

DATA CLEANING

```
In [2]: import pandas as pd
import numpy as np
df = pd.DataFrame(data={
    'Temperature': [25, np.nan, 34, 20, 15, 19, 18, np.nan, 25],
    'Humidity': [20, 20, 12, np.nan, 13, 15, np.nan, np.nan, 20],
    'Situation': ['Humid', np.nan, 'Too-Sunny', 'cool', 'Too Cool', 'Normal', 'Normal', np.nan, 'Humid']
})
```

```
In [3]: df
```

```
Out[3]:
```

	Temperature	Humidity	Situation
0	25.0	20.0	Humid
1	NaN	20.0	NaN
2	34.0	12.0	Too-Sunny
3	20.0	NaN	cool
4	15.0	13.0	Too Cool
5	19.0	15.0	Normal
6	18.0	NaN	Normal
7	NaN	NaN	NaN
8	25.0	20.0	Humid

```
In [4]: df.isna()
```

```
Out[4]:
```

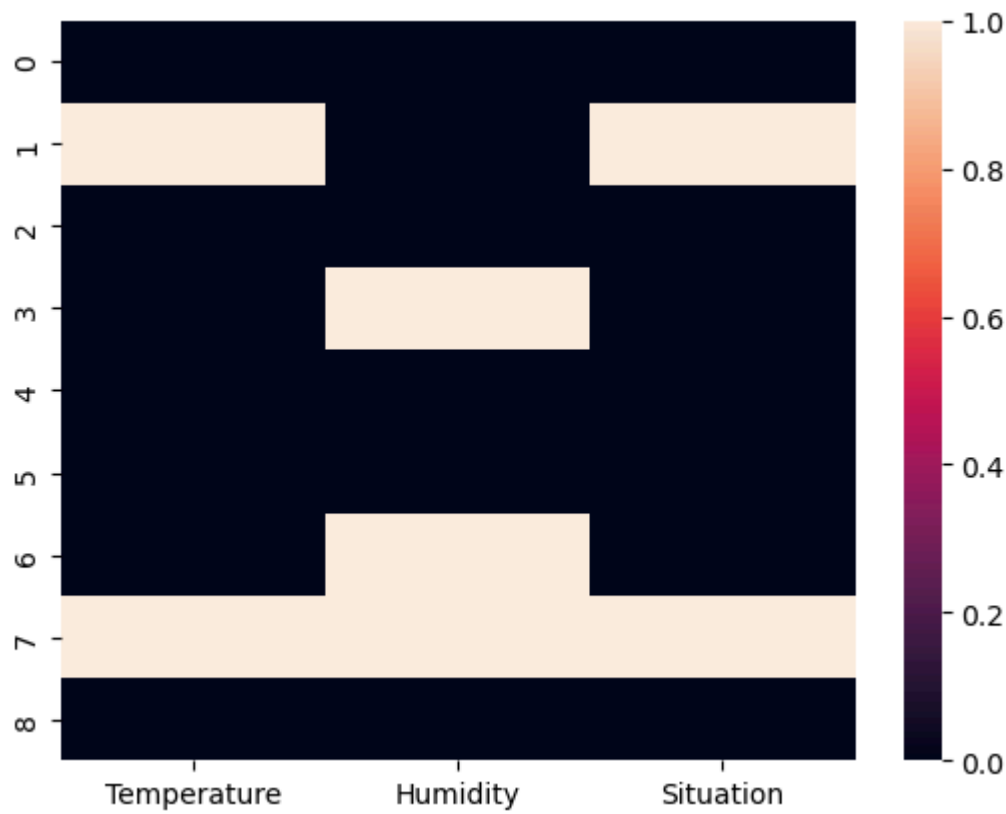
	Temperature	Humidity	Situation
0	False	False	False
1	True	False	True
2	False	False	False
3	False	True	False
4	False	False	False
5	False	False	False
6	False	True	False
7	True	True	True
8	False	False	False

```
In [5]: df.isna().sum()
```

```
Out[5]: Temperature    2
Humidity              3
Situation             2
dtype: int64
```

```
In [6]: import seaborn as sns
sns.heatmap(df.isnull())
```

Out[6]: <Axes: >



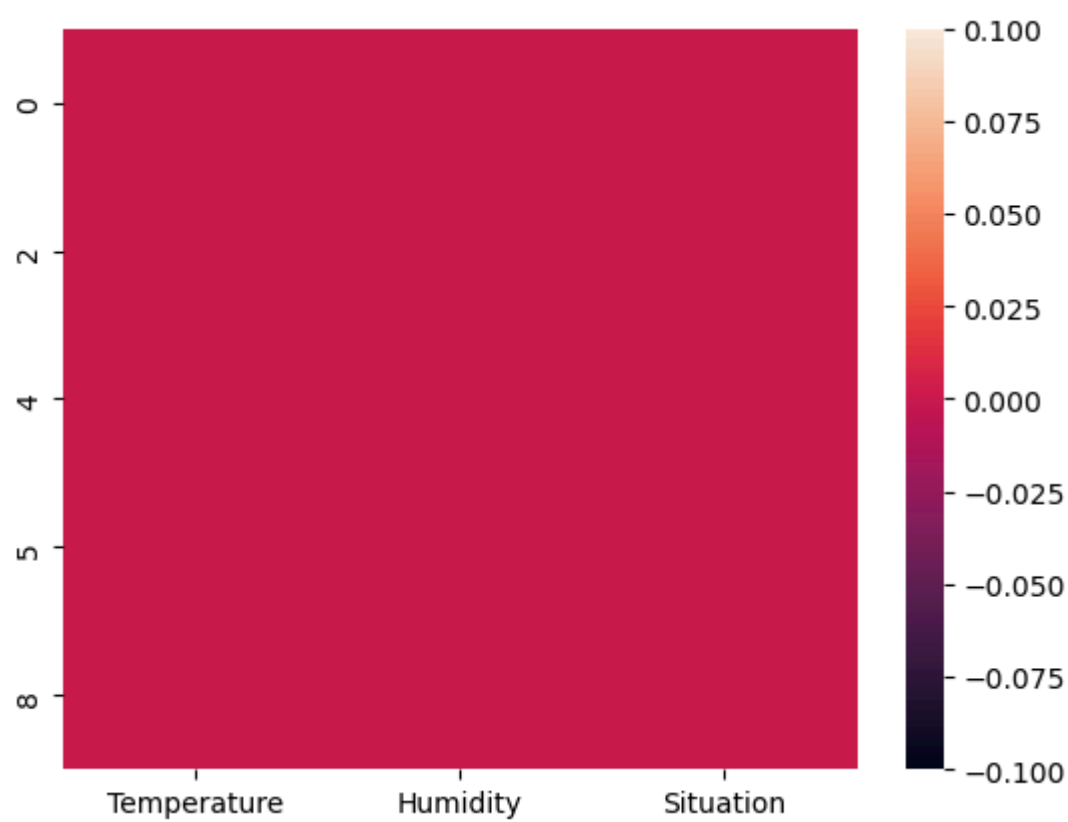
```
In [7]: ndf = df.dropna()  
ndf
```

Out[7]:

