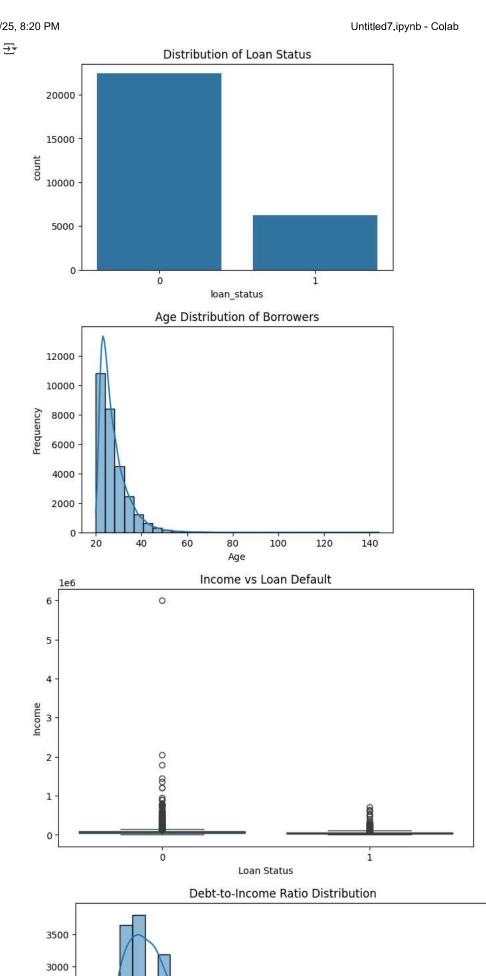
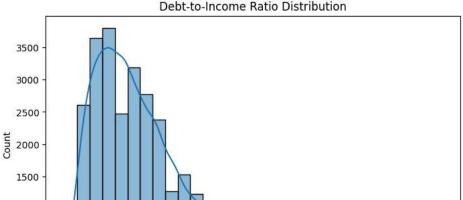
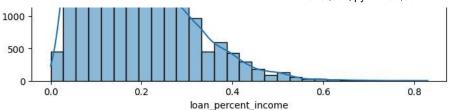
```
#Import libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
#Load dataset (change path if needed)
df = pd.read_csv("/content/credit_risk_dataset.csv")
#Basic info
print("Shape:", df.shape)
print(df.head())
print(df.info())
print(df['loan_status'].value_counts())  # target column
     Shape: (32581, 12)
       person_age person_income person_home_ownership person_emp_length \
     0
                           59000
                                                  RENT
     1
                21
                                                   OWN
     2
               25
                            9600
                                              MORTGAGE
                                                                      1.0
                            65500
                                                  RENT
     3
               23
                                                                      4.0
                           54400
     4
               24
                                                  RENT
                                                                      8.0
       loan_intent loan_grade loan_amnt loan_int_rate loan_status \
     0
         PERSONAL
                           D
                                  35000
                                                 16.02
         EDUCATION
                           В
                                   1000
                                                 11.14
                                                                  0
     2
          MEDICAL
                                   5500
                                                 12.87
                                                                  1
                           C
     3
           MEDICAL
                                  35000
                                                 15.23
     4
          MEDICAL
                           C
                                  35000
                                                 14.27
        loan_percent_income cb_person_default_on_file cb_person_cred_hist_length
     0
                      0.59
                      0.10
     1
                                                   N
     2
                      0.57
                                                   N
                                                                               3
     3
                      0.53
                                                   N
                                                                               2
                      0.55
                                                    Υ
     <class 'pandas.core.frame.DataFrame'>
     RangeIndex: 32581 entries, 0 to 32580
     Data columns (total 12 columns):
     #
         Column
                                     Non-Null Count Dtype
     0
                                     32581 non-null int64
          person age
          person_income
                                     32581 non-null int64
      1
          person_home_ownership
                                     32581 non-null object
      3
          person_emp_length
                                     31686 non-null float64
          loan_intent
                                     32581 non-null object
      5
          loan_grade
                                     32581 non-null object
                                     32581 non-null int64
          loan_amnt
                                     29465 non-null
          loan_int_rate
                                                     float64
                                     32581 non-null int64
         loan_status
         loan_percent_income
                                     32581 non-null float64
      10 cb_person_default_on_file
                                     32581 non-null
                                                     obiect
      11 cb_person_cred_hist_length 32581 non-null int64
     dtypes: float64(3), int64(5), object(4)
     memory usage: 3.0+ MB
     None
     loan_status
     0
          25473
          7108
     Name: count, dtype: int64
#Check missing values
print(df.isnull().sum())
#Drop rows with missing data (if very few)
df = df.dropna()
#Convert categorical variables if needed
cat_cols = df.select_dtypes(include=['object']).columns
print("Categorical columns:", cat_cols)
#Example: Encode categorical features
df = pd.get_dummies(df, drop_first=True)
→ person_age
     person_income
                                     0
     person_home_ownership
