```
import numpy as np
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn import svm
from sklearn.metrics import accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns
```

```
dataset = pd.read_csv("diabetes.csv")
dataset.head()
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedi
0	6	148	72	35	0	33.6	
1	1	85	66	29	0	26.6	
2	8	183	64	0	0	23.3	
3	1	89	66	23	94	28.1	
4	0	137	40	35	168	43.1	

```
dataset.isnull().sum()
```

```
Pregnancies                 0
Glucose                     0
BloodPressure               0
SkinThickness               0
Insulin                     0
BMI                         0
DiabetesPedigreeFunction    0
Age                         0
Outcome                     0
dtype: int64
```

```
dataset.describe()
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outc
count	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.00(
mean	3.845052	120.894531	69.105469	20.536458	79.799479	31.992578	0.471876	33.240885	0.34
std	3.369578	31.972618	19.355807	15.952218	115.244002	7.884160	0.331329	11.760232	0.47(
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.078000	21.000000	0.00(
25%	1.000000	99.000000	62.000000	0.000000	0.000000	27.300000	0.243750	24.000000	0.00(
50%	3.000000	117.000000	72.000000	23.000000	30.500000	32.000000	0.372500	29.000000	0.00(
75%	6.000000	140.250000	80.000000	32.000000	127.250000	36.600000	0.626250	41.000000	1.00(
max	17.000000	199.000000	122.000000	99.000000	846.000000	67.100000	2.420000	81.000000	1.00(

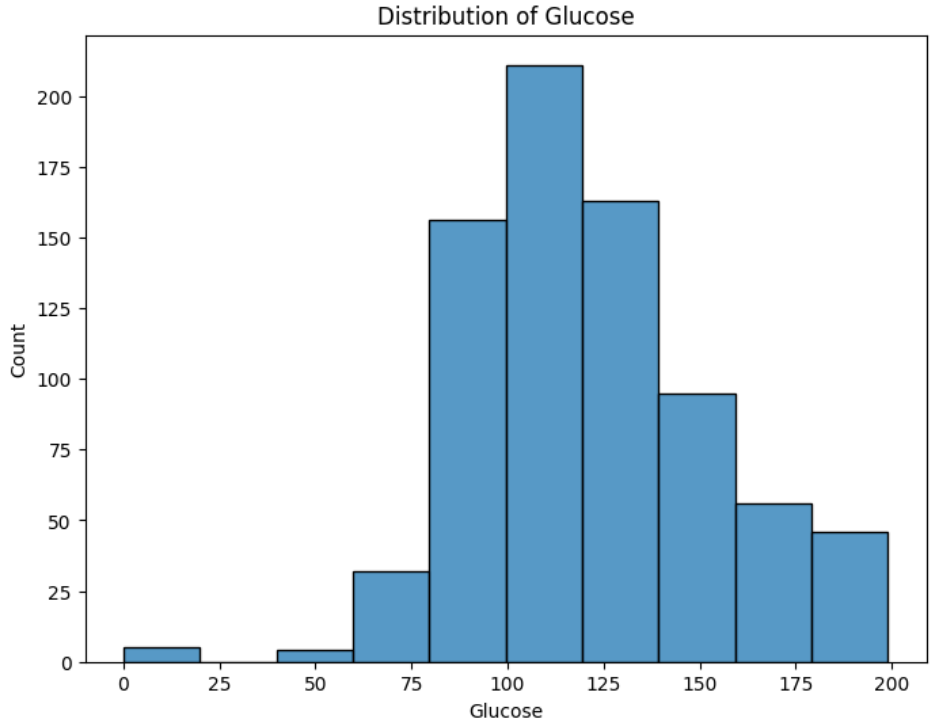
```
dataset.groupby("Outcome").mean()
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
Outcome								
0	3.298000	109.980000	68.184000	19.664000	68.792000	30.304200	0.429734	31.190000
1	4.865672	141.257463	70.824627	22.164179	100.335821	35.142537	0.550500	37.067164

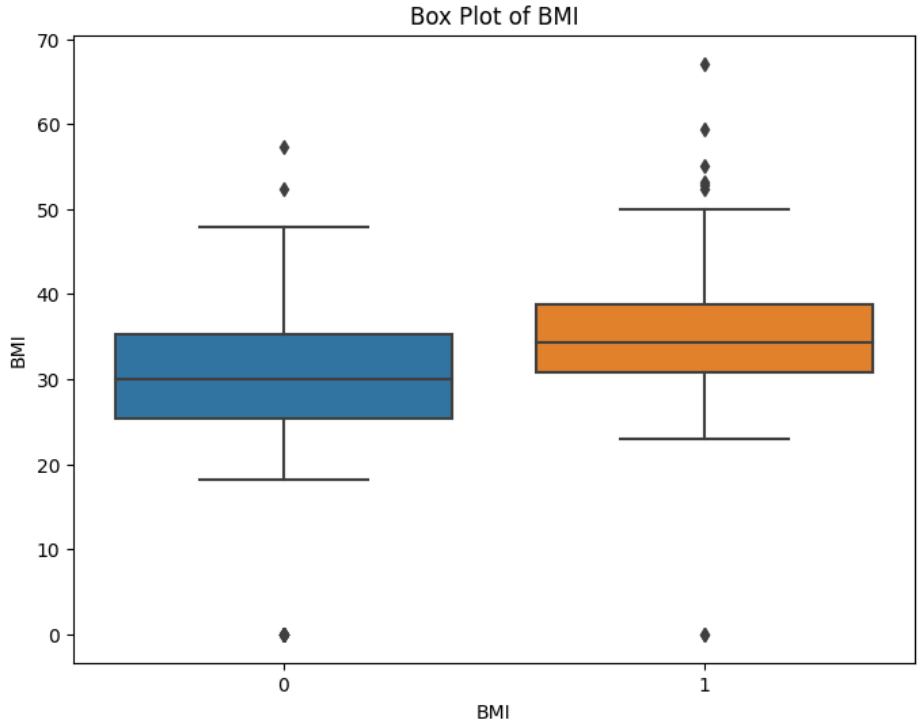
```
dataset['Outcome'].value_counts()
```