	Temperature	Humidity	Situation
0	25.0	20.0	Humid
2	34.0	12.0	Too-Sunny
4	15.0	13.0	Too Cool
5	19.0	15.0	Normal
8	25.0	20.0	Humid

```
In [8]: sns.heatmap(ndf.isnull())
```

Out[8]: <Axes: >



```
In [9]: fuldf = df.dropna(how="all")
```

```
In [10]: fuldf
```

```
Out[10]:
```

	Temperature	Humidity	Situation
0	25.0	20.0	Humid
1	NaN	20.0	NaN
2	34.0	12.0	Too-Sunny
3	20.0	NaN	cool
4	15.0	13.0	Too Cool
5	19.0	15.0	Normal
6	18.0	NaN	Normal
8	25.0	20.0	Humid

```
In [11]: df
```

Out[11]:

	Temperature	Humidity	Situation
0	25.0	20.0	Humid
1	NaN	20.0	NaN
2	34.0	12.0	Too-Sunny
3	20.0	NaN	cool
4	15.0	13.0	Too Cool
5	19.0	15.0	Normal
6	18.0	NaN	Normal
7	NaN	NaN	NaN
8	25.0	20.0	Humid

In [12]:

```
ffill = fuldf.fillna(method="ffill")
```

C:\Users\bhanu\AppData\Local\Temp\ipykernel_5160\3103576806.py:1: FutureWarning: DataFrame.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() instead.

```
ffill = fuldf.fillna(method="ffill")
```

In [13]:

```
ffill
```

Out[13]:

	Temperature	Humidity	Situation
0	25.0	20.0	Humid
1	25.0	20.0	Humid
2	34.0	12.0	Too-Sunny
3	20.0	12.0	cool
4	15.0	13.0	Too Cool
5	19.0	15.0	Normal
6	18.0	15.0	Normal
8	25.0	20.0	Humid

In [15]:

```
fuldf
```

Out[15]:

	Temperature	Humidity	Situation
0	25.0	20.0	Humid
1	NaN	20.0	NaN
2	34.0	12.0	Too-Sunny
3	20.0	NaN	cool
4	15.0	13.0	Too Cool
5	19.0	15.0	Normal
6	18.0	NaN	Normal
8	25.0	20.0	Humid

```
In [16]: meanv = fuldf["Humidity"].mean()
```

```
In [17]: meanv
```

```
Out[17]: 16.666666666666668
```

```
In [18]: fuldf["Humidity"].fillna(meanv)
```

```
Out[18]: 0    20.000000
1    20.000000
2    12.000000
3    16.666667
4    13.000000
5    15.000000
6    16.666667
8    20.000000
Name: Humidity, dtype: float64
```

```
In [19]: fuldf
```

```
Out[19]:
```

	Temperature	Humidity	Situation
0	25.0	20.0	Humid
1	NaN	20.0	NaN
2	34.0	12.0	Too-Sunny
3	20.0	NaN	cool
4	15.0	13.0	Too Cool
5	19.0	15.0	Normal
6	18.0	NaN	Normal
8	25.0	20.0	Humid

```
In [20]: fuldf["Humidity"].fillna(meanv, inplace=True)
```

C:\Users\bhanu\AppData\Local\Temp\ipykernel_5160\3854768861.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
fuldf["Humidity"].fillna(meanv, inplace=True)
```

```
In [21]: fuldf
```

```
Out[21]:
```

	Temperature	Humidity	Situation
0	25.0	20.000000	Humid
1	NaN	20.000000	NaN
2	34.0	12.000000	Too-Sunny
3	20.0	16.666667	cool
4	15.0	13.000000	Too Cool
5	19.0	15.000000	Normal
6	18.0	16.666667	Normal
8	25.0	20.000000	Humid

```
In [22]: fuldf.drop_duplicates()
```

Out[22]:

	Temperature	Humidity	Situation
0	25.0	20.000000	Humid
1	NaN	20.000000	NaN
2	34.0	12.000000	Too-Sunny
3	20.0	16.666667	cool
4	15.0	13.000000	Too Cool
5	19.0	15.000000	Normal
6	18.0	16.666667	Normal

```
In [23]: dset = pd.read_csv("dc_dataset.csv")
```

```
In [24]: dset
```

Out[24]:

	ID	No_Pation	Gender	AGE	Urea	Cr	HbA1c	Chol	TG	HDL	LDL	VLDL	BMI	CLASS
0	502	17975	F	50.0	4.7	46.0	4.9	4.2	0.9	2.4	1.4	0.5	24.0	N
1	735	34221	M	26.0	4.5	62.0	4.9	3.7	1.4	1.1	2.1	0.6	23.0	N
2	420	47975	F	50.0	4.7	46.0	4.9	4.2	0.9	2.4	1.4	0.5	24.0	N
3	680	87656	F	50.0	4.7	46.0	4.9	4.2	0.9	2.4	1.4	0.5	24.0	N
4	504	34223	M	33.0	7.1	46.0	4.9	4.9	1.0	0.8	2.0	0.4	21.0	N
...
1004	191	454316	M	55.0	NaN	62.0	6.8	5.3	2.0	1.0	3.5	0.9	30.1	Y
1005	192	454316	M	55.0	4.8	88.0	NaN	5.7	4.0	0.9	3.3	1.8	30.0	Y
1006	193	454316	M	62.0	6.3	82.0	6.7	5.3	2.0	1.0	3.5	NaN	30.1	Y
1007	194	454316	F	57.0	4.1	70.0	9.3	5.3	3.3	1.0	1.4	1.3	29.0	Y
1008	195	4543	f	55.0	4.1	34.0	13.9	5.4	1.6	1.6	3.1	0.7	33.0	Y

1009 rows × 14 columns

