

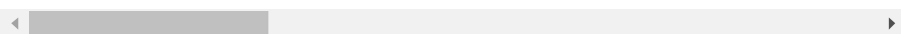
```
import pandas as pd
url='https://raw.githubusercontent.com/milaan9/93_Python_Data_Analytics_Projects/main/007_Breast_Cancer_Prediction_with_ML/data/data.csv'
df=pd.read_csv(url)
```

df



	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothne
0	842302	M	17.99	10.38	122.80	1001.0	
1	842517	M	20.57	17.77	132.90	1326.0	
2	84300903	M	19.69	21.25	130.00	1203.0	
3	84348301	M	11.42	20.38	77.58	386.1	
4	84358402	M	20.29	14.34	135.10	1297.0	
...	
564	926424	M	21.56	22.39	142.00	1479.0	
565	926682	M	20.13	28.25	131.20	1261.0	
566	926954	M	16.60	28.08	108.30	858.1	
567	927241	M	20.60	29.33	140.10	1265.0	
568	92751	B	7.76	24.54	47.92	181.0	

569 rows × 32 columns



df.head()

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactn
0	842302	M	17.99	10.38	122.80	1001.0	0.11840	
1	842517	M	20.57	17.77	132.90	1326.0	0.08474	
2	84300903	M	19.69	21.25	130.00	1203.0	0.10960	
3	84348301	M	11.42	20.38	77.58	386.1	0.14250	
4	84358402	M	20.29	14.34	135.10	1297.0	0.10030	

5 rows × 32 columns

df.head().columns

```
Index(['id', 'diagnosis', 'radius_mean', 'texture_mean', 'perimeter_mean',
       'area_mean', 'smoothness_mean', 'compactness_mean', 'concavity_mean',
       'concave points_mean', 'symmetry_mean', 'fractal_dimension_mean',
       'radius_se', 'texture_se', 'perimeter_se', 'area_se', 'smoothness_se',
       'compactness_se', 'concavity_se', 'concave points_se', 'symmetry_se',
       'fractal_dimension_se', 'radius_worst', 'texture_worst',
       'perimeter_worst', 'area_worst', 'smoothness_worst',
       'compactness_worst', 'concavity_worst', 'concave points_worst',
       'symmetry_worst', 'fractal_dimension_worst'],
      dtype='object')
```

```
x=df.iloc[:,2:]
x.head()
```

	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_m
0	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.3
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.0
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.1
3	11.42	20.38	77.58	386.1	0.14250	0.28300	0.2

```
y=df[['diagnosis']]
```

```
y.value_counts()
```

diagnosis	
B	357
M	212
dtype: int64	

```
y.sample(7)
```

diagnosis	
220	B
436	B
0	M
87	M
10	M
449	M
278	B

```
from sklearn.feature_selection import SelectKBest,SelectPercentile
from sklearn.feature_selection import chi2
x_kbest = SelectKBest(chi2, k=5).fit_transform(x, y)
percentile_model = SelectPercentile(chi2, percentile=20)
x_percentile = percentile_model.fit_transform(x, y)
```

```
x_kbest
```

array([[122.8 , 1001. , 153.4 , 184.6 , 2019.],
[132.9 , 1326. , 74.08, 158.8 , 1956.],
[130. , 1203. , 94.03, 152.5 , 1709.],
...,
[108.3 , 858.1 , 48.55, 126.7 , 1124.],
[140.1 , 1265. , 86.22, 184.6 , 1821.],
[47.92, 181. , 19.15, 59.16, 268.6]])

```
x_kbest = pd.DataFrame(x_kbest)
x_percentile = pd.DataFrame(x_percentile)
```

```
x_kbest.head()
```

	0	1	2	3	4
0	122.80	1001.0	153.40	184.60	2019.0
1	132.90	1326.0	74.08	158.80	1956.0
2	130.00	1203.0	94.03	152.50	1709.0
3	77.58	386.1	27.23	98.87	567.7
4	135.10	1297.0	94.44	152.20	1575.0

```
x_percentile.head()
```

	0	1	2	3	4	5
0	122.80	1001.0	153.40	25.38	184.60	2019.0
1	132.90	1326.0	74.08	24.99	158.80	1956.0
2	130.00	1203.0	94.03	23.57	152.50	1709.0

```
from sklearn.decomposition import PCA
```

```
pca = PCA(n_components = 5)
x_pca = pca.fit_transform(x)
```

```
x_pca = pd.DataFrame(x_pca)
```

```
x_pca.head()
```

	0	1	2	3	4
0	1160.142574	-293.917544	48.578398	-8.711975	32.000486
1	1269.122443	15.630182	-35.394534	17.861283	-4.334874
2	995.793889	39.156743	-1.709753	4.199340	-0.466529
3	-407.180803	-67.380320	8.672848	-11.759867	7.115461
4	930.341180	189.340742	1.374801	8.499183	7.613289

[Colab paid products](#) - [Cancel contracts here](#)

