```
# Import necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.datasets import load breast cancer
from sklearn.preprocessing import StandardScaler
from sklearn.feature selection import SelectKBest, f classif
# Load the breast cancer dataset
data = load breast cancer()
df = pd.DataFrame(data.data, columns=data.feature names)
df['target'] = data.target
df
     mean radius mean texture mean perimeter mean area mean
smoothness \
           17.99
                         10.38
                                         122.80
                                                     1001.0
0.11840
                         17.77
1
           20.57
                                         132.90
                                                     1326.0
0.08474
           19.69
                         21.25
                                         130.00
                                                     1203.0
0.10960
3
           11.42
                         20.38
                                          77.58
                                                     386.1
0.14250
4
           20.29
                         14.34
                                         135.10
                                                     1297.0
0.10030
. .
                          22.39
564
           21.56
                                         142.00
                                                     1479.0
0.11100
           20.13
                         28.25
                                         131.20
                                                     1261.0
565
0.09780
566
           16.60
                         28.08
                                         108.30
                                                     858.1
0.08455
567
           20.60
                         29.33
                                         140.10
                                                     1265.0
0.11780
568
                         24.54
                                          47.92
                                                     181.0
            7.76
0.05263
     mean compactness
                       mean concavity mean concave points mean
symmetry \
0
              0.27760
                               0.30010
                                                     0.14710
0.2419
              0.07864
                               0.08690
                                                     0.07017
1
0.1812
              0.15990
                               0.19740
                                                     0.12790
0.2069
3
              0.28390
                               0.24140
                                                     0.10520
```

0.2597 4 0.13280 0.19800 0.10430 0.1809 564 0.11590 0.24390 0.13890 0.1726 0.10340 0.14400 0.07701	
0.1809 564 0.11590 0.24390 0.13890 0.1726	
564 0.11590 0.24390 0.13890 0.1726	
564 0.11590 0.24390 0.13890 0.1726	
564 0.11590 0.24390 0.13890 0.1726	
6.5 0.10040 0.14400 0.00701	
565 0.10340 0.14400 0.09791	
0.1752	
566 0.10230 0.09251 0.05302 0.1590	
567 0.27700 0.35140 0.15200	
0.2397	
568 0.04362 0.00000 0.00000	
0.1587	
mean fractal dimension worst radius worst texture 0 0.07871 25.380 17.33	\
1 0.05667 24.990 23.41 2 0.05999 23.570 25.53 3 0.09744 14.910 26.50 4 0.05883 22.540 16.67	
3 0.09744 14.910 26.50	
4 0.05883 22.540 16.67	
564	
564 0.05623 25.450 26.40 565 0.05533 23.690 38.25	
566 0.05648 18.980 34.12	
567 0.07016 25.740 39.42	
568 0.05884 9.456 30.37	
worst perimeter worst area worst smoothness worst compa	ctness
\ 0	.66560
0 104.00 2013.0 0.10220 0	.00500
1 158.80 1956.0 0.12380 0	.18660
2 152 50 1700 0 0 14440	42.450
2 152.50 1709.0 0.14440 0	.42450
3 98.87 567.7 0.20980 0	.86630
5 55167 55717 5126565 5	
	.20500
4 152.20 1575.0 0.13740 0	
4 152.20 1575.0 0.13740 0	
4 152.20 1575.0 0.13740 0 	
	.21130
564 166.10 2027.0 0.14100 0	.21130
564 166.10 2027.0 0.14100 0 565 155.00 1731.0 0.11660 0	.21130
564 166.10 2027.0 0.14100 0 565 155.00 1731.0 0.11660 0	.21130

