

## Data Collection and Preprocessing Phase

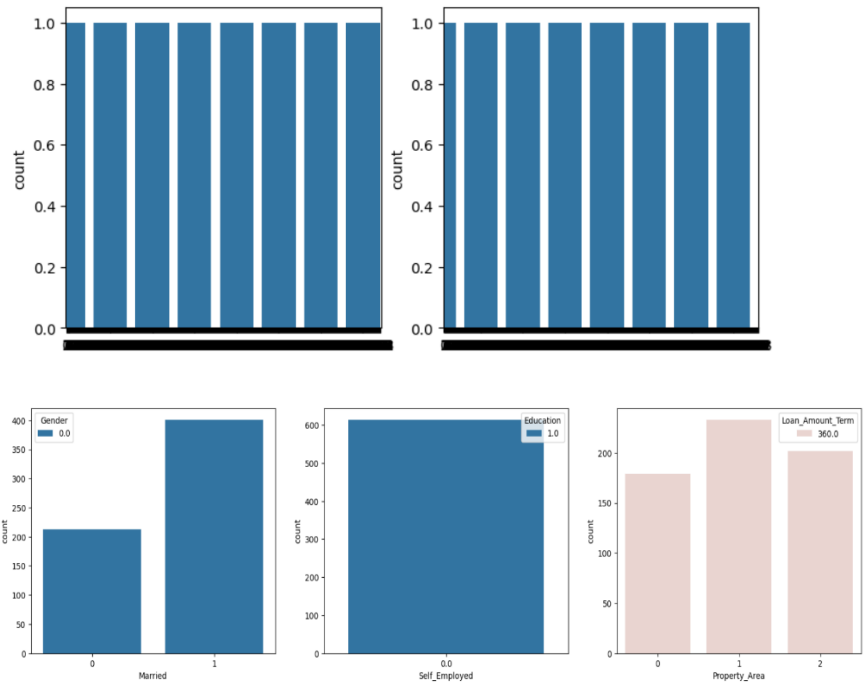
Date	11 July 2024
Team ID	740023
Project Title	Smart Lender - Applicant Credibility Prediction For Loan Approval
Maximum Marks	6 Marks

### Data Exploration and Preprocessing Template

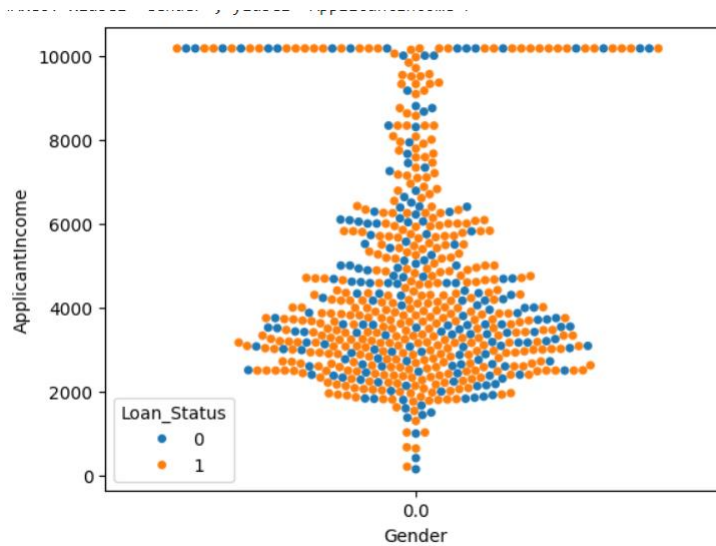
Data exploration and preprocessing for loan approval prediction begins with thorough exploratory data analysis (EDA) to understand distributions and relationships among features like credit scores, income levels, and loan amounts. Addressing missing data is critical; imputation methods such as mean, median, or predictive modeling are applied. Numerical features are scaled to ensure uniformity in influence, while categorical variables are encoded (e.g., one-hot encoding) for compatibility with machine learning models. These steps ensure the dataset is ready for training robust models that predict loan approval outcomes accurately based on processed data.

