**Case Study Python**

**By Afreen Sultana**

import pandas as pd

import matplotlib.pyplot as plt

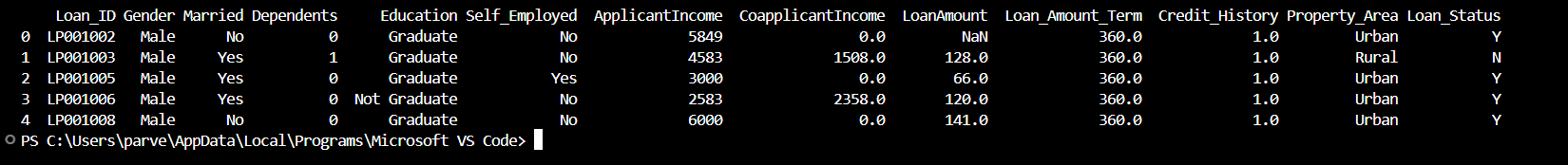
import seaborn as sns

**1.Loading Data in Pandas DataFrame**

df = pd.read\_csv(r"C:\Users\parve\Downloads\LoanData (1).csv")

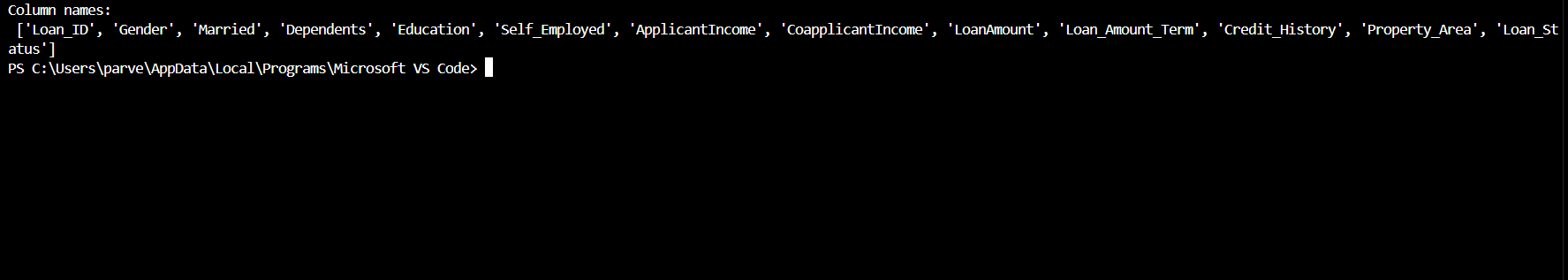
**2.Printing rows of the Data**

print(df.head())



**3.Printing the column names of the DataFrame**

print("Column names:\n",df.columns.tolist())



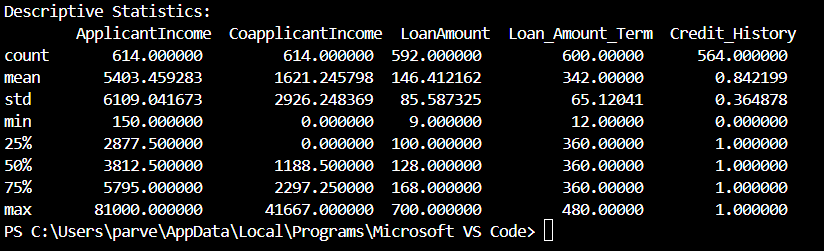
**4.Summary of Data Frame**

print("DataFrame Summary:\n",df.info())



**5.Descriptive Statistical Measures of a DataFrame**

print("Descriptive Statistics:\n",df.describe())



**6.Missing Data Handing**

print("\nMissing values in each column:\n", df.isnull().sum())

df['Gender'].fillna(df['Gender'].mode()[0], inplace=True)

df['Married'].fillna(df['Married'].mode()[0], inplace=True)

df['LoanAmount'].fillna(df['LoanAmount'].median(), inplace=True)

df['Self\_Employed'].fillna(df['Self\_Employed'].mode()[0], inplace=True)

df['Credit\_History'].fillna(df['Credit\_History'].mode()[0], inplace=True)

df.dropna(inplace=True)

print("\nAfter handling missing values:\n", df.isnull().sum())

