26.98	27.52	27.66	27.28	26.38	26.26	24.72	22.11	18.46
7	7	1908	18.35	19.73	22.93	27.06	28.07	28.49	27.16	26.41	25.74	24.25	21.06	18.15
8	8	1909	17.79	19.05	23.40	25.76	27.97	27.67	26.56	26.43	25.47	24.37	22.01	18.86
9	9	1910	18.14	19.72	22.90	25.96	28.36	27.72	26.93	26.61	25.98	24.04	20.72	18.05
10	10	1911	18.52	19.18	22.05	26.00	28.55	28.02	27.44	27.04	26.22	24.57	21.10	18.76
11	11	1912	18.60	20.84	22.93	26.21	28.30	28.53	27.49	26.68	25.81	24.44	21.00	18.44
12	12	1913	18.20	19.98	22.15	26.61	27.95	27.91	27.00	26.80	26.02	24.35	20.92	18.70
13	13	1914	18.96	19.66	22.63	25.73	28.24	28.46	27.29	26.49	26.38	23.97	21.87	18.73
14	14	1915	17.93	19.00	22.69	26.23	29.17	28.58	27.77	27.32	26.82	25.46	22.18	18.31
15	15	1916	18.52	19.68	24.24	26.84	28.16	27.81	27.08	26.77	26.00	24.60	21.03	18.17
16	16	1917	18.16	19.94	22.46	25.29	26.97	27.41	27.15	26.64	25.79	23.85	20.80	18.39
17	17	1918	17.25	19.58	22.64	25.26	28.27	27.58	27.74	26.97	26.00	24.54	21.71	18.37
18	18	1919	18.78	19.15	23.15	26.19	28.20	28.30	27.28	26.84	25.84	24.36	21.48	18.58
19	19	1920	18.27	19.16	23.02	25.73	27.06	27.76	27.48	26.61	26.34	24.85	21.84	18.78
20	20	1921	18.61	22.12	24.26	27.11	30.78	28.04	27.40	26.59	25.92	23.84	21.10	19.24
21	21	1922	18.30	20.50	23.65	26.55	29.80	28.03	27.26	26.95	26.04	23.94	21.28	18.39
22	22	1923	18.23	19.35	23.21	26.81	28.33	28.95	26.89	26.71	25.89	23.89	21.23	19.08
23	23	1924	18.06	19.97	24.12	26.87	27.43	28.60	27.50	26.91	25.72	24.50	21.04	18.97
24	24	1925	17.53	19.08	23.50	27.24	27.82	27.76	27.01	26.64	26.29	24.61	21.29	18.63
25	25	1926	18.65	20.70	22.87	25.30	27.69	28.97	27.65	27.08	25.95	24.41	20.59	18.80
26	26	1927	18.15	19.40	22.71	26.13	27.70	27.93	27.23	26.66	26.10	24.70	21.16	19.41
27	27	1928	18.66	20.13	23.23	26.44	28.88	28.22	27.28	26.90	26.32	24.92	21.77	18.73
28	28	1929	18.25	19.00	24.05	26.44	28.58	28.07	27.26	26.64	26.41	24.12	21.71	18.44
29	29	1930	17.75	18.72	23.36	25.94	28.11	27.84	27.31	26.85	26.17	24.60	21.41	19.11
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
87	87	1988	18.58	20.83	23.51	26.60	28.60	28.37	27.12	26.91	26.48	24.79	21.63	19.64
88	88	1989	17.68	19.85	23.00	25.95	28.13	27.72	26.96	26.86	26.38	24.84	21.80	19.23
89	89	1990	18.56	20.33	22.79	26.15	27.71	28.11	27.06	26.99	26.55	24.51	22.08	19.78
90	90	1991	17.99	20.43	23.58	26.13	28.35	28.17	27.59	26.88	26.65	24.56	21.60	19.44
91	91	1992	18.36	19.77	22.95	26.21	27.75	28.40	27.31	26.78	26.11	24.70	22.01	19.42
92	92	1993	18.37	20.84	22.71	26.50	28.73	28.50	27.24	27.29	26.34	24.88	22.20	19.55
93	93	1994	19.33	20.20	24.14	26.04	28.72	28.56	27.34	27.01	26.31	24.56	22.06	19.30
94	94	1995	19.18	21.48	24.26	27.23	29.34	29.88	27.80	27.40	27.15	26.00	22.99	20.65
95	95	1996	19.81	21.73	25.13	26.67	28.37	28.08	27.42	26.77	26.64	24.50	21.81	18.96
96	96	1997	17.86	19.88	23.64	25.55	27.86	28.33	28.01	27.27	26.81	24.48	22.05	19.31