```
In [25]: dset.isna().sum()
```

Out[25]:

ID	0
No_Pation	0
Gender	0
AGE	1
Urea	1
Cr	2
HbA1c	3
Chol	2
TG	2
HDL	1
LDL	2
VLDL	1
BMI	0
CLASS	0
dtype:	int64

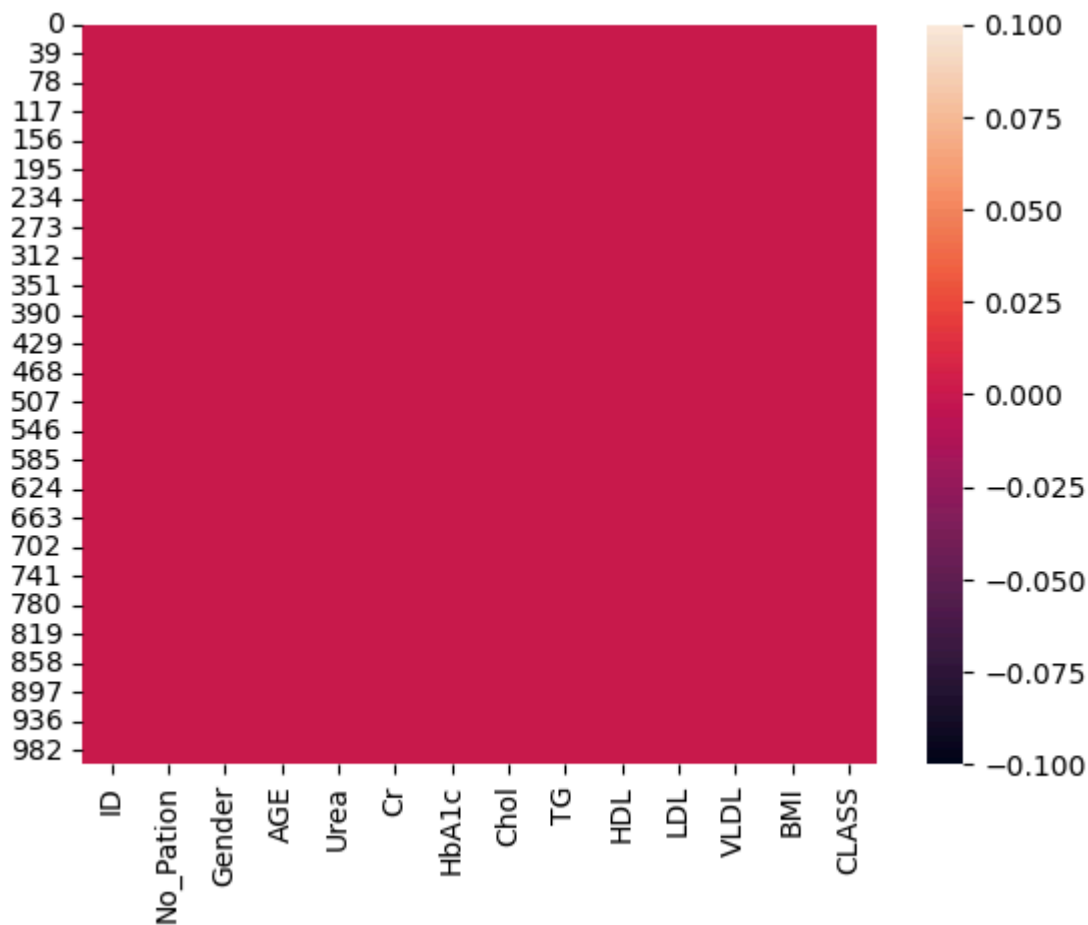
```
In [26]: ddset = dset.dropna()
```

```
In [27]: ddset.isna().sum()
```

```
Out[27]: ID          0
         No_Pation   0
         Gender      0
         AGE         0
         Urea        0
         Cr          0
         HbA1c       0
         Chol        0
         TG          0
         HDL         0
         LDL         0
         VLDL        0
         BMI         0
         CLASS       0
         dtype: int64
```

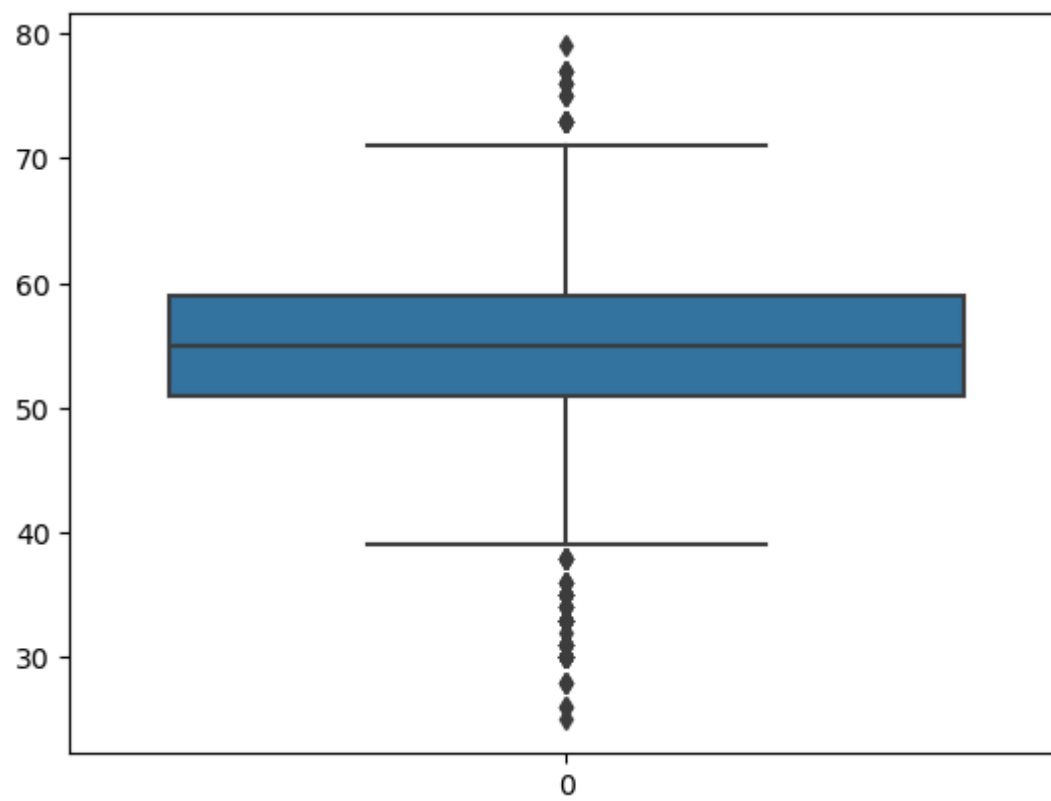
```
In [28]: sns.heatmap(ddset.isnull())
```

