Loan Eligibility Code in Data Science

1. Import Necessary Libraries:

```python
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.preprocessing import LabelEncoder

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy_score, confusion_matrix
```

2. Load the Dataset:

```python
data = pd.read_csv("loan_data.csv")  # Replace with your file path
```

3. Exploratory Data Analysis (EDA):

```python
# Check for null values:

print(data.isnull().sum())


# Summary statistics:

print(data.describe())


# Visualizations:
```

```
sns.countplot(x='Loan_Status', data=data)

plt.show()
```

4. Data Preprocessing:

```
# Handle missing values:

data.fillna(data.median(), inplace=True)


# Encode categorical variables:

le = LabelEncoder()

for column in ['Gender', 'Married', 'Education', 'Property_Area']:

    data[column] = le.fit_transform(data[column])
```

5. Split Data:

```
X = data.drop('Loan_Status', axis=1)

y = data['Loan_Status']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

6. Build a Model:

```
# Train a Random Forest Classifier:

model = RandomForestClassifier(random_state=42)

model.fit(X_train, y_train)


# Predictions:

y_pred = model.predict(X_test)
```

7. Evaluate the Model:

```python
print("Accuracy:", accuracy_score(y_test, y_pred))

print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred))
```

8. Deployment and Save Model:

```python
import joblib

joblib.dump(model, 'loan_eligibility_model.pkl')
```