In [4]:
```
import numpy as np
import pandas as pd
import seaborn as sns
```

In [5]:
```
df=pd.read_csv("data.csv")
```

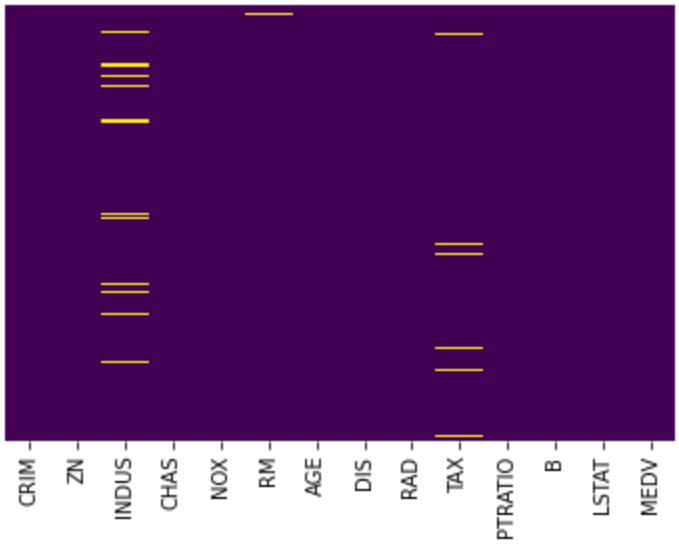
In [6]:
```
df
```

Out[6]:

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MEDV
0	0.00632	18.0	2.31	0	0.538	6.575	65.2	4.0900	1	296	15.3	396.90	4.98	24.0
1	0.02731	0.0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.90	9.14	21.6
2	0.02729	0.0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.03	34.7
3	0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	33.4
4	0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.90	5.33	36.2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
506	0.98765	0.0	12.50	0	0.561	6.980	89.0	2.0980	3	320	23.0	396.00	12.00	12.0
507	0.23456	0.0	12.50	0	0.561	6.980	76.0	2.6540	3	320	23.0	343.00	25.00	32.0
508	0.44433	0.0	12.50	0	0.561	6.123	98.0	2.9870	3	320	23.0	343.00	21.00	54.0
509	0.77763	0.0	12.70	0	0.561	6.222	34.0	2.5430	3	329	23.0	343.00	76.00	67.0
510	0.65432	0.0	12.80	0	0.561	6.760	67.0	2.9870	3	345	23.0	321.00	45.00	24.0

511 rows × 14 columns

In [15]:
```
sns.heatmap(df.isnull(),yticklabels=False,cbar=False,cmap='viridis');
```



In [11]:
```
i1 = np.random.choice(a=df.index, size=35)
i2 = np.random.choice(a=df.index, size=20)
```

In [12]:
```
df.loc[i1, 'INDUS'] = np.nan
df.loc[i2, 'TAX'] = np.nan
```

In [16]:
```
df.isna().sum()
```

Out[16]:
```
CRIM        0
```

```
ZN          0
INDUS      33
CHAS        0
NOX         0
RM          5
AGE         0
DIS         0
RAD         0
TAX        20
PTRATIO     0
B           0
LSTAT       0
MEDV        0
dtype: int64
```

In [20]:
```
from sklearn.impute import KNNImputer
imputer = KNNImputer(n_neighbors=3)
imputed=imputer.fit_transform(df)
df_imputed = pd.DataFrame(imputed, columns=df.columns)
```

In [19]:
```
df
```

Out[19]:

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MEDV
0	0.00632	18.0	2.31	0	0.538	6.575	65.2	4.0900	1	296.0	15.3	396.90	4.98	24.0
1	0.02731	0.0	NaN	0	0.469	6.421	78.9	4.9671	2	242.0	17.8	396.90	9.14	21.6
2	0.02729	0.0	7.07	0	0.469	7.185	61.1	4.9671	2	NaN	17.8	392.83	4.03	34.7
3	0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222.0	18.7	394.63	2.94	33.4
4	0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222.0	18.7	396.90	5.33	36.2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
506	0.98765	0.0	12.50	0	0.561	6.980	89.0	2.0980	3	NaN	23.0	396.00	12.00	12.0
507	0.23456	0.0	12.50	0	0.561	6.980	76.0	2.6540	3	320.0	23.0	343.00	25.00	32.0
508	0.44433	0.0	12.50	0	0.561	6.123	98.0	2.9870	3	320.0	23.0	343.00	21.00	54.0
509	0.77763	0.0	12.70	0	0.561	6.222	34.0	2.5430	3	329.0	23.0	343.00	76.00	67.0
510	0.65432	0.0	12.80	0	0.561	6.760	67.0	2.9870	3	345.0	23.0	321.00	45.00	24.0

511 rows × 14 columns

In [21]:
```
df_imputed
```

Out[21]:

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MED
0	0.00632	18.0	2.310000	0.0	0.538	6.575	65.2	4.0900	1.0	296.000000	15.3	396.90	4.98	24
1	0.02731	0.0	9.806667	0.0	0.469	6.421	78.9	4.9671	2.0	242.000000	17.8	396.90	9.14	21
2	0.02729	0.0	7.070000	0.0	0.469	7.185	61.1	4.9671	2.0	246.333333	17.8	392.83	4.03	34
3	0.03237	0.0	2.180000	0.0	0.458	6.998	45.8	6.0622	3.0	222.000000	18.7	394.63	2.94	33
4	0.06905	0.0	2.180000	0.0	0.458	7.147	54.2	6.0622	3.0	222.000000	18.7	396.90	5.33	36
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
506	0.98765	0.0	12.500000	0.0	0.561	6.980	89.0	2.0980	3.0	331.666667	23.0	396.00	12.00	12
507	0.23456	0.0	12.500000	0.0	0.561	6.980	76.0	2.6540	3.0	320.000000	23.0	343.00	25.00	32

508	0.44433	0.0	12.500000	0.0	0.561	6.123	98.0	2.9870	3.0	320.000000	23.0	343.00	21.00	54
509	0.77763	0.0	12.700000	0.0	0.561	6.222	34.0	2.5430	3.0	329.000000	23.0	343.00	76.00	67
510	0.65432	0.0	12.800000	0.0	0.561	6.760	67.0	2.9870	3.0	345.000000	23.0	321.00	45.00	24

511 rows × 14 columns

In [22]:
```
df_imputed.isna().sum()
```

Out[22]:
```
CRIM       0
ZN         0
INDUS      0
CHAS       0
NOX        0
RM         0
AGE        0
DIS        0
RAD        0
TAX        0
PTRATIO    0
B          0
LSTAT      0
MEDV       0
dtype: int64
```

In [ ]:
```

```