Write data to csv file

In [5]:
```
df.to_csv('./WeatherIndiaTemp.csv', index=False)
```

Read data from excle file

In [6]:
```
pd.read_excel('./demo.xlsx')
```

Out[6]:

	A	B	C	D
0	0	1	2	3
1	4	5	6	7
2	8	9	10	11
3	12	13	14	15

In [7]:
```
df_excle = pd.read_excel('./demo.xlsx', sheet_name='Sheet1')
df_excle
```

Out[7]:

	A	B	C	D
0	0	1	2	3
1	4	5	6	7
2	8	9	10	11
3	12	13	14	15

In [8]:
```
df_excle.to_excel('demoexcle.xlsx', sheet_name='sheet2')
```

Read data from html

In [9]:
```
html_data = pd.read_html('https://www.fdic.gov/bank/individual/failed/banklist.html')
```

In [10]:
```
html_data
```

Out[10]: [

	Bank Name	City	ST	CERT	Acquiring Institution	Closing Date
0	Ericson State Bank	Ericson	NE	18265	Farmers and Merchants Bank	February 14, 2020
1	City National Bank of New Jersey	Newark	NJ	21111	Industrial Bank	November 1, 2019
2	Resolute Bank	Maumee	OH	58317	Buckeye State Bank	October 25, 2019
3	Louisa Community Bank	Louisa				
4	The Enloe State Bank	Cooper				
5	Washington Federal Bank for Savings	Chicago				
6	The Farmers and Merchants State Bank of Argonia	Argonia				
7	Fayette County Bank	Saint Elmo				
8	Guaranty Bank, (d/b/a BestBank in Georgia & Mi...	Milwaukee				
9	First NBC Bank	New Orleans				
10	Proficio Bank	Cottonwood Heights				
11	Seaway Bank and Trust Company	Chicago				
12	Harvest Community Bank	Pennsville				
13	Allied Bank	Mulberry				
14	The Woodbury Banking Company	Woodbury				
15	First CornerStone Bank	King of Prussia				
16	Trust Company Bank	Memphis				
17	North Milwaukee State Bank	Milwaukee				
18	Hometown National Bank	Longview				
19	The Bank of Georgia	Peachtree City				
20	Premier Bank	Denver				
21	Edgebrook Bank	Chicago				
22	Doral Bank En Español	San Juan				
23	Capitol City Bank & Trust Company	Atlanta				
24	Highland Community Bank	Chicago				
25	First National Bank of Crestview	Crestview				
26	Northern Star Bank	Mankato				
27	Frontier Bank, FSB D/B/A El Paseo Bank	Palm Desert				
28	The National Republic Bank of Chicago	Chicago				
29	NBRS Financial	Rising Sun				
..	...	...				
530	ANB Financial, NA	Bentonville				
531	Hume Bank	Hume				
532	Douglass National Bank	Kansas City				
533	Miami Valley Bank	Lakeview				
534	NetBank	Alpharetta				
535	Metropolitan Savings Bank	Pittsburgh				
536	Bank of Ephraim	Ephraim				
537	Reliance Bank	White Plains				
538	Guaranty National Bank of Tallahassee	Tallahassee				
539	Dollar Savings Bank	Newark				
540	Pulaski Savings Bank	Philadelphia				
541	First National Bank of Blanchardville	Blanchardville				
542	Southern Pacific Bank	Torrance				
543	Farmers Bank of Cheneyville	Cheneyville				
544	Bank of Alamo	Alamo				
545	AmTrade International Bank En Español	Atlanta				
546	Universal Federal Savings Bank	Chicago				
547	Connecticut Bank of Commerce	Stamford				
548	New Century Bank	Shelby Township				
549	Net 1st National Bank	Boca Raton				
550	NextBank, NA	Phoenix				
551	Oakwood Deposit Bank Co.	Oakwood				
552	Bank of Sierra Blanca	Sierra Blanca				
553	Hamilton Bank, NA En Español	Miami				
554	Sinclair National Bank	Gravette				
555	Superior Bank, FSB	Hinsdale				
556	Malta National Bank	Malta				
557	First Alliance Bank & Trust Co.	Manchester				
558	National State Bank of Metropolis	Metropolis				
559	Bank of Honolulu	Honolulu				

