

```

import pandas as pd
a=pd.read_csv("/content/breast_cancer_survival.csv")
print(a)
print(a.head())
print(a.columns)
target_variable = 'Patient_Status'
features = a.columns[a.columns != target_variable]
print('Target Variable:', target_variable)
print('Features:',features)

```

	Surgery_type	Date_of_Surgery	Date_of_Last_Visit	\
0	Other	20-May-18	26-Aug-18	
1	Other	26-Apr-18	25-Jan-19	
2	Lumpectomy	24-Aug-18	08-Apr-20	
3	Other	16-Nov-18	28-Jul-20	
4	Lumpectomy	12-Dec-18	05-Jan-19	
..	...	...	...	
329	Lumpectomy	15-Jan-19	27-Mar-20	
330	Modified Radical Mastectomy	25-Jul-18	23-Apr-19	
331	Simple Mastectomy	26-Mar-19	11-Oct-19	
332	Lumpectomy	26-Nov-18	05-Dec-18	
333	Modified Radical Mastectomy	04-Feb-19	10-Aug-19	

  

	Patient_Status
0	Alive
1	Dead
2	Alive
3	Alive
4	Alive
..	...
329	Alive

```
Index(['Age', 'Gender', 'Protein1', 'Protein2', 'Protein3', 'Protein4',  
      'Tumour_Stage', 'Histology', 'ER status', 'PR status', 'HER2 status',  
      'Surgery_type', 'Date_of_Surgery', 'Date_of_Last_Visit',  
      'Patient_Status'],  
      dtype='object')  
Target Variable: Patient_Status  
Features: Index(['Age', 'Gender', 'Protein1', 'Protein2', 'Protein3', 'Protein4',  
      'Tumour_Stage', 'Histology', 'ER status', 'PR status', 'HER2 status',  
      'Surgery_type', 'Date_of_Surgery', 'Date_of_Last_Visit'],  
      dtype='object')
```

```
b=a.fillna(0)
```

```
b
```

in2	Protein3	Protein4	Tumour_Stage	Histology	ER status	PR status	HER2 status	Surgery_type
000	0.007972	-0.048340	II	Infiltrating Ductal Carcinoma	Positive	Positive	Negative	Other
020	-0.498030	-0.507320	II	Infiltrating Ductal Carcinoma	Positive	Positive	Negative	Other
100	-0.370190	0.010815	II	Infiltrating Ductal Carcinoma	Positive	Positive	Negative	Lumpectomy
143	-0.370380	0.132190	I	Infiltrating Ductal Carcinoma	Positive	Positive	Negative	Other
110	-0.543970	-0.390210	II	Infiltrating Ductal Carcinoma	Positive	Positive	Positive	Lumpectomy
...	...	...	...	...	...	...	...	...
050	0.024751	0.280320	II	Infiltrating Ductal Carcinoma	Positive	Positive	Positive	Lumpectomy
147	0.472370	-0.523870	I	Infiltrating Ductal Carcinoma	Positive	Positive	Positive	Modified Radical Mastectomy
150	-0.332850	0.857860	II	Infiltrating Ductal Carcinoma	Positive	Positive	Negative	Simp Mastectomy
180	-0.366570	-0.107820	II	Infiltrating Lobular Carcinoma	Positive	Positive	Negative	Lumpectomy
180	0.318830	0.836050	II	Infiltrating Ductal Carcinoma	Positive	Positive	Negative	Modified Radical Mastectomy

Next steps:

 [View recommended plots](#)

```
y=b['Patient_Status']
print(y)

0    Alive
1     Dead
2    Alive
3    Alive
4    Alive
...
```

```

329     Alive
330     Alive
331     Dead
332     Alive
333     Dead
Name: Patient_Status, Length: 334, dtype: object

```

```

x=b.drop('Patient_Status',axis=1)
print(x)

```

	Age	Gender	Protein1	Protein2	Protein3	Protein4	Tumour_Stage	\
0	42	FEMALE	0.952560	2.15000	0.007972	-0.048340	II	
1	54	FEMALE	0.000000	1.38020	-0.498030	-0.507320	II	
2	63	FEMALE	-0.523030	1.76400	-0.370190	0.010815	II	
3	78	FEMALE	-0.876180	0.12943	-0.370380	0.132190	I	
4	42	FEMALE	0.226110	1.74910	-0.543970	-0.390210	II	
..	...	...	...	...	...	...	...	
329	59	FEMALE	0.024598	1.40050	0.024751	0.280320	II	
330	41	FEMALE	0.100120	-0.46547	0.472370	-0.523870	I	
331	54	FEMALE	0.753820	1.64250	-0.332850	0.857860	II	
332	74	FEMALE	0.972510	1.42680	-0.366570	-0.107820	II	
333	66	FEMALE	0.286380	1.39980	0.318830	0.836050	II	