568		59.16	2	268.6		0.08996	õ		0.06444		
	worst	concavity	worst	concave	points	worst	symmetr	·y \			
0		0.7119			0.2654		0.460				
1		0.2416 0.4504			0.1860 0.2430		0.275 0.361				
2		0.6869			0.2575		0.663				
4		0.4000			0.1625		0.236	4			
 564		0.4107			0.2216		0.206				
565		0.3215			0.1628		0.257				
566		0.3403			0.1418		0.221				
567 568		0.9387 0.0000			0.2650 0.0000		0.408 0.287				
500		0.0000			0.0000		0.207	_			
	worst	fractal d									
0 1			0.11890								
2			0.08758								
3 4			0.17300								
4			0.07678								
564			0.07115								
565			0.06637								
566 567			0.07820								
568			0.07039								
[569	rows	k 30 column	ns l								
_	[569 rows x 30 columns]										
# Display the first few rows of the dataset											
df.he	ead()										
	an rad		texture	e mean p	perimete	er mear	n area	mean			
smoot 0	hness:	\ 7.99	10.38	}	122.8	30 -	1001.0				
0.118			10.50	,	122.(1001.0				
1		9.57	17.77	1	132.9	90 :	1326.0				
0.084 2		9.69	21.25		130.0	າດ 1	1203.0				
0.109		7.05	21.25	•	150.		1205.0				
3		1.42	20.38	3	77.	58	386.1				
0.142 4		9.29	14.34		135.3	10	1297.0				
0.100		7123	17.35		1331.		123710				
mo)	nnactnass	moon of	ncavity	moon	concave	nointa	moan			
ille	an Col	mpactness	mean CC	meavity	mean (concave	POTITES	mean			

```
symmetry \
                              0.3001
                                                   0.14710
0
            0.27760
0.2419
            0.07864
                              0.0869
                                                   0.07017
1
0.1812
            0.15990
                              0.1974
                                                   0.12790
0.2069
            0.28390
                              0.2414
                                                   0.10520
0.2597
            0.13280
                              0.1980
                                                   0.10430
0.1809
   mean fractal dimension ... worst texture worst perimeter worst
area \
                   0.07871
                                          17.33
                                                           184.60
2019.0
                   0.05667
                                          23.41
                                                           158.80
1956.0
                   0.05999
                                          25.53
                                                           152.50
1709.0
                   0.09744
                                          26.50
                                                            98.87
567.7
                   0.05883
                                          16.67
                                                           152.20
1575.0
   worst smoothness worst compactness worst concavity worst concave
points
                                 0.6656
                                                   0.7119
0
             0.1622
0.2654
             0.1238
                                 0.1866
                                                   0.2416
0.1860
                                 0.4245
                                                   0.4504
             0.1444
0.2430
             0.2098
                                 0.8663
                                                   0.6869
0.2575
             0.1374
                                 0.2050
                                                   0.4000
0.1625
   worst symmetry
                   worst fractal dimension
                                             target
0
           0.4601
                                     0.11890
                                                   0
1
           0.2750
                                     0.08902
                                                   0
2
           0.3613
                                     0.08758
                                                   0
3
           0.6638
                                     0.17300
                                                   0
           0.2364
                                     0.07678
[5 rows x 31 columns]
# Task 3: Remove constant columns
df = df.loc[:, (df != df.iloc[0]).any()]
```

```
# Task 4: Encode categorical features (if any)
# In this dataset, encoding is not necessary as all features are
numerical