In [11]:
```
type(html_data)
```

Out[11]:
```
list
```

In [12]:
```
html_data[0]
```

Out[12]:

	Bank Name	City	ST	CERT	Acquiring Institution	Closing Date
0	Ericson State Bank	Ericson	NE	18265	Farmers and Merchants Bank	February 14, 2020
1	City National Bank of New Jersey	Newark	NJ	21111	Industrial Bank	November 1, 2019
2	Resolute Bank	Maumee	OH	58317	Buckeye State Bank	October 25, 2019
3	Louisa Community Bank	Louisa	KY	58112	Kentucky Farmers Bank Corporation	October 25, 2019
4	The Enloe State Bank	Cooper	TX	10716	Legend Bank, N. A.	May 31, 2019
5	Washington Federal Bank for Savings	Chicago	IL	30570	Royal Savings Bank	December 15, 2017
6	The Farmers and Merchants State Bank of Argonia	Argonia	KS	17719	Conway Bank	October 13, 2017
7	Fayette County Bank	Saint Elmo	IL	1802	United Fidelity Bank, fsb	May 26, 2017
8	Guaranty Bank, (d/b/a BestBank in Georgia & Mi...	Milwaukee	WI	30003	First-Citizens Bank & Trust Company	May 5, 2017
9	First NBC Bank	New Orleans	LA	58302	Whitney Bank	April 28, 2017
10	Proficio Bank	Cottonwood Heights	UT	35495	Cache Valley Bank	March 3, 2017
11	Seaway Bank and Trust Company	Chicago	IL	19328	State Bank of Texas	January 27, 2017
12	Harvest Community Bank	Pennsville	NJ	34951	First-Citizens Bank & Trust Company	January 13, 2017
13	Allied Bank	Mulberry	AR	91	Today's Bank	September 23, 2016
14	The Woodbury Banking Company	Woodbury	GA	11297	United Bank	August 19, 2016
15	First CornerStone Bank	King of Prussia	PA	35312	First-Citizens Bank & Trust Company	May 6, 2016
16	Trust Company Bank	Memphis	TN	9956	The Bank of Fayette County	April 29, 2016
17	North Milwaukee State Bank	Milwaukee	WI	20364	First-Citizens Bank & Trust Company	March 11, 2016
18	Hometown National Bank	Longview	WA	35156	Twin City Bank	October 2, 2015
19	The Bank of Georgia	Peachtree City	GA	35259	Fidelity Bank	October 2, 2015
20	Premier Bank	Denver	CO	34112	United Fidelity Bank, fsb	July 10, 2015
21	Edgebrook Bank	Chicago	IL	57772	Republic Bank of Chicago	May 8, 2015
22	Doral Bank En Español	San Juan	PR	32102	Banco Popular de Puerto Rico	February 27, 2015
23	Capitol City Bank & Trust Company	Atlanta	GA	33938	First-Citizens Bank & Trust Company	February 13, 2015
24	Highland Community Bank	Chicago	IL	20290	United Fidelity Bank, fsb	January 23, 2015
25	First National Bank of Crestview	Crestview	FL	17557	First NBC Bank	January 16, 2015
26	Northern Star Bank	Mankato	MN	34983	BankVista	December 19, 2014
27	Frontier Bank, FSB D/B/A El Paseo Bank	Palm Desert	CA	34738	Bank of Southern California, N.A.	November 7, 2014
28	The National Republic Bank of Chicago	Chicago	IL	916	State Bank of Texas	October 24, 2014

In [13]:
```
import psycopg2 as pg
import pandas.io.sql as psql
```

In [28]:
```
connection = pg.connect("host=localhost dbname=kinder user=chandan password=
08/08/2019")
```

In [29]:
```
connection
```

Out[29]:
```
<connection object at 0x7f6f93f4bd58; dsn: 'user=chandan password=xxx dbname=
kinder host=localhost', closed: 0>
```

In [30]:
```
product = psql.read_sql('SELECT * FROM product_product', connection)
```

In [31]:
```
product
```

Out[31]:

	id	message_main_attachment_id	default_code	active	product_tmpl_id	barcode	volume	weight	crea
0	1	None	INRF	True	1	None	None	None	

In [24]:
```
product_category = psql.read_sql_query('select * from product_category', con
nection)
```

