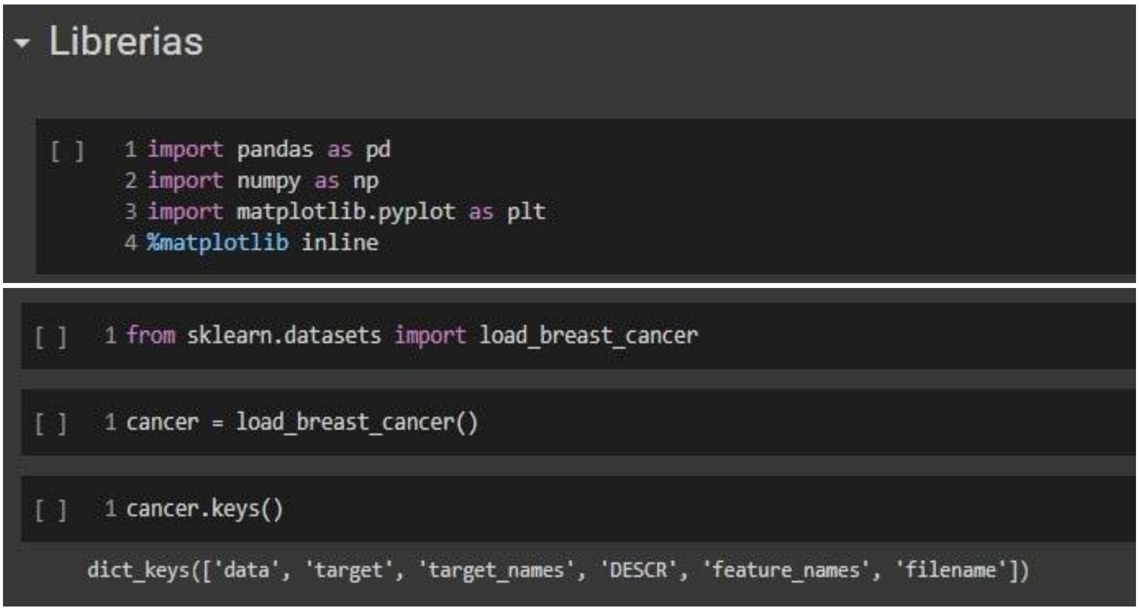
Recurso de Profundización

La siguiente figura muestra cómo implementar los componentes principales sobre el conjunto de datos de cáncer disponible en la librería de scikit-learn.

Implementación de componentes principales



Librerias

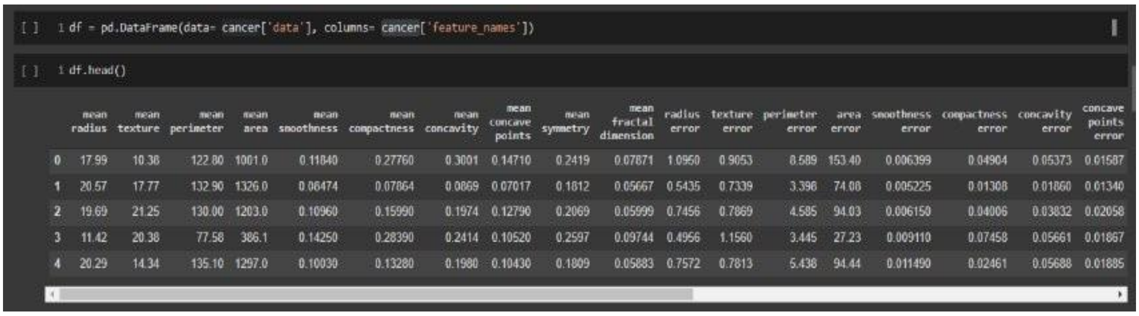
```
[ ] import pandas as pd
    import numpy as np
    import matplotlib.pyplot as plt
    %matplotlib inline
```

```
[ ] from sklearn.datasets import load_breast_cancer
```

```
[ ] cancer = load_breast_cancer()
```

```
[ ] cancer.keys()

dict_keys(['data', 'target', 'target_names', 'DESCR', 'feature_names', 'filename'])
```