```
0    500
1    268
Name: Outcome, dtype: int64
```

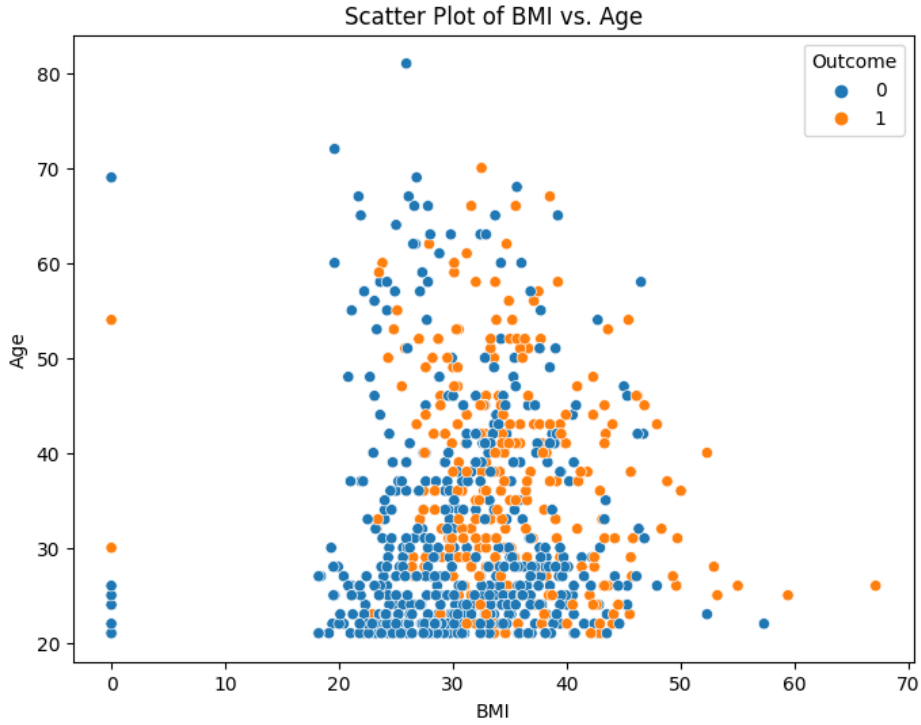
```
# Histogram of Glucose feature
plt.figure(figsize=(8, 6))
sns.histplot(dataset['Glucose'], bins=10)
plt.title("Distribution of Glucose")
plt.xlabel("Glucose")
plt.ylabel("Count")
plt.show()
```




```
# Box plot of BMI feature
plt.figure(figsize=(8, 6))
sns.boxplot(data=dataset, x='Outcome', y='BMI')
plt.title("Box Plot of BMI")
plt.xlabel("BMI")
plt.show()
```



```
plt.figure(figsize=(8, 6))
sns.scatterplot(data=dataset, x='BMI', y='Age', hue='Outcome')
plt.title("Scatter Plot of BMI vs. Age")
plt.xlabel("BMI")
plt.ylabel("Age")
plt.show()
```




```
X = dataset.drop('Outcome', axis = 1).values
y = dataset['Outcome'].values
```

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,y, test_size = 0.25)
```

```
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

```
classifier = svm.SVC()
param_grid = {'kernel' : ['linear', 'rbf','polynomial'], 'C': [0.1, 1, 10]}
```

```
from sklearn.model_selection import GridSearchCV
grid_search = GridSearchCV(classifier, param_grid)
```

```
grid_search.fit(X_train, y_train)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/model_selection/_validation.py:378: FitFailedWarning:
15 fits failed out of a total of 45.
The score on these train-test partitions for these parameters will be set to nan.
If these failures are not expected, you can try to debug them by setting error_score='raise'.

Below are more details about the failures:
--------------------------------------------------------------------------------
15 fits failed with the following error:
Traceback (most recent call last):
  File "/usr/local/lib/python3.10/dist-packages/sklearn/model_selection/_validation.py", line 686, in _fit_and_score
    estimator.fit(X_train, y_train, **fit_params)
  File "/usr/local/lib/python3.10/dist-packages/sklearn/svm/_base.py", line 180, in fit
    self._validate_params()
  File "/usr/local/lib/python3.10/dist-packages/sklearn/base.py", line 600, in _validate_params
    validate_parameter_constraints(
  File "/usr/local/lib/python3.10/dist-packages/sklearn/utils/_param_validation.py", line 97, in validate_parameter_constr
    raise InvalidParameterError(
sklearn.utils._param_validation.InvalidParameterError: The 'kernel' parameter of SVC must be a str among {'rbf', 'linear',

  warnings.warn(some_fits_failed_message, FitFailedWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/model_selection/_search.py:952: UserWarning: One or more of the test score
 0.77086957 0.72574213        nan]
  warnings.warn(
```

▸ **GridSearchCV**

```
y_prediction = grid_search.predict(X_test)
```

▸ SVC

```
accuracy_score(y_prediction, y_test)
```

```
0.734375
```

```
from sklearn.metrics import classification_report
print(classification_report(y_prediction, y_test))
```

```
                   precision    recall  f1-score   support

               0       0.84      0.77      0.81       137
               1       0.53      0.64      0.58        55

        accuracy                           0.73       192
       macro avg       0.69      0.71      0.69       192
    weighted avg       0.75      0.73      0.74       192
```

✓ 0s completed at 11:05 AM