Handling CSV File

In [1]:
```
import pandas as pd
```

In [3]:
```
# Dataset uploaded from files
df = pd.read_csv("Machine Learning/restaurant_orders.csv")
df
```

Out[3]:

	Order ID	Customer Name	Food Item	Category	Quantity	Price	Payment Method	Order Time
0	2268	Mary Vega DDS	Pasta	Main	5	16.52	Cash	2025-02-02 14:28:41
1	3082	Brandon Myers	Brownie	Dessert	4	17.27	Debit Card	2025-06-08 10:57:47
2	3160	Margaret Wells	Pasta	Main	1	3.37	Credit Card	2025-03-04 07:41:41
3	1272	Michael Matthews	Pasta	Main	5	2.20	Online Payment	2025-05-15 12:43:45
4	9447	Connor Williams	Soup	Starter	1	12.23	Cash	2025-03-15 14:25:56
...	...	...	...	...	...	...	...	...
495	6323	Alyssa Anthony	Pizza	Main	1	21.31	Cash	2025-01-15 19:21:02
496	9836	Jerry Pineda	Soup	Starter	3	15.99	Debit Card	2025-07-15 15:00:19
497	1202	Brandy Smith	Pasta	Main	2	8.54	Credit Card	2025-08-03 23:47:28
498	7876	Ivan Haynes	Soup	Starter	5	20.54	Credit Card	2025-07-23 08:10:06
499	1509	Amber Mendez	Ice Cream	Dessert	2	18.86	Online Payment	2025-08-09 05:11:27

500 rows × 8 columns

In [4]:
```
# Dataset uploaded from open url
url = "https://people.sc.fsu.edu/~jburkardt/data/csv/hw_200.csv"
```

```
df = pd.read_csv(url)
df
```

Out[4]:

	Index	Height(Inches)"	"Weight(Pounds)"
0	1	65.78	112.99
1	2	71.52	136.49
2	3	69.40	153.03
3	4	68.22	142.34
4	5	67.79	144.30
...	...	...	...
195	196	65.80	120.84
196	197	66.11	115.78
197	198	68.24	128.30
198	199	68.02	127.47
199	200	71.39	127.88

200 rows × 3 columns

In [5]:
```
# Dataset uploaded from URL
df = pd.read_csv("Machine Learning/Medical cost dataset.txt", sep='\t')
df
```

Out[5]:

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
...	...	...	...	...	...	...	...
1333	50	male	30.970	3	no	northwest	10600.54830
1334	18	female	31.920	0	no	northeast	2205.98080
1335	18	female	36.850	0	no	southeast	1629.83350
1336	21	female	25.800	0	no	southwest	2007.94500
1337	61	female	29.070	0	yes	northwest	29141.36030

1338 rows × 7 columns

In [6]:
```
# To change the name of the headings
df = pd.read_csv("Machine Learning/Medical cost dataset.txt", sep='\t', names=['
df
```

Out[6]:

	S.No	Age	Sex	BMI	Children	Smoker	Region	Charges
0	age	sex	bmi	children	smoker	region	charges	NaN
1	19	female	27.9	0	yes	southwest	16884.924	NaN
2	18	male	33.77	1	no	southeast	1725.5523	NaN
3	28	male	33	3	no	southeast	4449.462	NaN
4	33	male	22.705	0	no	northwest	21984.47061	NaN
...	...	...	...	...	...	...	...	...
1334	50	male	30.97	3	no	northwest	10600.5483	NaN
1335	18	female	31.92	0	no	northeast	2205.9808	NaN
1336	18	female	36.85	0	no	southeast	1629.8335	NaN
1337	21	female	25.8	0	no	southwest	2007.945	NaN
1338	61	female	29.07	0	yes	northwest	29141.3603	NaN

1339 rows × 8 columns

In [8]:
```
#Index column
df = pd.read_csv("Machine Learning/Medical cost dataset.txt", sep='\t',index_col
df
```

Out[8]:

age	sex	bmi	children	smoker	region	charges
19	female	27.900	0	yes	southwest	16884.92400
18	male	33.770	1	no	southeast	1725.55230
28	male	33.000	3	no	southeast	4449.46200
33	male	22.705	0	no	northwest	21984.47061
32	male	28.880	0	no	northwest	3866.85520
...	...	...	...	...	...	...
50	male	30.970	3	no	northwest	10600.54830
18	female	31.920	0	no	northeast	2205.98080
18	female	36.850	0	no	southeast	1629.83350
21	female	25.800	0	no	southwest	2007.94500
61	female	29.070	0	yes	northwest	29141.36030

1338 rows × 6 columns

In [10]:
```
# Use Column
df = pd.read_csv("Machine Learning/Medical cost dataset.txt", sep='\t',usecols=[
df
```

