DATE	**25 OCT 2023**
TEAM ID	**344**
PROJECT NAME	**AI BASED DIABETES PREDICTION**
NAME	**gowtham.p**

PROJECT NAME :AI BASED DIABETES PREDICTION
PHASE 4;

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, confusion_matrix,
precision_score, recall_score
```

In [2]:

```
# Load the dataset
data = pd.read_csv("input/diabetes-data-set/diabetes.csv")
data.head()
```

Out[2]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome

0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

```
summary_stats = data.describe()
summary_stats
```

```
Out[3]:
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
count	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000

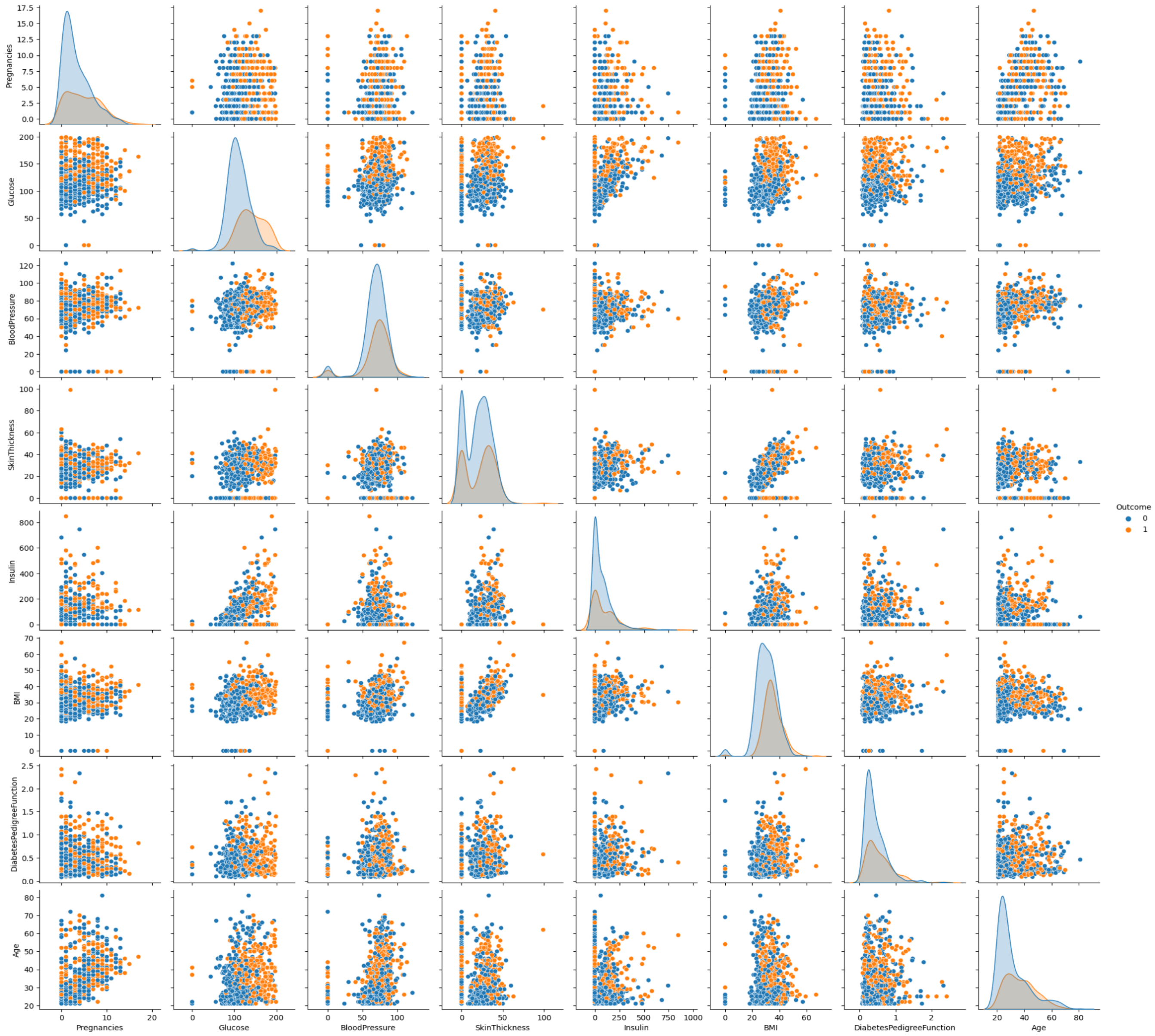
mean	3.845052	120.894531	69.105469	20.536458	79.799479	31.992578	0.471876	33.240885	0.348958
std	3.369578	31.972618	19.355807	15.952218	115.244002	7.884160	0.331329	11.760232	0.476951
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.078000	21.000000	0.000000
25%	1.000000	99.000000	62.000000	0.000000	0.000000	27.300000	0.243750	24.000000	0.000000
50%	3.000000	117.000000	72.000000	23.000000	30.500000	32.000000	0.372500	29.000000	0.000000
75%	6.000000	140.250000	80.000000	32.000000	127.250000	36.600000	0.626250	41.000000	1.000000
max	17.000000	199.000000	122.000000	99.000000	846.000000	67.100000	2.420000	81.000000	1.000000

```
class_distribution = data['Outcome'].value_counts()
class_distribution
```