Section	Description																																																																																																																					
Data Overview	<div>Dimensions: 614 rows × 13 columns</div> <div>Descriptive statistics:</div> <table><thead><tr><th></th><th>Loan_ID</th><th>Gender</th><th>Married</th><th>Education</th><th>Self_Employed</th><th>ApplicantIncome</th><th>CoapplicantIncome</th><th>LoanAmount</th><th>Loan_Amount_Term</th><th>Credit_History</th><th>Property_Area</th><th>Loan_Status</th></tr></thead><tbody><tr><td>count</td><td>614.000000</td><td>614.0</td><td>614.000000</td><td>614.0</td><td>614.0</td><td>614.000000</td><td>614.000000</td><td>614.000000</td><td>614.0</td><td>614.0</td><td>614.000000</td><td>614.000000</td></tr><tr><td>mean</td><td>1999.785016</td><td>0.0</td><td>0.653094</td><td>1.0</td><td>0.0</td><td>4617.111564</td><td>1419.700733</td><td>137.078990</td><td>360.0</td><td>1.0</td><td>1.037459</td><td>0.687296</td></tr><tr><td>std</td><td>567.097644</td><td>0.0</td><td>0.476373</td><td>0.0</td><td>0.0</td><td>2479.851729</td><td>1624.606410</td><td>55.847766</td><td>0.0</td><td>0.0</td><td>0.787482</td><td>0.463973</td></tr><tr><td>min</td><td>1002.000000</td><td>0.0</td><td>0.000000</td><td>1.0</td><td>0.0</td><td>150.000000</td><td>0.000000</td><td>9.000000</td><td>360.0</td><td>1.0</td><td>0.000000</td><td>0.000000</td></tr><tr><td>25%</td><td>1532.750000</td><td>0.0</td><td>0.000000</td><td>1.0</td><td>0.0</td><td>2877.500000</td><td>0.000000</td><td>100.250000</td><td>360.0</td><td>1.0</td><td>0.000000</td><td>0.000000</td></tr><tr><td>50%</td><td>1993.500000</td><td>0.0</td><td>1.000000</td><td>1.0</td><td>0.0</td><td>3812.500000</td><td>1188.500000</td><td>125.000000</td><td>360.0</td><td>1.0</td><td>1.000000</td><td>1.000000</td></tr><tr><td>75%</td><td>2476.750000</td><td>0.0</td><td>1.000000</td><td>1.0</td><td>0.0</td><td>5795.000000</td><td>2297.250000</td><td>164.750000</td><td>360.0</td><td>1.0</td><td>2.000000</td><td>1.000000</td></tr><tr><td>max</td><td>2990.000000</td><td>0.0</td><td>1.000000</td><td>1.0</td><td>0.0</td><td>10171.250000</td><td>5743.125000</td><td>261.500000</td><td>360.0</td><td>1.0</td><td>2.000000</td><td>1.000000</td></tr></tbody></table>		Loan_ID	Gender	Married	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Loan_Status	count	614.000000	614.0	614.000000	614.0	614.0	614.000000	614.000000	614.000000	614.0	614.0	614.000000	614.000000	mean	1999.785016	0.0	0.653094	1.0	0.0	4617.111564	1419.700733	137.078990	360.0	1.0	1.037459	0.687296	std	567.097644	0.0	0.476373	0.0	0.0	2479.851729	1624.606410	55.847766	0.0	0.0	0.787482	0.463973	min	1002.000000	0.0	0.000000	1.0	0.0	150.000000	0.000000	9.000000	360.0	1.0	0.000000	0.000000	25%	1532.750000	0.0	0.000000	1.0	0.0	2877.500000	0.000000	100.250000	360.0	1.0	0.000000	0.000000	50%	1993.500000	0.0	1.000000	1.0	0.0	3812.500000	1188.500000	125.000000	360.0	1.0	1.000000	1.000000	75%	2476.750000	0.0	1.000000	1.0	0.0	5795.000000	2297.250000	164.750000	360.0	1.0	2.000000	1.000000	max	2990.000000	0.0	1.000000	1.0	0.0	10171.250000	5743.125000	261.500000	360.0	1.0	2.000000	1.000000
	Loan_ID	Gender	Married	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Loan_Status																																																																																																										
count	614.000000	614.0	614.000000	614.0	614.0	614.000000	614.000000	614.000000	614.0	614.0	614.000000	614.000000																																																																																																										
mean	1999.785016	0.0	0.653094	1.0	0.0	4617.111564	1419.700733	137.078990	360.0	1.0	1.037459	0.687296																																																																																																										
std	567.097644	0.0	0.476373	0.0	0.0	2479.851729	1624.606410	55.847766	0.0	0.0	0.787482	0.463973																																																																																																										
min	1002.000000	0.0	0.000000	1.0	0.0	150.000000	0.000000	9.000000	360.0	1.0	0.000000	0.000000																																																																																																										
25%	1532.750000	0.0	0.000000	1.0	0.0	2877.500000	0.000000	100.250000	360.0	1.0	0.000000	0.000000																																																																																																										
50%	1993.500000	0.0	1.000000	1.0	0.0	3812.500000	1188.500000	125.000000	360.0	1.0	1.000000	1.000000																																																																																																										
75%	2476.750000	0.0	1.000000	1.0	0.0	5795.000000	2297.250000	164.750000	360.0	1.0	2.000000	1.000000																																																																																																										
max	2990.000000	0.0	1.000000	1.0	0.0	10171.250000	5743.125000	261.500000	360.0	1.0	2.000000	1.000000																																																																																																										
Univariate Analysis	<div><div><div>KUMPLU1L{"dx1s": d}, dx=dx, colut="KUM_COLUT", "KUM_KWS")</div></div><div></div></div>																																																																																																																					