```
Out[28]: <Axes: >
```



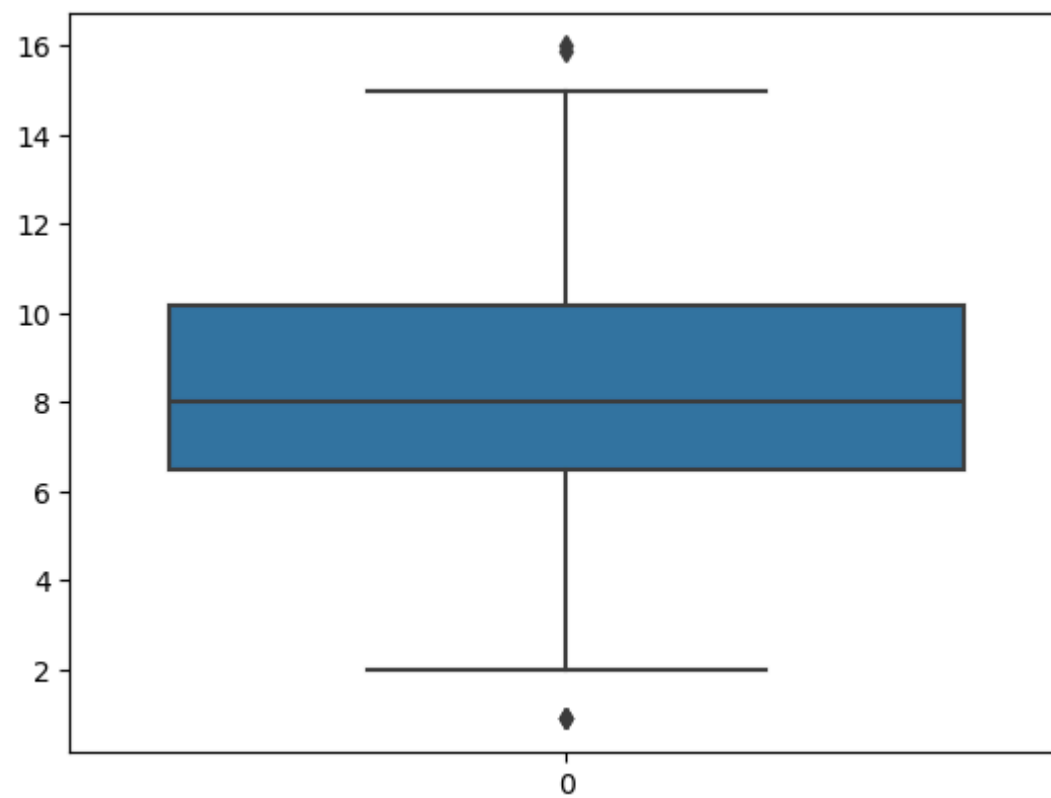
```
In [29]: sns.boxplot(ddset['AGE'])
```

```
Out[29]: <Axes: >
```



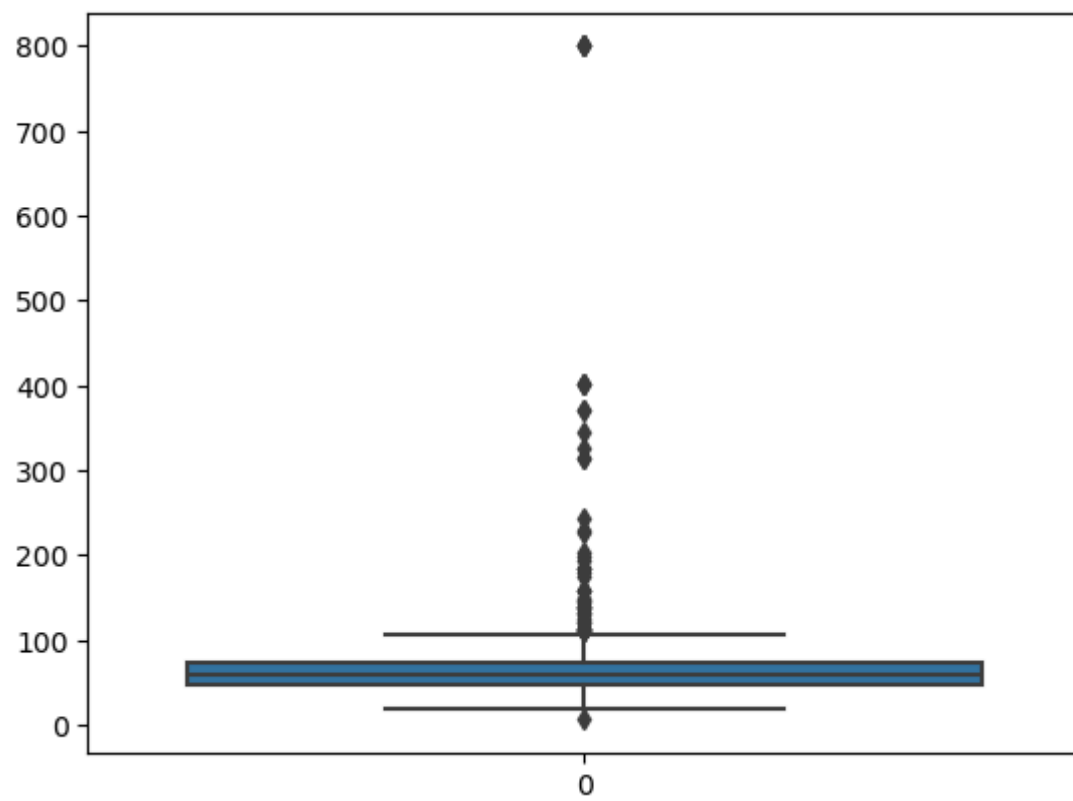
```
In [30]: sns.boxplot(ddset['HbA1c'])
```