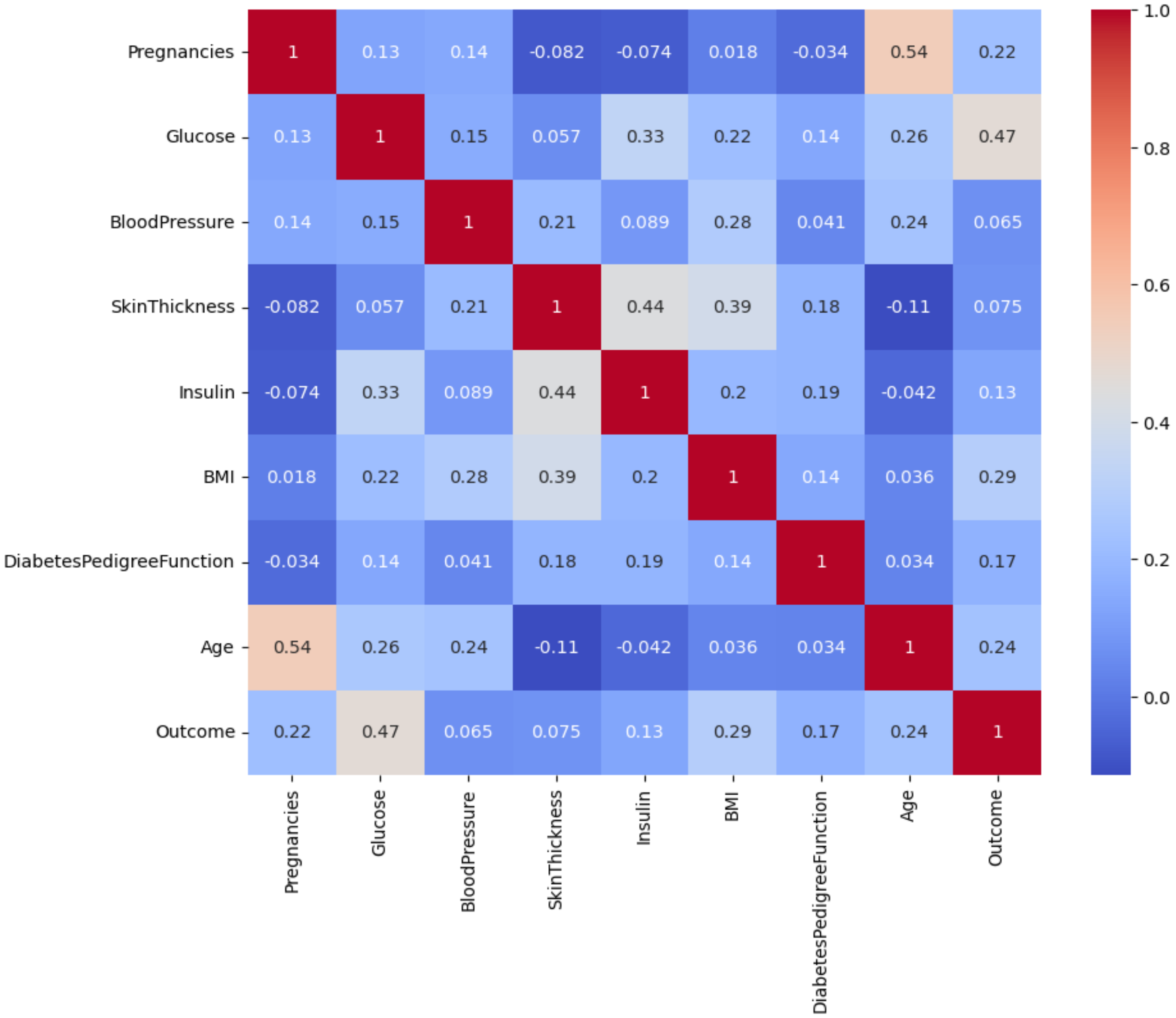
Output

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

```
sns.pairplot(data, hue='Outcome', diag_kind='kde')
plt.show()
```



```
correlation_matrix = data.corr()
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.show()
```



```
X = data.drop("Outcome", axis=1)
y = data["Outcome"]

X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, random_state=42)

rf_classifier = RandomForestClassifier(n_estimators=100,
random_state=42)
```

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

output

```
RandomForestClassifier(random_state=42)
```

```
y_pred = rf_classifier.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)
confusion = confusion_matrix(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)

tn, fp, fn, tp = confusion.ravel()
specificity = tn / (tn + fp)
```

code

```
print("Accuracy:", accuracy)
print("Confusion Matrix:")
print(confusion)
print("Precision:", precision)
print("Recall:", recall)
```

```
print("Specificity:", specificity)
```

```
Accuracy: 0.7532467532467533
Confusion Matrix:
[[121  30]
 [ 27  53]]
Precision: 0.6385542168674698
Recall: 0.6625
Specificity: 0.8013245033112583
```