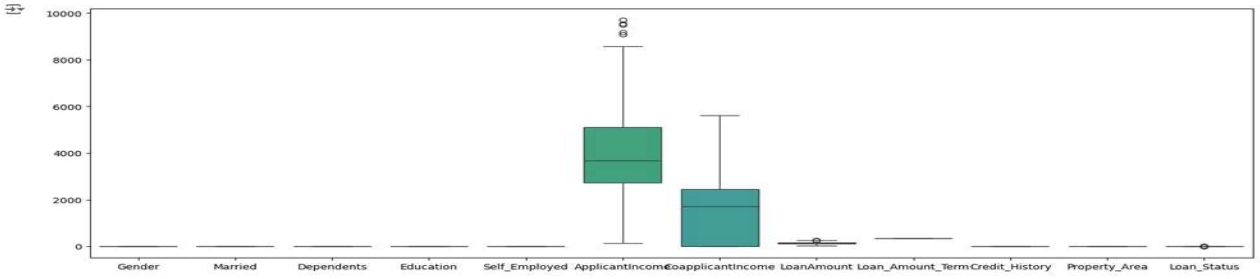
Smote On Outlier Analysis,Box Plot,Histograms,ROC curve, strengths and weaknesses in terms of sensitivity, specificity, precision, and accuracy

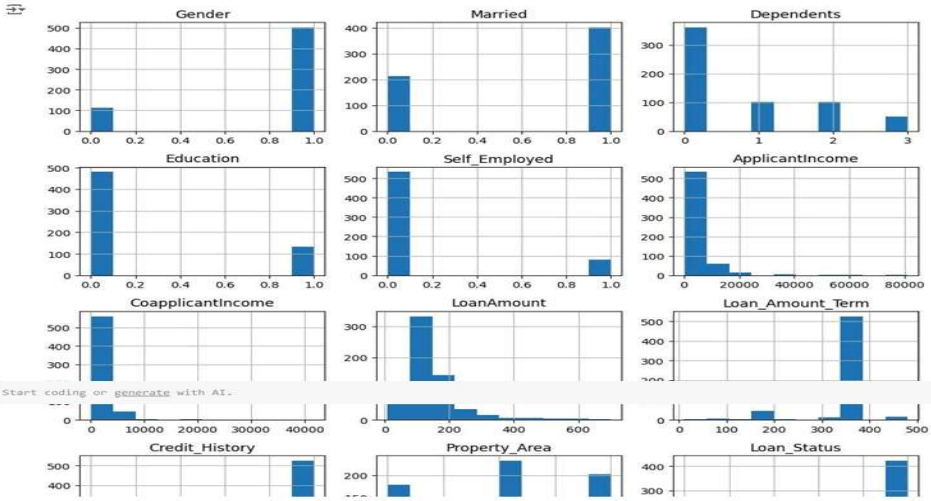
Date	3 August 2024
Team ID	740293
Project Title	Loan Sanction Amount Prediction Data With Ml
Maximum Marks	5 Marks