# Task 5: Feature scaling
numerical_features = df.drop('target', axis=1).columns
scaler = StandardScaler()
df[numerical_features] = scaler.fit_transform(df[numerical_features])

# Task 6: Correlation analysis
correlation_matrix = df.corr()
plt.figure(figsize=(12, 10))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm',
fmt=".2f")
plt.title("Correlation Matrix")
plt.show()
```

```
Correlation Matrix
                                                                                           1.0
      mean texture <del>0.32</del>.0<mark>0.33.33.002.24.30.29.07.08.28.39.28.26.0D.19.14.16.0D.05.35.99.36.34.08.28.30.30.1D.12.2</mark>
    mean perimeter -1.00,331.00,99.21.50,720.80,180.20,650.09,69.7-0.20,25.25.45.08.00,970.30,970.99.150,460.50,70,150.050,7
        mean area -0.99.320.99.00.180.50.69.80.150.20.7-0.00.7-0.80.170.20.20.350.07.00.09.09.29.960.90.170.39.50.70.140.00.7
                                                                                           - 0.8
   mean concavity = 0.60.30.72.60.50.88.000.92.500.34.60.08.60.60.10.62.60.10.45.60.30.73.60.45.73.83.80.42.550.70
                                                                                           - 0.6
 mean fractal dimension -0.3-0.08.26.20.50.50.50.30.10.46.00.00.16.040.09.40.56.45.340.35.60.25.05.25.25.50.46.35.18.35.70.01
       radius error -0.60.28.69.70.30.50.60.70.30.00.00.20.20.970.90.10.30.30.50.20.20.70.19.70.70.14.29.30.50.09.050.5
      perimeter error - 0.60.28.69.70.30.50.60.70.30.04.90.22.000.94.15.420.30.50.20.24.70.24.70.20.70.70.15.340.420.55.10.09.5
        area error -0.7-0.26.7-4.80.25.40.62.60.220.00.90.10.94.00.08.28.270.420.130.13.7-0.20.7-6.80.130.280.390.5-4.070.020.5
                                                                                           - 0.2
   smoothness error -0.20.00.20.10.33.14.10.03.19.40.16.40.15.08.00.34.20.33.40.43.23.00.20.18.30.06.06.10.10.10.07
   compactness error -0.2 D.1 9.2 5.2 D.3 2 D.3 2 D.4 6 0.4 9.4 2 D.5 6 3 6 2 3 .4 2 D.2 8 3 4 .0 0 8 0 7 0 .3 9 .8 0 .2 0 .1 4 .2 6 .2 0 .2 5 .6 5 6 0.4 8 .2 6 .5 0 .2 9
     concavity error -0.19.14.23.2D.25.50.60.440.340.45.33.19.36.2D.20.80.00.70.30.70.19.10.23.19.1D.48.60.440.20.440.25
 concave points error +0.38.16.4D.370.38.640.68.60.390.340.510.29.560.420.350.740.72.000.350.600.380.390.340.220.480.550.60.140.310.4
                                                                                           - 0.0
     fractal dimension error -0.04.05.0-0.02.28.50.45.26.30.60.25.28.24.15.40.80.75.60.27.00.04.06.00.00.00.17.39.38.22.11
      - -0.2
      worst perimeter -0.90.30.970.90.24.59.70.80.220.20.70.10.70.70.20.20.20.39.10.00.99.37.000.90.20.50.60.80.20.10.70
        worst area -0.90.340.940.90.20.50.680.80.180.20.750.00.730.810.180.20.190.340.1-0.00.980.350.980.00.20.440.540.70.20.050.7
                                                                                           -0.4
   worst smoothness -0.12.0.0.15.1.2.80.570.45.450.50.140.00.130.130.130.230.170.220.00.170.220.230.240.210.00.570.520.550.490.650
  worst compactness 0.4D.28.46.39.40.8D.75.60.4D.46.290.09.340.280.00.60.48.45.06.390.48.36.530.440.57
    worst concavity 4.533.30.50.50.43.82.83.75.43.33.33.33.33.00.42.39.00.64.60.53.04.33.57.37.60.54.52.82.00.80.53
 -0.6
    worst symmetry -0.1@.1D.1@.140.3@.5D.4D.30.70.30.09.1B.1D.070.1D.2@.2@.140.3@.1D.240.20.2D.2D.49.60.50.50.40.00.540.4
target 0.7-9.42.7-9.70.30.60.70.70.38.010.50.010.50.50.070.29.25.40.00.00.79.40.79.79.42.59.60.79.40.31.00
```

```
# Task 7: Feature selection
X = df.drop('target', axis=1)
y = df['target']
k_best_selector = SelectKBest(score_func=f_classif, k=5)
X_new = k_best_selector.fit_transform(X, y)
# Display the selected features
selected_features = X.columns[k_best_selector.get_support()]
print("\nSelected Features:")
print(selected_features)
Selected Features:
Index(['mean perimeter', 'mean concave points', 'worst radius',
```

radius error texture error perimeter error area error smoothness error -

mean concavity

mean smoothness

mean symmetry mean fractal dimension worst perimeter worst area -

worst smoothness

error

fractal dimension

error

concave points error symmetry error

compactness error

worst compactness worst concavity -

worst concave points

worst fractal dimension

```
'worst perimeter', 'worst concave points'],
dtype='object')
```