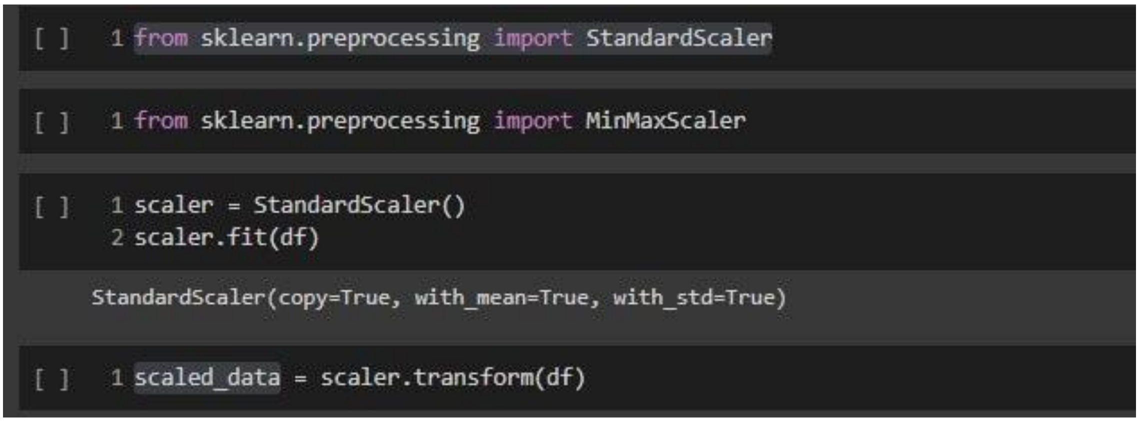


```
[ ] df = pd.DataFrame(data= cancer['data'], columns= cancer['feature_names'])
```

```
[ ] df.head()
```

```
     mean     mean       mean    mean        mean         mean       mean     mean      mean       mean  radius  texture  perimeter    area  smoothness  compactness  concavity  concave
   radius  texture  perimeter    area  smoothness  compactness  concavity  concave  symmetry    fractal   error    error      error   error       error        error      error   points
                                                                             points             dimension                                                                          error
0   17.99    10.38     122.80  1001.0     0.11840      0.27760     0.3001  0.14710    0.2419    0.07871  1.0950   0.9053      8.589  153.40    0.006399      0.04904    0.05373  0.01587
1   20.57    17.77     132.90  1326.0     0.08474      0.07864     0.0869  0.07017    0.1812    0.05667  0.5435   0.7339      3.398   74.08    0.005225      0.01308    0.01860  0.01340
2   19.69    21.25     130.00  1203.0     0.10960      0.15990     0.1974  0.12790    0.2069    0.05999  0.7456   0.7869      4.585   94.03    0.006150      0.04006    0.03832  0.02058
3   11.42    20.38      77.58   386.1     0.14250      0.28390     0.2414  0.10520    0.2597    0.09744  0.4956   1.1560      3.445   27.23    0.009110      0.07458    0.05661  0.01867
4   20.29    14.34     135.10  1297.0     0.10030      0.13280     0.1980  0.10430    0.1809    0.05883  0.7572   0.7813      5.438   94.44    0.011490      0.02461    0.05688  0.01885
```

Profundiza más



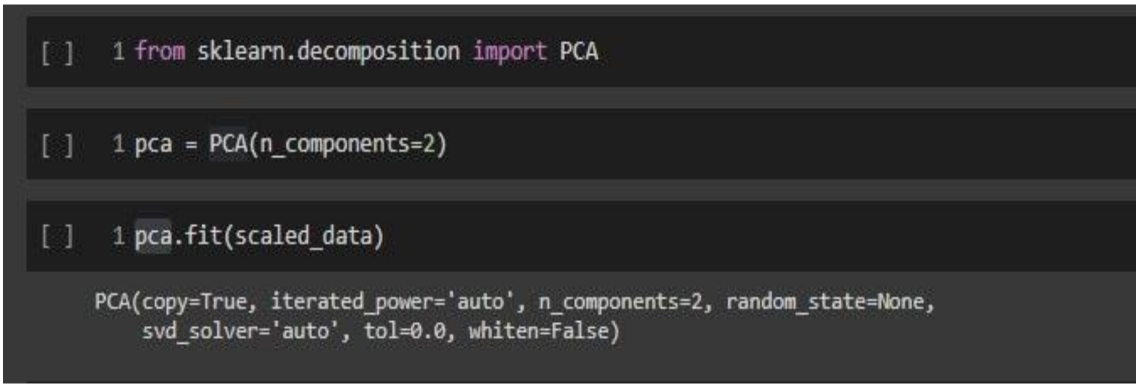
```
[ ]  from sklearn.preprocessing import StandardScaler
```

```
[ ]  from sklearn.preprocessing import MinMaxScaler
```

```
[ ]  scaler = StandardScaler()
     scaler.fit(df)

  StandardScaler(copy=True, with_mean=True, with_std=True)
```

```
[ ]  scaled_data = scaler.transform(df)
```