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
data

Q1 = data.quantile(0.25)
Q3 = data.quantile(0.75)
IQR = Q3 - Q1

data_out = data[~((data < (Q1 - 1.5 * IQR)) |(data > (Q3 + 1.5 * IQR))).any(axis=1)]
plt.figure(figsize=(18, 6))
sns.boxplot(data=data_out)
plt.show()
```

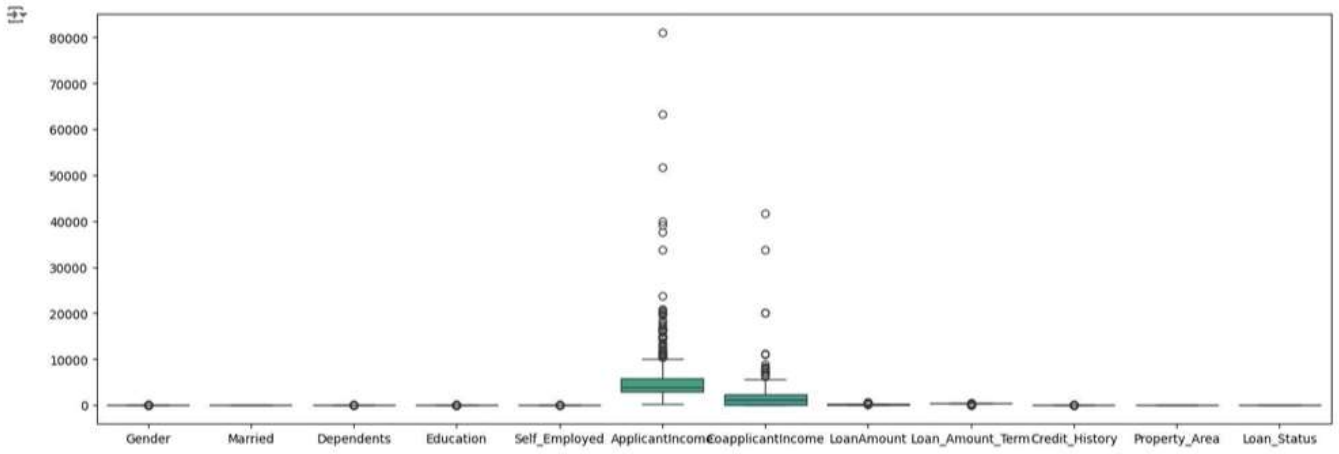
Start coding or generate with AI.

```
import matplotlib.pyplot as plt
data.hist(figsize=(10, 10))
plt.tight_layout()
plt.show()
```

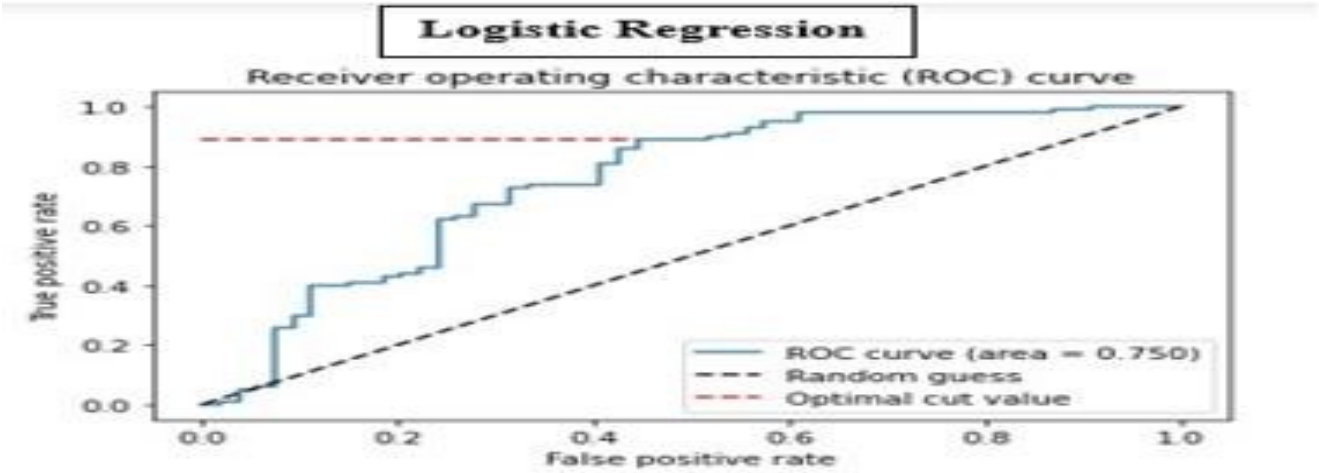



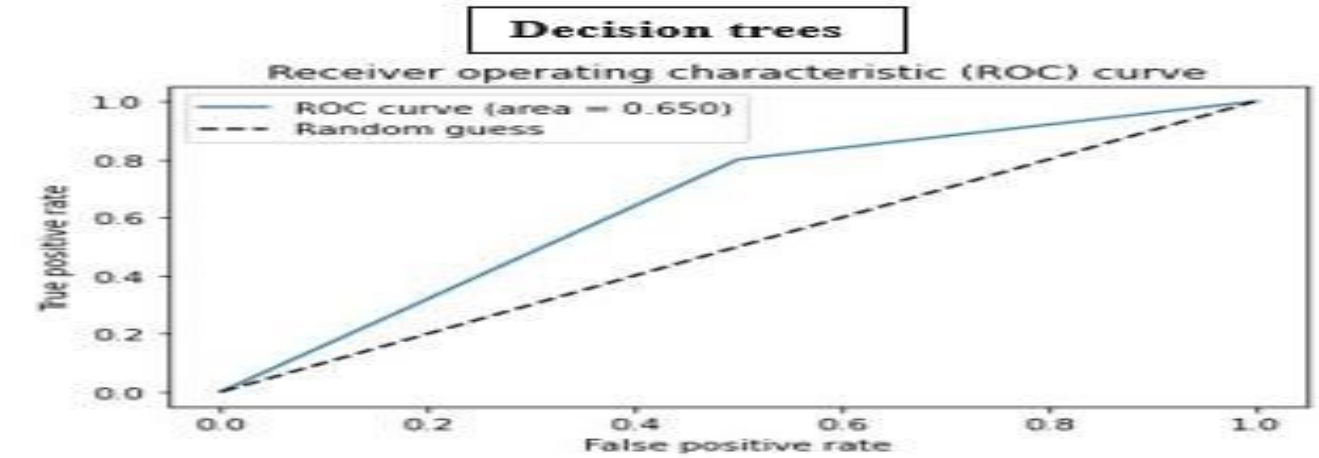
Start coding or generate with AI.

