

```
In [12]: import pandas as pd

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/wine/wine.data'
wine = pd.read_csv(url)

wine
```

Out[12]:

		1	14.23	1.71	2.43	15.6	127	2.8	3.06	.28	2.29	5.64	1.04	3.92	1065
0	1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28	4.38	1.05	3.40	1050	
1	1	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30	2.81	5.68	1.03	3.17	1185	
2	1	14.37	1.95	2.50	16.8	113	3.85	3.49	0.24	2.18	7.80	0.86	3.45	1480	
3	1	13.24	2.59	2.87	21.0	118	2.80	2.69	0.39	1.82	4.32	1.04	2.93	735	
4	1	14.20	1.76	2.45	15.2	112	3.27	3.39	0.34	1.97	6.75	1.05	2.85	1450	
...
172	3	13.71	5.65	2.45	20.5	95	1.68	0.61	0.52	1.06	7.70	0.64	1.74	740	
173	3	13.40	3.91	2.48	23.0	102	1.80	0.75	0.43	1.41	7.30	0.70	1.56	750	
174	3	13.27	4.28	2.26	20.0	120	1.59	0.69	0.43	1.35	10.20	0.59	1.56	835	
175	3	13.17	2.59	2.37	20.0	120	1.65	0.68	0.53	1.46	9.30	0.60	1.62	840	
176	3	14.13	4.10	2.74	24.5	96	2.05	0.76	0.56	1.35	9.20	0.61	1.60	560	

177 rows × 14 columns

Delete the first, fourth, seventh, ninth, eleventh, thirteenth and fourteenth columns.

```
In [13]: #first column = 0, second column = 1, etc

wine = wine.drop(wine.columns[[0,3,6,8,10,12,13]], axis=1)

wine
```

Out[13]:

		14.23	1.71	15.6	127	3.06	2.29	1.04
0	13.20	1.78	11.2	100	2.76	1.28	1.05	
1	13.16	2.36	18.6	101	3.24	2.81	1.03	
2	14.37	1.95	16.8	113	3.49	2.18	0.86	
3	13.24	2.59	21.0	118	2.69	1.82	1.04	
4	14.20	1.76	15.2	112	3.39	1.97	1.05	
...
172	13.71	5.65	20.5	95	0.61	1.06	0.64	
173	13.40	3.91	23.0	102	0.75	1.41	0.70	
174	13.27	4.28	20.0	120	0.69	1.35	0.59	
175	13.17	2.59	20.0	120	0.68	1.46	0.60	
176	14.13	4.10	24.5	96	0.76	1.35	0.61	

177 rows × 7 columns

Assign the columns as below:

- alcohol
- malic_acid
- alcalinity_of_ash
- magnesium
- flavanoids
- proanthocyanins
- hue

```
In [20]: wine = wine.rename({'14.23' : 'alcohol',
                             '1.71' : 'malic_acid',
                             '15.6' : 'alcalinity_of_ash',
                             '127' : 'magnesium',
                             '3.06' : 'flavanoids',
                             '2.29' : 'proanthocyanins',
                             '1.04' : 'hue'}, axis=1)

wine
```

Out[20]:

	alcohol	malic_acid	alcalinity_of_ash	magnesium	flavanoids	proanthocyanins	hue
0	10.00	1.78	11.2	100.0	2.76	1.28	1.05
1	10.00	2.36	18.6	101.0	3.24	2.81	1.03
2	10.00	1.95	16.8	113.0	3.49	2.18	0.86
3	13.24	2.59	21.0	10.0	2.69	1.82	1.04
4	14.20	1.76	15.2	10.0	3.39	1.97	1.05
...
172	13.71	5.65	20.5	95.0	0.61	1.06	0.64
173	13.40	3.91	23.0	102.0	0.75	1.41	0.70
174	13.27	4.28	20.0	120.0	0.69	1.35	0.59
175	13.17	2.59	20.0	120.0	0.68	1.46	0.60
176	14.13	4.10	24.5	96.0	0.76	1.35	0.61

177 rows × 7 columns

Set the values of the first 3 rows in the alcohol column as NaN

```
In [15]: import numpy as np

wine.loc[:2, 'alcohol'] = np.nan
wine
```

Out[15]:

	alcohol	malic_acid	alcalinity_of_ash	magnesium	flavanoids	proanthocyanins	hue
0	NaN	1.78	11.2	100	2.76	1.28	1.05
1	NaN	2.36	18.6	101	3.24	2.81	1.03
2	NaN	1.95	16.8	113	3.49	2.18	0.86
3	13.24	2.59	21.0	118	2.69	1.82	1.04
4	14.20	1.76	15.2	112	3.39	1.97	1.05
...
172	13.71	5.65	20.5	95	0.61	1.06	0.64
173	13.40	3.91	23.0	102	0.75	1.41	0.70
174	13.27	4.28	20.0	120	0.69	1.35	0.59
175	13.17	2.59	20.0	120	0.68	1.46	0.60
176	14.13	4.10	24.5	96	0.76	1.35	0.61

177 rows × 7 columns

Now set the value of the rows 3 and 4 of the magnesium column as NaN

```
In [16]: wine.loc[3:4, 'magnesium'] = np.nan
wine
```

Out[16]:

	alcohol	malic_acid	alcalinity_of_ash	magnesium	flavanoids	proanthocyanins	hue
0	NaN	1.78	11.2	100.0	2.76	1.28	1.05
1	NaN	2.36	18.6	101.0	3.24	2.81	1.03
2	NaN	1.95	16.8	113.0	3.49	2.18	0.86
3	13.24	2.59	21.0	NaN	2.69	1.82	1.04
4	14.20	1.76	15.2	NaN	3.39	1.97	1.05
...
172	13.71	5.65	20.5	95.0	0.61	1.06	0.64
173	13.40	3.91	23.0	102.0	0.75	1.41	0.70
174	13.27	4.28	20.0	120.0	0.69	1.35	0.59
175	13.17	2.59	20.0	120.0	0.68	1.46	0.60
176	14.13	4.10	24.5	96.0	0.76	1.35	0.61

177 rows × 7 columns

Fill in the null values (NaN) with the number 10 in the alcohol column and 100 in magnesium column.

```
In [17]: wine['alcohol'] = wine['alcohol'].fillna(10)
wine['magnesium'] = wine['magnesium'].fillna(10)

wine
```

Out[17]:

	alcohol	malic_acid	alcalinity_of_ash	magnesium	flavanoids	proanthocyanins	hue
0	10.00	1.78	11.2	100.0	2.76	1.28	1.05
1	10.00	2.36	18.6	101.0	3.24	2.81	1.03
2	10.00	1.95	16.8	113.0	3.49	2.18	0.86
3	13.24	2.59	21.0	10.0	2.69	1.82	1.04
4	14.20	1.76	15.2	10.0	3.39	1.97	1.05
...
172	13.71	5.65	20.5	95.0	0.61	1.06	0.64
173	13.40	3.91	23.0	102.0	0.75	1.41	0.70
174	13.27	4.28	20.0	120.0	0.69	1.35	0.59
175	13.17	2.59	20.0	120.0	0.68	1.46	0.60
176	14.13	4.10	24.5	96.0	0.76	1.35	0.61

177 rows × 7 columns

Count the number of missing values in the entire dataset.

```
In [18]: #if we put a single function sum() returns us the null values of each
#column but asks us for the null values of the entire dataset

wine.isnull().sum().sum()
```

Out[18]:

```
In [ ]: install nbconvert[webpdf]
```

```
In [ ]:
```