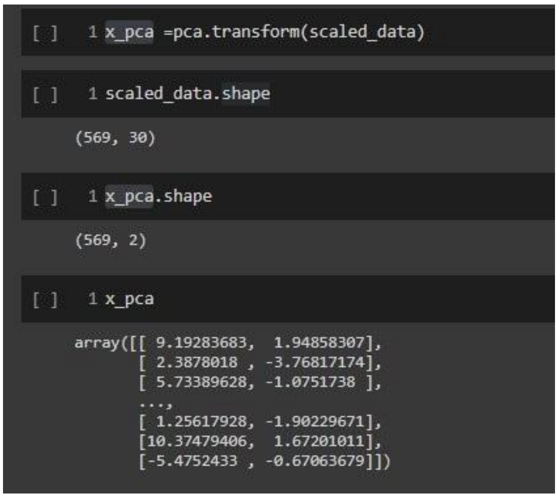


```
[ ]  from sklearn.decomposition import PCA
```

```
[ ]  pca = PCA(n_components=2)
```

```
[ ]  pca.fit(scaled_data)

  PCA(copy=True, iterated_power='auto', n_components=2, random_state=None,
      svd_solver='auto', tol=0.0, whiten=False)
```



```
[ ]  x_pca =pca.transform(scaled_data)
```

```
[ ]  scaled_data.shape

  (569, 30)
```

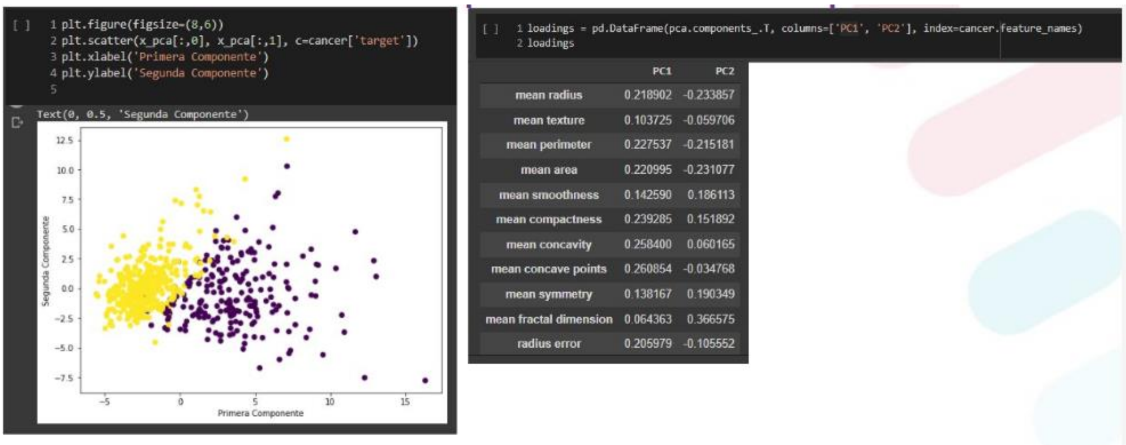
```
[ ]  x_pca.shape

  (569, 2)
```

```
[ ]  x_pca

  array([[ 9.19283683,  1.94858307],
         [ 2.3878018 , -3.76817174],
         [ 5.73389628, -1.0751738 ],
         ...,
         [ 1.25617928, -1.90229671],
         [10.37479406,  1.67201011],
         [-5.4752433 , -0.67063679]])
```

Profundiza más



```
[ ]  plt.figure(figsize-(8,6))
     plt.scatter(x_pca[:,0], x_pca[:,1], c=cancer['target'])
     plt.xlabel('Primera Componente')
     plt.ylabel('Segunda Componente')

Text(0, 0.5, 'Segunda Componente')
```

```
[ ]  loadings = pd.DataFrame(pca.components_.T, columns=['PC1', 'PC2'], index=cancer.feature_names)
     loadings
```

	PC1	PC2
mean radius	0.218902	-0.233857
mean texture	0.103725	-0.059706
mean perimeter	0.227537	-0.215181
mean area	0.220995	-0.231077
mean smoothness	0.142590	0.186113
mean compactness	0.239285	0.151892
mean concavity	0.258400	0.060165
mean concave points	0.260854	-0.034768
mean symmetry	0.138167	0.190349
mean fractal dimension	0.064363	0.366575
radius error	0.205979	-0.105552

Fuente: Elaboración propia