In [37]:
```
product_category.drop('id', axis=1, inplace=True)
product_category
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
<ipython-input-37-af317c3f71fb> in <module>
----> 1 product_category.drop('id', axis=1, inplace=True)
      2 product_category

~/anaconda3/lib/python3.7/site-packages/pandas/core/frame.py in drop(self, la
bels, axis, index, columns, level, inplace, errors)
   3938                                           index=index, columns=colum
ns,
   3939                                           level=level, inplace=inpla
ce,
-> 3940                                           errors=errors)
   3941
   3942     @rewrite_axis_style_signature('mapper', [('copy', True),

~/anaconda3/lib/python3.7/site-packages/pandas/core/generic.py in drop(self,
labels, axis, index, columns, level, inplace, errors)
   3778         for axis, labels in axes.items():
   3779             if labels is not None:
-> 3780                 obj = obj._drop_axis(labels, axis, level=level, error
s=errors)
   3781
   3782         if inplace:

~/anaconda3/lib/python3.7/site-packages/pandas/core/generic.py in _drop_axis
(self, labels, axis, level, errors)
   3810                 new_axis = axis.drop(labels, level=level, errors=erro
rs)
   3811             else:
-> 3812                 new_axis = axis.drop(labels, errors=errors)
   3813             result = self.reindex(**{axis_name: new_axis})
   3814

~/anaconda3/lib/python3.7/site-packages/pandas/core/indexes/base.py in drop(s
elf, labels, errors)
   4963             if errors != 'ignore':
   4964                 raise KeyError(
-> 4965                     '{} not found in axis'.format(labels[mask]))
   4966             indexer = indexer[~mask]
   4967         return self.delete(indexer)

KeyError: "['id'] not found in axis"
```

In [36]:
```
psql.to_sql(product_category, 'product_category', connection)
```

```
---------------------------------------------------------------------------
ProgrammingError                          Traceback (most recent call last)
~/anaconda3/lib/python3.7/site-packages/pandas/io/sql.py in execute(self, *ar
gs, **kwargs)
   1430             else:
-> 1431                 cur.execute(*args)
   1432             return cur

ProgrammingError: relation "sqlite_master" does not exist
LINE 1: SELECT name FROM sqlite_master WHERE type='table' AND name=?...
                         ^


During handling of the above exception, another exception occurred:

DatabaseError                             Traceback (most recent call last)
<ipython-input-36-a111587276f6> in <module>
----> 1 psql.to_sql(product_category, 'product_category', connection)
      2 product_category

~/anaconda3/lib/python3.7/site-packages/pandas/io/sql.py in to_sql(frame, nam
e, con, schema, if_exists, index, index_label, chunksize, dtype, method)
    458     pandas_sql.to_sql(frame, name, if_exists=if_exists, index=index,
    459                       index_label=index_label, schema=schema,
--> 460                       chunksize=chunksize, dtype=dtype, method=metho
d)
    461
    462

~/anaconda3/lib/python3.7/site-packages/pandas/io/sql.py in to_sql(self, fram
e, name, if_exists, index, index_label, schema, chunksize, dtype, method)
   1544                                 if_exists=if_exists, index_label=index_la
bel,
   1545                                 dtype=dtype)
-> 1546         table.create()
   1547         table.insert(chunksize, method)
   1548

~/anaconda3/lib/python3.7/site-packages/pandas/io/sql.py in create(self)
    570
    571     def create(self):
--> 572         if self.exists():
    573             if self.if_exists == 'fail':
    574                 raise ValueError(

~/anaconda3/lib/python3.7/site-packages/pandas/io/sql.py in exists(self)
    558
    559     def exists(self):
--> 560         return self.pd_sql.has_table(self.name, self.schema)
    561
    562     def sql_schema(self):

~/anaconda3/lib/python3.7/site-packages/pandas/io/sql.py in has_table(self, n
ame, schema)
   1556                  "WHERE type='table' AND name={wld};").format(wld=wl
d)
   1557
-> 1558         return len(self.execute(query, [name, ]).fetchall()) > 0
   1559
   1560     def get_table(self, table_name, schema=None):

~/anaconda3/lib/python3.7/site-packages/pandas/io/sql.py in execute(self, *ar
gs, **kwargs)
   1443                 "Execution failed on sql '{sql}': {exc}".format(
   1444                     sql=args[0], exc=exc))
-> 1445             raise_with_traceback(ex)
   1446
   1447     @staticmethod
```