```
Out[30]: <Axes: >
```



```
In [31]: sns.boxplot(ddset['Cr'])
```

```
Out[31]: <Axes: >
```

DATA MANIPULATION

```
In [32]: hless = pd.read_csv("home1s.csv")
```

```
In [33]: hless
```

Out[33]:

	region	state	individuals	family_members	state_pop
0	East South Central	Alabama	2570	864	4887681
1	Pacific	Alaska	1434	582	735139
2	Mountain	Arizona	7259	2606	7158024
3	West South Central	Arkansas	2280	432	3009733
4	Pacific	California	109008	20964	39461588
5	Mountain	Colorado	7607	3250	5691287
6	New England	Connecticut	2280	1696	3571520
7	South Atlantic	Delaware	708	374	965479
8	South Atlantic	District of Columbia	3770	3134	701547
9	South Atlantic	Florida	21443	9587	21244317
10	South Atlantic	Georgia	6943	2556	10511131
11	Pacific	Hawaii	4131	2399	1420593
12	Mountain	Idaho	1297	715	1750536
13	East North Central	Illinois	6752	3891	12723071
14	East North Central	Indiana	3776	1482	6695497
15	West North Central	Iowa	1711	1038	3148618
16	West North Central	Kansas	1443	773	2911359
17	East South Central	Kentucky	2735	953	4461153
18	West South Central	Louisiana	2540	519	4659690
19	New England	Maine	1450	1066	1339057
20	South Atlantic	Maryland	4914	2230	6035802
21	New England	Massachusetts	6811	13257	6882635
22	East North Central	Michigan	5209	3142	9984072
23	West North Central	Minnesota	3993	3250	5606249
24	East South Central	Mississippi	1024	328	2981020
25	West North Central	Missouri	3776	2107	6121623
26	Mountain	Montana	983	422	1060665
27	West North Central	Nebraska	1745	676	1925614
28	Mountain	Nevada	7058	486	3027341
29	New England	New Hampshire	835	615	1353465
30	Mid-Atlantic	New Jersey	6048	3350	8886025
31	Mountain	New Mexico	1949	602	2092741
32	Mid-Atlantic	New York	39827	52070	19530351
33	South Atlantic	North Carolina	6451	2817	10381615
34	West North Central	North Dakota	467	75	758080
35	East North Central	Ohio	6929	3320	11676341

	region	state	individuals	family_members	state_pop
36	West South Central	Oklahoma	2823	1048	3940235
37	Pacific	Oregon	11139	3337	4181886
38	Mid-Atlantic	Pennsylvania	8163	5349	12800922
39	New England	Rhode Island	747	354	1058287
40	South Atlantic	South Carolina	3082	851	5084156
41	West North Central	South Dakota	836	323	878698
42	East South Central	Tennessee	6139	1744	6771631
43	West South Central	Texas	19199	6111	28628666
44	Mountain	Utah	1904	972	3153550
45	New England	Vermont	780	511	624358
46	South Atlantic	Virginia	3928	2047	8501286
47	Pacific	Washington	16424	5880	7523869
48	South Atlantic	West Virginia	1021	222	1804291
49	East North Central	Wisconsin	2740	2167	5807406
50	Mountain	Wyoming	434	205	577601

In [34]: `hless.shape`

Out[34]: (51, 5)

In [35]: `hless.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  -
0   region          51 non-null    object
1   state           51 non-null    object
2   individuals     51 non-null    int64
3   family_members  51 non-null    int64
4   state_pop       51 non-null    int64
dtypes: int64(3), object(2)
memory usage: 2.1+ KB
```

In [36]: `nhless = hless.convert_dtypes()`

In [37]: `nhless.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  -
0   region          51 non-null    string
1   state           51 non-null    string
2   individuals     51 non-null    Int64
3   family_members  51 non-null    Int64
4   state_pop       51 non-null    Int64
dtypes: Int64(3), string(2)
memory usage: 2.3 KB
```