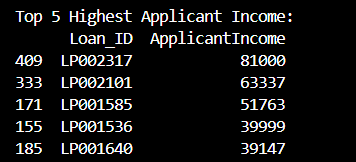
|  |  |
| --- | --- |
|  |  |

**7.Sorting DataFrame values**

sorted\_df = df.sort\_values(by='ApplicantIncome', ascending=False)

print("\nTop 5 Highest Applicant Income:")

print(sorted\_df[['Loan\_ID', 'ApplicantIncome']].head())



**8.Merge Data Frames**

merged\_df = pd.merge(df, df, on='Loan\_ID', suffixes=('\_left', '\_right'))

print("\nMerged DataFrame Shape:", merged\_df.shape)

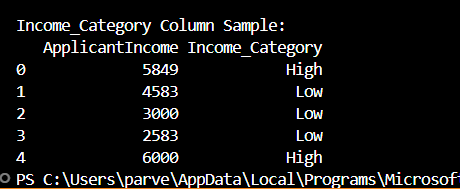


**9.Apply functions**

df['Income\_Category'] = df['ApplicantIncome'].apply(lambda x: 'High' if x > 5000 else 'Low')

print("\nIncome\_Category Column Sample:")

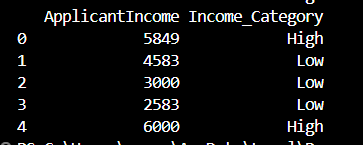
print(df[['ApplicantIncome', 'Income\_Category']].head())



**10.By using the lambda operator**

df['Income\_Category'] = df['ApplicantIncome'].map(lambda x: 'High' if x > 5000 else 'Low')

print(df[['ApplicantIncome', 'Income\_Category']].head())

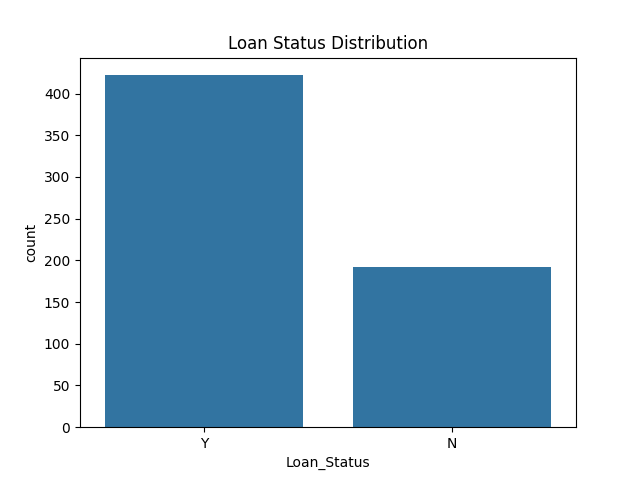


**11.Visualizing DataFrame**

sns.countplot(data=df, x='Loan\_Status')

plt.title("Loan Status Distribution")

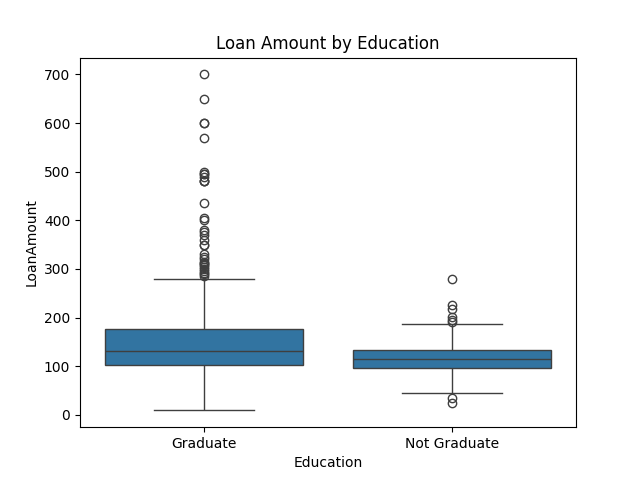
plt.show()



sns.boxplot(data=df, x='Education', y='LoanAmount')

plt.title("Loan Amount by Education")

plt.show()



sns.histplot(df['ApplicantIncome'], kde=True)

plt.title("Distribution of Applicant Income")

plt.show()