	Histology	ER status	PR status	HER2 status	\
0	Infiltrating Ductal Carcinoma	Positive	Positive	Negative	
1	Infiltrating Ductal Carcinoma	Positive	Positive	Negative	
2	Infiltrating Ductal Carcinoma	Positive	Positive	Negative	
3	Infiltrating Ductal Carcinoma	Positive	Positive	Negative	
4	Infiltrating Ductal Carcinoma	Positive	Positive	Positive	
..	...	...	...	...	
329	Infiltrating Ductal Carcinoma	Positive	Positive	Positive	
330	Infiltrating Ductal Carcinoma	Positive	Positive	Positive	
331	Infiltrating Ductal Carcinoma	Positive	Positive	Negative	
332	Infiltrating Lobular Carcinoma	Positive	Positive	Negative	
333	Infiltrating Ductal Carcinoma	Positive	Positive	Negative	

	Surgery_type	Date_of_Surgery	Date_of_Last_Visit
0	Other	20-May-18	26-Aug-18
1	Other	26-Apr-18	25-Jan-19
2	Lumpectomy	24-Aug-18	08-Apr-20
3	Other	16-Nov-18	28-Jul-20
4	Lumpectomy	12-Dec-18	05-Jan-19
..	...	...	...
329	Lumpectomy	15-Jan-19	27-Mar-20
330	Modified Radical Mastectomy	25-Jul-18	23-Apr-19
331	Simple Mastectomy	26-Mar-19	11-Oct-19
332	Lumpectomy	26-Nov-18	05-Dec-18
333	Modified Radical Mastectomy	04-Feb-19	10-Aug-19

```
[334 rows x 14 columns]
```

```

x['Gender'] = x['Gender'].replace({'FEMALE': 1, 'MALE':0})
print(x)

```

	Age	Gender	Protein1	Protein2	Protein3	Protein4	Tumour_Stage	\
0	42	1	0.952560	2.15000	0.007972	-0.048340	II	
1	54	1	0.000000	1.38020	-0.498030	-0.507320	II	
2	63	1	-0.523030	1.76400	-0.370190	0.010815	II	

3	78	1	-0.876180	0.12943	-0.370380	0.132190	I
4	42	1	0.226110	1.74910	-0.543970	-0.390210	II
..	...	...	...	...	...	...	...
329	59	1	0.024598	1.40050	0.024751	0.280320	II
330	41	1	0.100120	-0.46547	0.472370	-0.523870	I
331	54	1	0.753820	1.64250	-0.332850	0.857860	II
332	74	1	0.972510	1.42680	-0.366570	-0.107820	II
333	66	1	0.286380	1.39980	0.318830	0.836050	II

			Histology	ER status	PR status	HER2 status	\
0	Infiltrating	Ductal	Carcinoma	Positive	Positive	Negative	
1	Infiltrating	Ductal	Carcinoma	Positive	Positive	Negative	
2	Infiltrating	Ductal	Carcinoma	Positive	Positive	Negative	
3	Infiltrating	Ductal	Carcinoma	Positive	Positive	Negative	
4	Infiltrating	Ductal	Carcinoma	Positive	Positive	Positive	
..			...	...	...	...	
329	Infiltrating	Ductal	Carcinoma	Positive	Positive	Positive	
330	Infiltrating	Ductal	Carcinoma	Positive	Positive	Positive	
331	Infiltrating	Ductal	Carcinoma	Positive	Positive	Negative	
332	Infiltrating	Lobular	Carcinoma	Positive	Positive	Negative	
333	Infiltrating	Ductal	Carcinoma	Positive	Positive	Negative	

		Surgery_type	Date_of_Surgery	Date_of_Last_Visit
0		Other	20-May-18	26-Aug-18
1		Other	26-Apr-18	25-Jan-19
2		Lumpectomy	24-Aug-18	08-Apr-20
3		Other	16-Nov-18	28-Jul-20
4		Lumpectomy	12-Dec-18	05-Jan-19
..		...	...	...
329		Lumpectomy	15-Jan-19	27-Mar-20
330	Modified Radical	Mastectomy	25-Jul-18	23-Apr-19
331	Simple	Mastectomy	26-Mar-19	11-Oct-19
332		Lumpectomy	26-Nov-18	05-Dec-18
333	Modified Radical	Mastectomy	04-Feb-19	10-Aug-19