     person_emp_length
     loan intent
                                     0
     loan_grade
                                     0
                                     0
     loan_amnt
     loan_int_rate
                                   3116
     loan status
                                     0
     loan_percent_income
```

```
cb_person_default_on_file
     cb_person_cred_hist_length
     dtype: int64
     Categorical columns: Index(['person_home_ownership', 'loan_intent', 'loan_grade',
            'cb_person_default_on_file'],
           dtype='object')
#Distribution of Loan Status (Target)
plt.figure(figsize=(6, 4))
sns.countplot(x="loan_status", data=df)
plt.title("Distribution of Loan Status")
plt.show()
#Age Distribution
plt.figure(figsize=(6, 4))
sns.histplot(df['person_age'], bins=30, kde=True)
plt.title("Age Distribution of Borrowers")
plt.xlabel("Age")
plt.ylabel("Frequency")
plt.show()
#Income vs Loan Default
plt.figure(figsize=(8,5))
sns.boxplot(x="loan_status", y="person_income", data=df)
plt.title("Income vs Loan Default")
plt.xlabel("Loan Status")
plt.ylabel("Income")
plt.show()
#Debt-to-Income Ratio Distribution
plt.figure(figsize=(8, 5))
sns.histplot(df['loan_percent_income'], bins=30, kde=True)
plt.title("Debt-to-Income Ratio Distribution")
plt.show()
```







```
plt.figure(figsize=(10, 6))
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='coolwarm')
plt.title("Correlation Heatmap of Features")
plt.show()
```



Correlation Heatmap of Features 1.0 person age - 1 0.180.10.050.010.023.040.20.00050.20.340.90.0706.0106.0-340.600.00720.3501.00.004.90.90.004.90.78 person_income -0.18 1 0.140.20.00101.140.250.120.010.0380.10.0107045.030.014002800.6006401040103010300090027 person_emp_length -0.170.14 1 0.150.056.088 050.150.01802 0.26.0.00.028000 00 1020 0100 0 70 40.0607.00 700 990 10.02 loan_amnt 0.05 0.270.11 1 0.150.110.58.045.010.020.1+0.0101.040.002090.000008549.010.069.0902.070.0610.04 0.8 loan_int_rate 9.0±010004050.15 1 0.340.12.015.0±080060816.00608002.005.00±090200020070.480.330.20.130.5 loan_status -0.020.1-0.080.110.34 1 0.380.0060090.10.24 0.00.0380.060.021 0.72 0-00.016.330,180.10.0881.18 loan_percent_income -0.040.25.050.580.120.38 1 -0.00.010.040.302006999199015.00040082952Le-0.568059.04.036036 cb_person_cred_hist_length -0.860.120.15.045.015.015.015.015.015.015.003 1 .00060089027.070.06.0103.0350660044007.00098002400820085084 0.6 person_home_ownership_OTHER-0,000/3010.013001030103010300980-010006 110,017.00;00,700/00980153e-00.004/8e0050401010.010099.4002001 person home ownership OWNO.0022.00.027.0240060.0.041008901 1 0.200084015.004002795009000905.80500900140014 - 0.4 person_home_ownership_RENT -0.039.180.230.120.130.240.120.027058.29 1 .0045050.059.012050.036.064.054.03280602606.059 loan_intent_EDUCATION -0.0±0.017.0±0.00.00.0036.05.000203.404 1 -0.180.240.230.20.0003814.000001.0020058053 loan_intent_HOMEIMPROVEMENT 9.0766.0456.028.0420.020.035.019.060009.810.059.18 1 -0.170.160.10.068.00.026.00.0009.8059016 loan_intent_MEDICAL 6.010 @0.004800.3005806.010504030063010405 0.240.17 1 -0.230.0007.0004.0004.00043000300098 - 0.2 loan_intent_PERSONAL 6,0334014400.00000000070201.000403.53 e6003020701.0.230.160.22 1 -0.70,00901.0010007.00994e6005029 loan_intent_VENTUREG-0009702890.D0030037070008206.7014.095.03-0.230.160.230.21 1 00-705090500.0004.30082893.40034 loan_grade_B-0,007/2001.0008.204830075050;04220044840.0050290730506780286077009.007 1-0.340.250.1-0.0539030.32 0.0 loan_grade_C0<mark>.0935</mark>00:40 42:02<mark>0.37</mark>0.03.fl.e900597.D9341009.50634.0124005570446.010.00<mark>0.34 1 -</mark>0.1-8.0838.040:02<mark>6.42</mark> loan_grade_D e.012.0144.007069.480.3B.0680098010005854.0106026600400-73301-0.250.18 1-0.0680050.016.32 loan_grade_ED-0004916.017092.330.18.0590024010.015028000.01906410070040.12.083063110.015006.14 loan_grade_F 6.00%0108009.9770.2 0.1 0.0%00%269.40€4602600200%8607.8009.4€65505004%06-0.01 1 .008.962 -0.2loan_grade_G0-004300%010206D.1B.08%056066%00266014966600556053909138e005054==0.022801660**0**830=**11**0.047 cb_person_default_on_file_Y0-090780027020.040.50.18.03600840100001.459006391.600098929030.320.420.320.140.06204 loan_grade_E person_income loan_amnt loan_int_rate loan status person_cred_hist_length loan_intent_VENTURE oan grade F cb_person_default_on_file_Y person_emp_length loan percent income person_home_ownership_OTHER person_home_ownership_OWN loan_intent_EDUCATION person home ownership RENT oan_intent_HOMEIMPROVEMENT loan_intent_MEDICAL oan intent PERSONAL loan grade loan grade loan_grade_ oan grade e

```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix
# Features & Target
X = df.drop("loan_status", axis=1)
y = df["loan_status"]
```