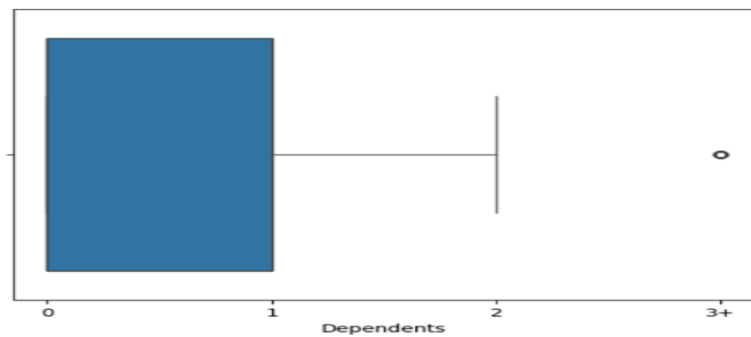
## Bivariate Analysis

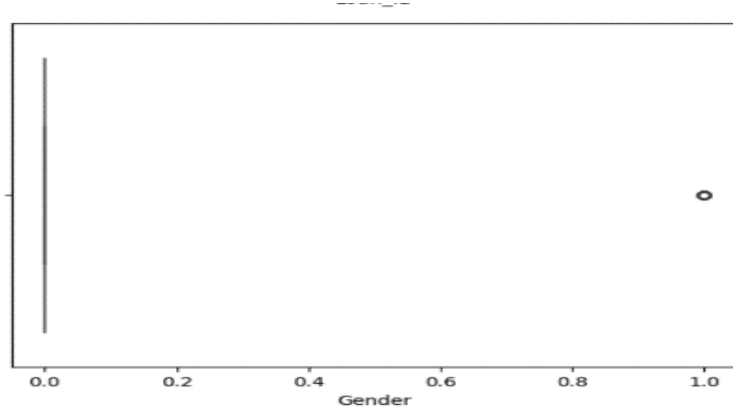


## Multivariate Analysis



## Outliers and Anomalies





## Data Preprocessing Code Screenshots

### Loading Data

```

reporting the dataset which is in the csv file
import pandas as pd
data = pd.read_csv('content/loan_prediction.csv')
data

```

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Loan_Status
0	LP001002	Male	No	0	Graduate	No	5849	0.0	NaN	360.0	1.0	Urban	Y
1	LP001003	Male	Yes	1	Graduate	No	4583	1508.0	128.0	360.0	1.0	Rural	N
2	LP001005	Male	Yes	0	Graduate	Yes	3000	0.0	66.0	360.0	1.0	Urban	Y
3	LP001006	Male	Yes	0	Not Graduate	No	2583	2358.0	120.0	360.0	1.0	Urban	Y
4	LP001008	Male	No	0	Graduate	No	6000	0.0	141.0	360.0	1.0	Urban	Y
...	...	...	...	...	...	...	...	...	...	...	...	...	...
609	LP002978	Female	No	0	Graduate	No	2900	0.0	71.0	360.0	1.0	Rural	Y
610	LP002979	Male	Yes	3+	Graduate	No	4106	0.0	40.0	180.0	1.0	Rural	Y
611	LP002983	Male	Yes	1	Graduate	No	8072	240.0	253.0	360.0	1.0	Urban	Y
612	LP002984	Male	Yes	2	Graduate	No	7583	0.0	187.0	360.0	1.0	Urban	Y
613	LP002990	Female	No	0	Graduate	Yes	4583	0.0	133.0	360.0	0.0	Semiurban	N

614 rows x 13 columns

### Handling Missing Data

```

#Code for handling with Null values
data['Gender'].fillna(data['Gender'].mode()[0],inplace=True)

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

[10] data['Dependents'].fillna(data['Dependents'].mode()[0],inplace=True)

[11] data['Self_Employed'].fillna(data['Self_Employed'].mode()[0],inplace=True)

[12] data['LoanAmount'].fillna(data['LoanAmount'].mode()[0],inplace=True)

[13] data['Loan_Amount_Term'].fillna(data['Loan_Amount_Term'].mode()[0],inplace=True)

[14] data['Credit_History'].fillna(data['Credit_History'].mode()[0],inplace=True)

```

### Data Transformation

```

#Handling categorical values
data['Gender']=data['Gender'].map({'Female':1,'Male':0})

[17] data['Property_Area']=data['Property_Area'].map({'Urban':2,'Semiurban':1,'Rural':0})

[18] data['Married']=data['Married'].map({'Yes':1,'No':0})

[19] data['Education']=data['Education'].map({'Graduate':1,'Not Graduate':0})

[20] data['Loan_Status']=data['Loan_Status'].map({'Y':1,'N':0})

```

Feature Engineering	Attached the codes in final submission.
Save Processed Data	-