```
In [38]: nhless.rename(columns = {'state_pop': 'state_population'})
```

Out[38]:

	region	state	individuals	family_members	state_population
0	East South Central	Alabama	2570	864	4887681
1	Pacific	Alaska	1434	582	735139
2	Mountain	Arizona	7259	2606	7158024
3	West South Central	Arkansas	2280	432	3009733
4	Pacific	California	109008	20964	39461588
5	Mountain	Colorado	7607	3250	5691287
6	New England	Connecticut	2280	1696	3571520
7	South Atlantic	Delaware	708	374	965479
8	South Atlantic	District of Columbia	3770	3134	701547
9	South Atlantic	Florida	21443	9587	21244317
10	South Atlantic	Georgia	6943	2556	10511131
11	Pacific	Hawaii	4131	2399	1420593
12	Mountain	Idaho	1297	715	1750536
13	East North Central	Illinois	6752	3891	12723071
14	East North Central	Indiana	3776	1482	6695497
15	West North Central	Iowa	1711	1038	3148618
16	West North Central	Kansas	1443	773	2911359
17	East South Central	Kentucky	2735	953	4461153
18	West South Central	Louisiana	2540	519	4659690
19	New England	Maine	1450	1066	1339057
20	South Atlantic	Maryland	4914	2230	6035802
21	New England	Massachusetts	6811	13257	6882635
22	East North Central	Michigan	5209	3142	9984072
23	West North Central	Minnesota	3993	3250	5606249
24	East South Central	Mississippi	1024	328	2981020
25	West North Central	Missouri	3776	2107	6121623
26	Mountain	Montana	983	422	1060665
27	West North Central	Nebraska	1745	676	1925614
28	Mountain	Nevada	7058	486	3027341
29	New England	New Hampshire	835	615	1353465
30	Mid-Atlantic	New Jersey	6048	3350	8886025
31	Mountain	New Mexico	1949	602	2092741
32	Mid-Atlantic	New York	39827	52070	19530351
33	South Atlantic	North Carolina	6451	2817	10381615
34	West North Central	North Dakota	467	75	758080
35	East North Central	Ohio	6929	3320	11676341

	region	state	individuals	family_members	state_population
36	West South Central	Oklahoma	2823	1048	3940235
37	Pacific	Oregon	11139	3337	4181886
38	Mid-Atlantic	Pennsylvania	8163	5349	12800922
39	New England	Rhode Island	747	354	1058287
40	South Atlantic	South Carolina	3082	851	5084156
41	West North Central	South Dakota	836	323	878698
42	East South Central	Tennessee	6139	1744	6771631
43	West South Central	Texas	19199	6111	28628666
44	Mountain	Utah	1904	972	3153550
45	New England	Vermont	780	511	624358
46	South Atlantic	Virginia	3928	2047	8501286
47	Pacific	Washington	16424	5880	7523869
48	South Atlantic	West Virginia	1021	222	1804291
49	East North Central	Wisconsin	2740	2167	5807406
50	Mountain	Wyoming	434	205	577601

In [39]:

nhless

Out[39]:

	region	state	individuals	family_members	state_pop
0	East South Central	Alabama	2570	864	4887681
1	Pacific	Alaska	1434	582	735139
2	Mountain	Arizona	7259	2606	7158024
3	West South Central	Arkansas	2280	432	3009733
4	Pacific	California	109008	20964	39461588
5	Mountain	Colorado	7607	3250	5691287
6	New England	Connecticut	2280	1696	3571520
7	South Atlantic	Delaware	708	374	965479
8	South Atlantic	District of Columbia	3770	3134	701547
9	South Atlantic	Florida	21443	9587	21244317
10	South Atlantic	Georgia	6943	2556	10511131
11	Pacific	Hawaii	4131	2399	1420593
12	Mountain	Idaho	1297	715	1750536
13	East North Central	Illinois	6752	3891	12723071
14	East North Central	Indiana	3776	1482	6695497
15	West North Central	Iowa	1711	1038	3148618
16	West North Central	Kansas	1443	773	2911359
17	East South Central	Kentucky	2735	953	4461153
18	West South Central	Louisiana	2540	519	4659690
19	New England	Maine	1450	1066	1339057
20	South Atlantic	Maryland	4914	2230	6035802
21	New England	Massachusetts	6811	13257	6882635
22	East North Central	Michigan	5209	3142	9984072
23	West North Central	Minnesota	3993	3250	5606249
24	East South Central	Mississippi	1024	328	2981020
25	West North Central	Missouri	3776	2107	6121623
26	Mountain	Montana	983	422	1060665
27	West North Central	Nebraska	1745	676	1925614
28	Mountain	Nevada	7058	486	3027341
29	New England	New Hampshire	835	615	1353465
30	Mid-Atlantic	New Jersey	6048	3350	8886025
31	Mountain	New Mexico	1949	602	2092741
32	Mid-Atlantic	New York	39827	52070	19530351
33	South Atlantic	North Carolina	6451	2817	10381615
34	West North Central	North Dakota	467	75	758080
35	East North Central	Ohio	6929	3320	11676341

	region	state	individuals	family_members	state_pop
36	West South Central	Oklahoma	2823	1048	3940235
37	Pacific	Oregon	11139	3337	4181886
38	Mid-Atlantic	Pennsylvania	8163	5349	12800922
39	New England	Rhode Island	747	354	1058287
40	South Atlantic	South Carolina	3082	851	5084156
41	West North Central	South Dakota	836	323	878698
42	East South Central	Tennessee	6139	1744	6771631
43	West South Central	Texas	19199	6111	28628666
44	Mountain	Utah	1904	972	3153550
45	New England	Vermont	780	511	624358
46	South Atlantic	Virginia	3928	2047	8501286
47	Pacific	Washington	16424	5880	7523869
48	South Atlantic	West Virginia	1021	222	1804291
49	East North Central	Wisconsin	2740	2167	5807406
50	Mountain	Wyoming	434	205	577601

In [40]: `nhless.rename(columns = {'state_pop':'state_population'}, inplace=True)`

In [41]: `nhless`

Out[41]:

	region	state	individuals	family_members	state_population
0	East South Central	Alabama	2570	864	4887681
1	Pacific	Alaska	1434	582	735139
2	Mountain	Arizona	7259	2606	7158024
3	West South Central	Arkansas	2280	432	3009733
4	Pacific	California	109008	20964	39461588
5	Mountain	Colorado	7607	3250	5691287
6	New England	Connecticut	2280	1696	3571520
7	South Atlantic	Delaware	708	374	965479
8	South Atlantic	District of Columbia	3770	3134	701547
9	South Atlantic	Florida	21443	9587	21244317
10	South Atlantic	Georgia	6943	2556	10511131
11	Pacific	Hawaii	4131	2399	1420593
12	Mountain	Idaho	1297	715	1750536
13	East North Central	Illinois	6752	3891	12723071
14	East North Central	Indiana	3776	1482	6695497
15	West North Central	Iowa	1711	1038	3148618
16	West North Central	Kansas	1443	773	2911359
17	East South Central	Kentucky	2735	953	4461153
18	West South Central	Louisiana	2540	519	4659690
19	New England	Maine	1450	1066	1339057
20	South Atlantic	Maryland	4914	2230	6035802
21	New England	Massachusetts	6811	13257	6882635
22	East North Central	Michigan	5209	3142	9984072
23	West North Central	Minnesota	3993	3250	5606249
24	East South Central	Mississippi	1024	328	2981020
25	West North Central	Missouri	3776	2107	6121623
26	Mountain	Montana	983	422	1060665
27	West North Central	Nebraska	1745	676	1925614
28	Mountain	Nevada	7058	486	3027341
29	New England	New Hampshire	835	615	1353465
30	Mid-Atlantic	New Jersey	6048	3350	8886025
31	Mountain	New Mexico	1949	602	2092741
32	Mid-Atlantic	New York	39827	52070	19530351
33	South Atlantic	North Carolina	6451	2817	10381615
34	West North Central	North Dakota	467	75	758080
35	East North Central	Ohio	6929	3320	11676341

	region	state	individuals	family_members	state_population
36	West South Central	Oklahoma	2823	1048	3940235
37	Pacific	Oregon	11139	3337	4181886
38	Mid-Atlantic	Pennsylvania	8163	5349	12800922
39	New England	Rhode Island	747	354	1058287
40	South Atlantic	South Carolina	3082	851	5084156
41	West North Central	South Dakota	836	323	878698
42	East South Central	Tennessee	6139	1744	6771631
43	West South Central	Texas	19199	6111	28628666
44	Mountain	Utah	1904	972	3153550
45	New England	Vermont	780	511	624358
46	South Atlantic	Virginia	3928	2047	8501286
47	Pacific	Washington	16424	5880	7523869
48	South Atlantic	West Virginia	1021	222	1804291
49	East North Central	Wisconsin	2740	2167	5807406
50	Mountain	Wyoming	434	205	577601

In [42]:

nhless.sort_values("region")

Out[42]:

	region	state	individuals	family_members	state_population
35	East North Central	Ohio	6929	3320	11676341
49	East North Central	Wisconsin	2740	2167	5807406
22	East North Central	Michigan	5209	3142	9984072
13	East North Central	Illinois	6752	3891	12723071
14	East North Central	Indiana	3776	1482	6695497
0	East South Central	Alabama	2570	864	4887681
42	East South Central	Tennessee	6139	1744	6771631
24	East South Central	Mississippi	1024	328	2981020
17	East South Central	Kentucky	2735	953	4461153
38	Mid-Atlantic	Pennsylvania	8163	5349	12800922
32	Mid-Atlantic	New York	39827	52070	19530351
30	Mid-Atlantic	New Jersey	6048	3350	8886025
44	Mountain	Utah	1904	972	3153550
31	Mountain	New Mexico	1949	602	2092741
28	Mountain	Nevada	7058	486	3027341
26	Mountain	Montana	983	422	1060665
50	Mountain	Wyoming	434	205	577601
2	Mountain	Arizona	7259	2606	7158024
12	Mountain	Idaho	1297	715	1750536
5	Mountain	Colorado	7607	3250	5691287
19	New England	Maine	1450	1066	1339057
21	New England	Massachusetts	6811	13257	6882635
45	New England	Vermont	780	511	624358
39	New England	Rhode Island	747	354	1058287
29	New England	New Hampshire	835	615	1353465
6	New England	Connecticut	2280	1696	3571520
37	Pacific	Oregon	11139	3337	4181886
47	Pacific	Washington	16424	5880	7523869
11	Pacific	Hawaii	4131	2399	1420593
4	Pacific	California	109008	20964	39461588
1	Pacific	Alaska	1434	582	735139
33	South Atlantic	North Carolina	6451	2817	10381615
8	South Atlantic	District of Columbia	3770	3134	701547
46	South Atlantic	Virginia	3928	2047	8501286
9	South Atlantic	Florida	21443	9587	21244317
10	South Atlantic	Georgia	6943	2556	10511131

	region	state	individuals	family_members	state_population
20	South Atlantic	Maryland	4914	2230	6035802
48	South Atlantic	West Virginia	1021	222	1804291
7	South Atlantic	Delaware	708	374	965479
40	South Atlantic	South Carolina	3082	851	5084156
41	West North Central	South Dakota	836	323	878698
15	West North Central	Iowa	1711	1038	3148618
34	West North Central	North Dakota	467	75	758080
16	West North Central	Kansas	1443	773	2911359
27	West North Central	Nebraska	1745	676	1925614
23	West North Central	Minnesota	3993	3250	5606249
25	West North Central	Missouri	3776	2107	6121623
36	West South Central	Oklahoma	2823	1048	3940235
43	West South Central	Texas	19199	6111	28628666
18	West South Central	Louisiana	2540	519	4659690
3	West South Central	Arkansas	2280	432	3009733