```
import matplotlib.pyplot as plt
import seaborn as sns
plt.figure(figsize=(18, 6))
sns.boxplot(data=data)
plt.show()
```

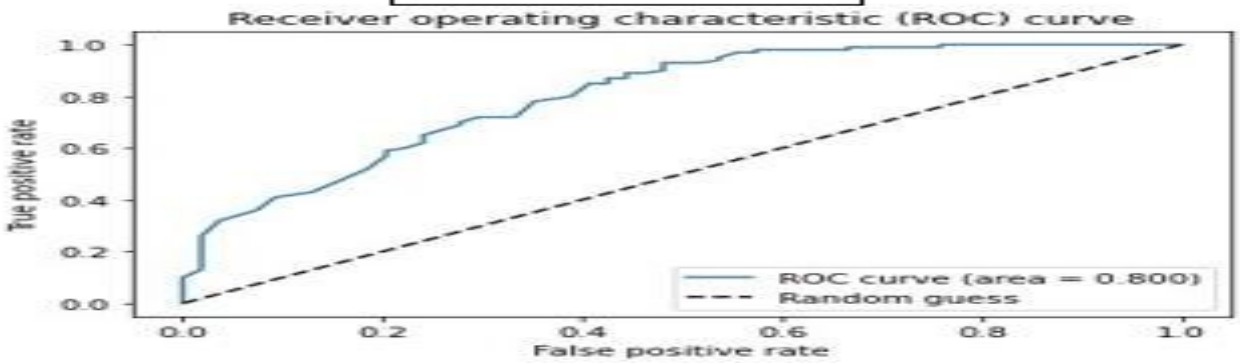


Among the models evaluated, random forest exhibits the highest AUC (0.80), suggesting superior discrimination ability compared to logistic regression (0.75) and decision tree (0.65). Therefore, based on the ROC curve analysis, the random forest model appears to be the most promising candidate for further exploration and deployment in this classification task.





Random Forest



Each model exhibits strengths and weaknesses in terms of sensitivity, specificity, precision, and accuracy. Logistic regression demonstrates high sensitivity and negative predictive value, while random forest shows competitive sensitivity and accuracy. Decision tree performs moderately across most metrics.

Logistic Regression

Metric	Value
Sensitivity	0.98
Specificity	0.388889
Positive Predictive Value	0.748092
Negative Predictive Value	0.913043
Accuracy	0.772727
Precision	0.748092

Decision Trees

Metric	Value
Sensitivity	0.8
Specificity	0.5
Positive Predictive Value	0.747664
Negative Predictive Value	0.574468
Accuracy	0.694805
Precision	0.747664

Random Forest

Metric	Value
Sensitivity	0.94
Specificity	0.462963
Positive Predictive Value	0.764228
Negative Predictive Value	0.806452
Accuracy	0.772727
Precision	0.764228