[334 rows x 14 columns]

```
x['ER status'] = x['ER status'].replace({'Positive': 1, 'Negative' : 0})
x['HER2 status'] = x['HER2 status'].replace({'Positive': 1, 'Negative' : 0})
x['PR status'] = x['PR status'].replace({'Positive': 1, 'Negative' : 0})
print(x)
```

	Age	Gender	Protein1	Protein2	Protein3	Protein4	Tumour_Stage	\
0	42	1	0.952560	2.15000	0.007972	-0.048340	II	
1	54	1	0.000000	1.38020	-0.498030	-0.507320	II	
2	63	1	-0.523030	1.76400	-0.370190	0.010815	II	
3	78	1	-0.876180	0.12943	-0.370380	0.132190	I	
4	42	1	0.226110	1.74910	-0.543970	-0.390210	II	
..	...	...	...	...	...	...	...	
329	59	1	0.024598	1.40050	0.024751	0.280320	II	
330	41	1	0.100120	-0.46547	0.472370	-0.523870	I	
331	54	1	0.753820	1.64250	-0.332850	0.857860	II	
332	74	1	0.972510	1.42680	-0.366570	-0.107820	II	
333	66	1	0.286380	1.39980	0.318830	0.836050	II	

  

			Histology	ER status	PR status	HER2 status	\
0	Infiltrating	Ductal	Carcinoma	1	1	0	
1	Infiltrating	Ductal	Carcinoma	1	1	0	
2	Infiltrating	Ductal	Carcinoma	1	1	0	

3	Infiltrating Ductal Carcinoma	1	1	0
4	Infiltrating Ductal Carcinoma	1	1	1
..	...	...	...	...
329	Infiltrating Ductal Carcinoma	1	1	1
330	Infiltrating Ductal Carcinoma	1	1	1
331	Infiltrating Ductal Carcinoma	1	1	0
332	Infiltrating Lobular Carcinoma	1	1	0
333	Infiltrating Ductal Carcinoma	1	1	0

	Surgery_type	Date_of_Surgery	Date_of_Last_Visit
0	Other	20-May-18	26-Aug-18
1	Other	26-Apr-18	25-Jan-19
2	Lumpectomy	24-Aug-18	08-Apr-20
3	Other	16-Nov-18	28-Jul-20
4	Lumpectomy	12-Dec-18	05-Jan-19
..	...	...	...
329	Lumpectomy	15-Jan-19	27-Mar-20
330	Modified Radical Mastectomy	25-Jul-18	23-Apr-19
331	Simple Mastectomy	26-Mar-19	11-Oct-19
332	Lumpectomy	26-Nov-18	05-Dec-18
333	Modified Radical Mastectomy	04-Feb-19	10-Aug-19

[334 rows x 14 columns]

```
d4=x.drop('Histology',axis=1)
d1=d4.drop('Tumour_Stage',axis=1)
d2=d1.drop('Surgery_type',axis=1)
d3=d2.drop('Date_of_Surgery',axis=1)
d=d3.drop('Date_of_Last_Visit',axis=1)
print(d)
```

	Age	Gender	Protein1	Protein2	Protein3	Protein4	ER status	\
0	42	1	0.952560	2.15000	0.007972	-0.048340	1	
1	54	1	0.000000	1.38020	-0.498030	-0.507320	1	
2	63	1	-0.523030	1.76400	-0.370190	0.010815	1	
3	78	1	-0.876180	0.12943	-0.370380	0.132190	1	
4	42	1	0.226110	1.74910	-0.543970	-0.390210	1	
..	...	...	...	...	...	...	...	
329	59	1	0.024598	1.40050	0.024751	0.280320	1	
330	41	1	0.100120	-0.46547	0.472370	-0.523870	1	
331	54	1	0.753820	1.64250	-0.332850	0.857860	1	
332	74	1	0.972510	1.42680	-0.366570	-0.107820	1	
333	66	1	0.286380	1.39980	0.318830	0.836050	1	

	PR status	HER2 status
0	1	0
1	1	0
2	1	0
3	1	0
4	1	1
..	...	...
329	1	1
330	1	1
331	1	0
332	1	0
333	1	0

[334 rows x 9 columns]

```
y = y.replace({'Alive':1, 'Dead':0})  
print(y)
```

```
0      1  
1      0  
2      1  
3      1  
4      1
```

```
..  
329    1  
330    1  
331    0  
332    1  
333    0
```

```
Name: Patient_Status, Length: 334, dtype: int64
```

```
import numpy as np
```