```
In [43]: nhless.sort_values("region", ascending=False)
```

Out[43]:

	region	state	individuals	family_members	state_population
36	West South Central	Oklahoma	2823	1048	3940235
3	West South Central	Arkansas	2280	432	3009733
18	West South Central	Louisiana	2540	519	4659690
43	West South Central	Texas	19199	6111	28628666
25	West North Central	Missouri	3776	2107	6121623
15	West North Central	Iowa	1711	1038	3148618
27	West North Central	Nebraska	1745	676	1925614
23	West North Central	Minnesota	3993	3250	5606249
41	West North Central	South Dakota	836	323	878698
16	West North Central	Kansas	1443	773	2911359
34	West North Central	North Dakota	467	75	758080
8	South Atlantic	District of Columbia	3770	3134	701547
48	South Atlantic	West Virginia	1021	222	1804291
40	South Atlantic	South Carolina	3082	851	5084156
20	South Atlantic	Maryland	4914	2230	6035802
7	South Atlantic	Delaware	708	374	965479
10	South Atlantic	Georgia	6943	2556	10511131
33	South Atlantic	North Carolina	6451	2817	10381615
9	South Atlantic	Florida	21443	9587	21244317
46	South Atlantic	Virginia	3928	2047	8501286
11	Pacific	Hawaii	4131	2399	1420593
47	Pacific	Washington	16424	5880	7523869
1	Pacific	Alaska	1434	582	735139
4	Pacific	California	109008	20964	39461588
37	Pacific	Oregon	11139	3337	4181886
19	New England	Maine	1450	1066	1339057
6	New England	Connecticut	2280	1696	3571520
45	New England	Vermont	780	511	624358
39	New England	Rhode Island	747	354	1058287
29	New England	New Hampshire	835	615	1353465
21	New England	Massachusetts	6811	13257	6882635
44	Mountain	Utah	1904	972	3153550
50	Mountain	Wyoming	434	205	577601
2	Mountain	Arizona	7259	2606	7158024
28	Mountain	Nevada	7058	486	3027341
26	Mountain	Montana	983	422	1060665

	region	state	individuals	family_members	state_population
5	Mountain	Colorado	7607	3250	5691287
12	Mountain	Idaho	1297	715	1750536
31	Mountain	New Mexico	1949	602	2092741
32	Mid-Atlantic	New York	39827	52070	19530351
30	Mid-Atlantic	New Jersey	6048	3350	8886025
38	Mid-Atlantic	Pennsylvania	8163	5349	12800922
17	East South Central	Kentucky	2735	953	4461153
42	East South Central	Tennessee	6139	1744	6771631
24	East South Central	Mississippi	1024	328	2981020
0	East South Central	Alabama	2570	864	4887681
14	East North Central	Indiana	3776	1482	6695497
13	East North Central	Illinois	6752	3891	12723071
22	East North Central	Michigan	5209	3142	9984072
35	East North Central	Ohio	6929	3320	11676341
49	East North Central	Wisconsin	2740	2167	5807406

In [44]:

nhless["region"]

```
Out[44]: 0      East South Central
1          Pacific
2          Mountain
3      West South Central
4          Pacific
5          Mountain
6          New England
7          South Atlantic
8          South Atlantic
9          South Atlantic
10         South Atlantic
11         Pacific
12         Mountain
13     East North Central
14     East North Central
15     West North Central
16     West North Central
17     East South Central
18     West South Central
19         New England
20         South Atlantic
21         New England
22     East North Central
23     West North Central
24     East South Central
25     West North Central
26         Mountain
27     West North Central
28         Mountain
29         New England
30         Mid-Atlantic
31         Mountain
32         Mid-Atlantic
33         South Atlantic
34     West North Central
35     East North Central
36     West South Central
37         Pacific
38         Mid-Atlantic
39         New England
40         South Atlantic
41     West North Central
42     East South Central
43     West South Central
44         Mountain
45         New England
46         South Atlantic
47         Pacific
48         South Atlantic
49     East North Central
50         Mountain
Name: region, dtype: string
```

```
In [45]: nhless.region
```

```

Out[45]: 0      East South Central
         1          Pacific
         2          Mountain
         3      West South Central
         4          Pacific
         5          Mountain
         6          New England
         7          South Atlantic
         8          South Atlantic
         9          South Atlantic
        10          South Atlantic
        11          Pacific
        12          Mountain
        13      East North Central
        14      East North Central
        15      West North Central
        16      West North Central
        17      East South Central
        18      West South Central
        19          New England
        20          South Atlantic
        21          New England
        22      East North Central
        23      West North Central
        24      East South Central
        25      West North Central
        26          Mountain
        27      West North Central
        28          Mountain
        29          New England
        30          Mid-Atlantic
        31          Mountain
        32          Mid-Atlantic
        33          South Atlantic
        34      West North Central
        35      East North Central
        36      West South Central
        37          Pacific
        38          Mid-Atlantic
        39          New England
        40          South Atlantic
        41      West North Central
        42      East South Central
        43      West South Central
        44          Mountain
        45          New England
        46          South Atlantic
        47          Pacific
        48          South Atlantic
        49      East North Central
        50          Mountain
        Name: region, dtype: string

```