Out[10]:

	age	bmi	charges
0	19	27.900	16884.92400
1	18	33.770	1725.55230
2	28	33.000	4449.46200
3	33	22.705	21984.47061
4	32	28.880	3866.85520
...	...	...	...
1333	50	30.970	10600.54830
1334	18	31.920	2205.98080
1335	18	36.850	1629.83350
1336	21	25.800	2007.94500
1337	61	29.070	29141.36030

1338 rows × 3 columns

In [16]:
```
# Skip rows
df = pd.read_csv("Machine Learning/Medical cost dataset.txt", sep='\t',skiprows=
df
```

Out[16]:

	18	male	33.77	1	no	southeast	1725.5523
0	28	male	33.000	3	no	southeast	4449.46200
1	33	male	22.705	0	no	northwest	21984.47061
2	32	male	28.880	0	no	northwest	3866.85520
3	31	female	25.740	0	no	southeast	3756.62160
4	46	female	33.440	1	no	southeast	8240.58960
...	...	...	...	...	...	...	...
1331	50	male	30.970	3	no	northwest	10600.54830
1332	18	female	31.920	0	no	northeast	2205.98080
1333	18	female	36.850	0	no	southeast	1629.83350
1334	21	female	25.800	0	no	southwest	2007.94500
1335	61	female	29.070	0	yes	northwest	29141.36030

1336 rows × 7 columns

In [17]:
```
# n rows
df = pd.read_csv("Machine Learning/Medical cost dataset.txt", sep='\t',nrows=50)
df
```

Out[17]:

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
5	31	female	25.740	0	no	southeast	3756.62160
6	46	female	33.440	1	no	southeast	8240.58960
7	37	female	27.740	3	no	northwest	7281.50560
8	37	male	29.830	2	no	northeast	6406.41070
9	60	female	25.840	0	no	northwest	28923.13692
10	25	male	26.220	0	no	northeast	2721.32080
11	62	female	26.290	0	yes	southeast	27808.72510
12	23	male	34.400	0	no	southwest	1826.84300
13	56	female	39.820	0	no	southeast	11090.71780
14	27	male	42.130	0	yes	southeast	39611.75770
15	19	male	24.600	1	no	southwest	1837.23700
16	52	female	30.780	1	no	northeast	10797.33620
17	23	male	23.845	0	no	northeast	2395.17155
18	56	male	40.300	0	no	southwest	10602.38500
19	30	male	35.300	0	yes	southwest	36837.46700
20	60	female	36.005	0	no	northeast	13228.84695
21	30	female	32.400	1	no	southwest	4149.73600
22	18	male	34.100	0	no	southeast	1137.01100
23	34	female	31.920	1	yes	northeast	37701.87680
24	37	male	28.025	2	no	northwest	6203.90175
25	59	female	27.720	3	no	southeast	14001.13380
26	63	female	23.085	0	no	northeast	14451.83515
27	55	female	32.775	2	no	northwest	12268.63225
28	23	male	17.385	1	no	northwest	2775.19215
29	31	male	36.300	2	yes	southwest	38711.00000
30	22	male	35.600	0	yes	southwest	35585.57600
31	18	female	26.315	0	no	northeast	2198.18985
32	19	female	28.600	5	no	southwest	4687.79700

	age	sex	bmi	children	smoker	region	charges
33	63	male	28.310	0	no	northwest	13770.09790
34	28	male	36.400	1	yes	southwest	51194.55914
35	19	male	20.425	0	no	northwest	1625.43375
36	62	female	32.965	3	no	northwest	15612.19335
37	26	male	20.800	0	no	southwest	2302.30000
38	35	male	36.670	1	yes	northeast	39774.27630
39	60	male	39.900	0	yes	southwest	48173.36100
40	24	female	26.600	0	no	northeast	3046.06200
41	31	female	36.630	2	no	southeast	4949.75870
42	41	male	21.780	1	no	southeast	6272.47720
43	37	female	30.800	2	no	southeast	6313.75900
44	38	male	37.050	1	no	northeast	6079.67150
45	55	male	37.300	0	no	southwest	20630.28351
46	18	female	38.665	2	no	northeast	3393.35635
47	28	female	34.770	0	no	northwest	3556.92230
48	60	female	24.530	0	no	southeast	12629.89670
49	36	male	35.200	1	yes	southeast	38709.17600

In [18]:
```
# Handling dataset
df = pd.read_csv("Machine Learning/Medical cost dataset.txt", sep='\t').info()
df
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1338 entries, 0 to 1337
Data columns (total 7 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   age       1338 non-null   int64  
 1   sex       1338 non-null   object 
 2   bmi       1338 non-null   float64
 3   children  1338 non-null   int64  
 4   smoker    1338 non-null   object 
 5   region    1338 non-null   object 
 6   charges   1338 non-null   float64
dtypes: float64(2), int64(2), object(3)
memory usage: 73.3+ KB
```