```

In [47]: nhless[["region","state"]]

```


Out[47]:

	region	state
0	East South Central	Alabama
1	Pacific	Alaska
2	Mountain	Arizona
3	West South Central	Arkansas
4	Pacific	California
5	Mountain	Colorado
6	New England	Connecticut
7	South Atlantic	Delaware
8	South Atlantic	District of Columbia
9	South Atlantic	Florida
10	South Atlantic	Georgia
11	Pacific	Hawaii
12	Mountain	Idaho
13	East North Central	Illinois
14	East North Central	Indiana
15	West North Central	Iowa
16	West North Central	Kansas
17	East South Central	Kentucky
18	West South Central	Louisiana
19	New England	Maine
20	South Atlantic	Maryland
21	New England	Massachusetts
22	East North Central	Michigan
23	West North Central	Minnesota
24	East South Central	Mississippi
25	West North Central	Missouri
26	Mountain	Montana
27	West North Central	Nebraska
28	Mountain	Nevada
29	New England	New Hampshire
30	Mid-Atlantic	New Jersey
31	Mountain	New Mexico
32	Mid-Atlantic	New York
33	South Atlantic	North Carolina
34	West North Central	North Dakota
35	East North Central	Ohio

	region	state
36	West South Central	Oklahoma
37	Pacific	Oregon
38	Mid-Atlantic	Pennsylvania
39	New England	Rhode Island
40	South Atlantic	South Carolina
41	West North Central	South Dakota
42	East South Central	Tennessee
43	West South Central	Texas
44	Mountain	Utah
45	New England	Vermont
46	South Atlantic	Virginia
47	Pacific	Washington
48	South Atlantic	West Virginia
49	East North Central	Wisconsin
50	Mountain	Wyoming

In [49]: `nhless["family_members"] > 5000`

```
Out[49]: 0      False
         1      False
         2      False
         3      False
         4       True
         5      False
         6      False
         7      False
         8      False
         9       True
        10      False
        11      False
        12      False
        13      False
        14      False
        15      False
        16      False
        17      False
        18      False
        19      False
        20      False
        21       True
        22      False
        23      False
        24      False
        25      False
        26      False
        27      False
        28      False
        29      False
        30      False
        31      False
        32       True
        33      False
        34      False
        35      False
        36      False
        37      False
        38       True
        39      False
        40      False
        41      False
        42      False
        43       True
        44      False
        45      False
        46      False
        47       True
        48      False
        49      False
        50      False
Name: family_members, dtype: boolean
```

```
In [50]: nhless[["family_members"] > 5000]
```

```
-----
TypeError                                Traceback (most recent call last)
Cell In[50], line 1
----> 1 nhless[["family_members"] > 5000]

TypeError: '>' not supported between instances of 'list' and 'int'
```

```
In [51]: nhless[nhless["family_members"] > 5000]
```

Out[51]:

	region	state	individuals	family_members	state_population
4	Pacific	California	109008	20964	39461588
9	South Atlantic	Florida	21443	9587	21244317
21	New England	Massachusetts	6811	13257	6882635
32	Mid-Atlantic	New York	39827	52070	19530351
38	Mid-Atlantic	Pennsylvania	8163	5349	12800922
43	West South Central	Texas	19199	6111	28628666
47	Pacific	Washington	16424	5880	7523869

In [52]:

```
nhless[nhless.family_members > 5000]
```

Out[52]:

	region	state	individuals	family_members	state_population
4	Pacific	California	109008	20964	39461588
9	South Atlantic	Florida	21443	9587	21244317
21	New England	Massachusetts	6811	13257	6882635
32	Mid-Atlantic	New York	39827	52070	19530351
38	Mid-Atlantic	Pennsylvania	8163	5349	12800922
43	West South Central	Texas	19199	6111	28628666
47	Pacific	Washington	16424	5880	7523869

In [53]:

```
indexhless = nhless.set_index(["region"])
```

In [54]:

```
indexhless
```

Out[54]:

	state	individuals	family_members	state_population
region				
East South Central	Alabama	2570	864	4887681
Pacific	Alaska	1434	582	735139
Mountain	Arizona	7259	2606	7158024
West South Central	Arkansas	2280	432	3009733
Pacific	California	109008	20964	39461588
Mountain	Colorado	7607	3250	5691287
New England	Connecticut	2280	1696	3571520
South Atlantic	Delaware	708	374	965479
South Atlantic	District of Columbia	3770	3134	701547
South Atlantic	Florida	21443	9587	21244317
South Atlantic	Georgia	6943	2556	10511131
Pacific	Hawaii	4131	2399	1420593
Mountain	Idaho	1297	715	1750536
East North Central	Illinois	6752	3891	12723071
East North Central	Indiana	3776	1482	6695497
West North Central	Iowa	1711	1038	3148618
West North Central	Kansas	1443	773	2911359
East South Central	Kentucky	2735	953	4461153
West South Central	Louisiana	2540	519	4659690
New England	Maine	1450	1066	1339057
South Atlantic	Maryland	4914	2230	6035802
New England	Massachusetts	6811	13257	6882635
East North Central	Michigan	5209	3142	9984072
West North Central	Minnesota	3993	3250	5606249
East South Central	Mississippi	1024	328	2981020
West North Central	Missouri	3776	2107	6121623
Mountain	Montana	983	422	1060665
West North Central	Nebraska	1745	676	1925614
Mountain	Nevada	7058	486	3027341
New England	New Hampshire	835	615	1353465
Mid-Atlantic	New Jersey	6048	3350	8886025
Mountain	New Mexico	1949	602	2092741
Mid-Atlantic	New York	39827	52070	19530351
South Atlantic	North Carolina	6451	2817	10381615
West North Central	North Dakota	467	75	758080

	state	individuals	family_members	state_population
region				
East North Central	Ohio	6929	3320	11676341
West South Central	Oklahoma	2823	1048	3940235
Pacific	Oregon	11139	3337	4181886
Mid-Atlantic	Pennsylvania	8163	5349	12800922
New England	Rhode Island	747	354	1058287
South Atlantic	South Carolina	3082	851	5084156
West North Central	South Dakota	836	323	878698
East South Central	Tennessee	6139	1744	6771631
West South Central	Texas	19199	6111	28628666
Mountain	Utah	1904	972	3153550
New England	Vermont	780	511	624358
South Atlantic	Virginia	3928	2047	8501286
Pacific	Washington	16424	5880	7523869
South Atlantic	West Virginia	1021	222	1804291
East North Central	Wisconsin	2740	2167	5807406
Mountain	Wyoming	434	205	577601

In []: