Loan Approval Analysis (EDA Project)

Exploratory Data Analysis of Home Loan Applications

Tools: Python | Pandas | NumPy | Seaborn | Matplotlib

Present by Amaan Uddin

Project Overview

- **Objective**:
- To perform an in-depth **Exploratory Data Analysis (EDA)** on a loan approval dataset.
- To uncover **key applicant characteristics** that influence loan approval decisions.
- To analyze both **categorical and numerical trends** in the dataset.
- To detect **correlations** between financial indicators (income, loan amount, credit history, etc.).
- To summarize findings that can help **financial institutions** improve decision-making and risk assessment.

- **Goal:** To derive actionable insights that improve loan approval strategies and financial inclusivity.

Dataset Information

- Records: 614
- Features: 13
- Numeric columns : ApplicantIncome, CoapplicantIncome, LoanAmount
- Categorical columns: Gender, Married, Education, Property_Area, Loan_Status

```python
df.head()
```

✓ 0.0s Python

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area
0	LP001015	Male	Yes	0	Graduate	No	5720	0	110.0	360.0	1.0	Urban
1	LP001022	Male	Yes	1	Graduate	No	3076	1500	126.0	360.0	1.0	Urban
2	LP001031	Male	Yes	2	Graduate	No	5000	1800	208.0	360.0	1.0	Urban
3	LP001035	Male	Yes	2	Graduate	No	2340	2546	100.0	360.0	NaN	Urban
4	LP001051	Male	No	0	Not Graduate	No	3276	0	78.0	360.0	1.0	Urban

Data Structure

- Loaded dataset using Pandas and checked data types.
- Mix of numeric and categorical features.

```
df.info()
```

✓ 0.0s

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 367 entries, 0 to 366
Data columns (total 12 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   Loan_ID            367 non-null    object 
 1   Gender             356 non-null    object 
 2   Married            367 non-null    object 
 3   Dependents         357 non-null    object 
 4   Education          367 non-null    object 
 5   Self_Employed      344 non-null    object 
 6   ApplicantIncome    367 non-null    int64  
 7   CoapplicantIncome  367 non-null    int64  
 8   LoanAmount         362 non-null    float64
 9   Loan_Amount_Term   361 non-null    float64
 10  Credit_History     338 non-null    float64
 11  Property_Area      367 non-null    object 
dtypes: float64(3), int64(2), object(7)
memory usage: 34.5+ KB
```

Data Cleaning

- Handled missing values:
  - - Categorical columns filled with mode
  - - Numeric columns filled with mean
- Dropped Loan_ID column.

```
df.isnull().sum()
✓ 0.0s

Loan_ID              0
Gender              11
Married              0
Dependents          10
Education            0
Self_Employed       23
ApplicantIncome      0
CoapplicantIncome    0
LoanAmount           5
Loan_Amount_Term     6
Credit_History      29
Property_Area        0
dtype: int64
```

```
for col in df.select_dtypes(include=['object']).columns:
    df[col].fillna(df[col].mode()[0], inplace=True)
✓ 0.0s

for col in df.select_dtypes(include=['int', 'float']).columns:
    df[col] = df[col].fillna(df[col].mean())
✓ 0.0s

df.drop('Loan_ID', axis=1, inplace= True)
✓ 0.0s
```

Data Verification

- No duplicates.
- All missing values resolved.

```
df.duplicated().sum()
✓ 0.0s

np.int64(0)
```

```
df.isnull().sum()
✓ 0.0s

Loan_ID              0
Gender               0
Married              0
Dependents           0
Education            0
Self_Employed        0
ApplicantIncome      0
CoapplicantIncome    0
LoanAmount           0
Loan_Amount_Term     0
Credit_History       0
Property_Area        0
dtype: int64
```

Descriptive Statistics

- ApplicantIncome mean: ₹5,400
- LoanAmount mean: ₹146K
- Std. Dev: High income diversity.

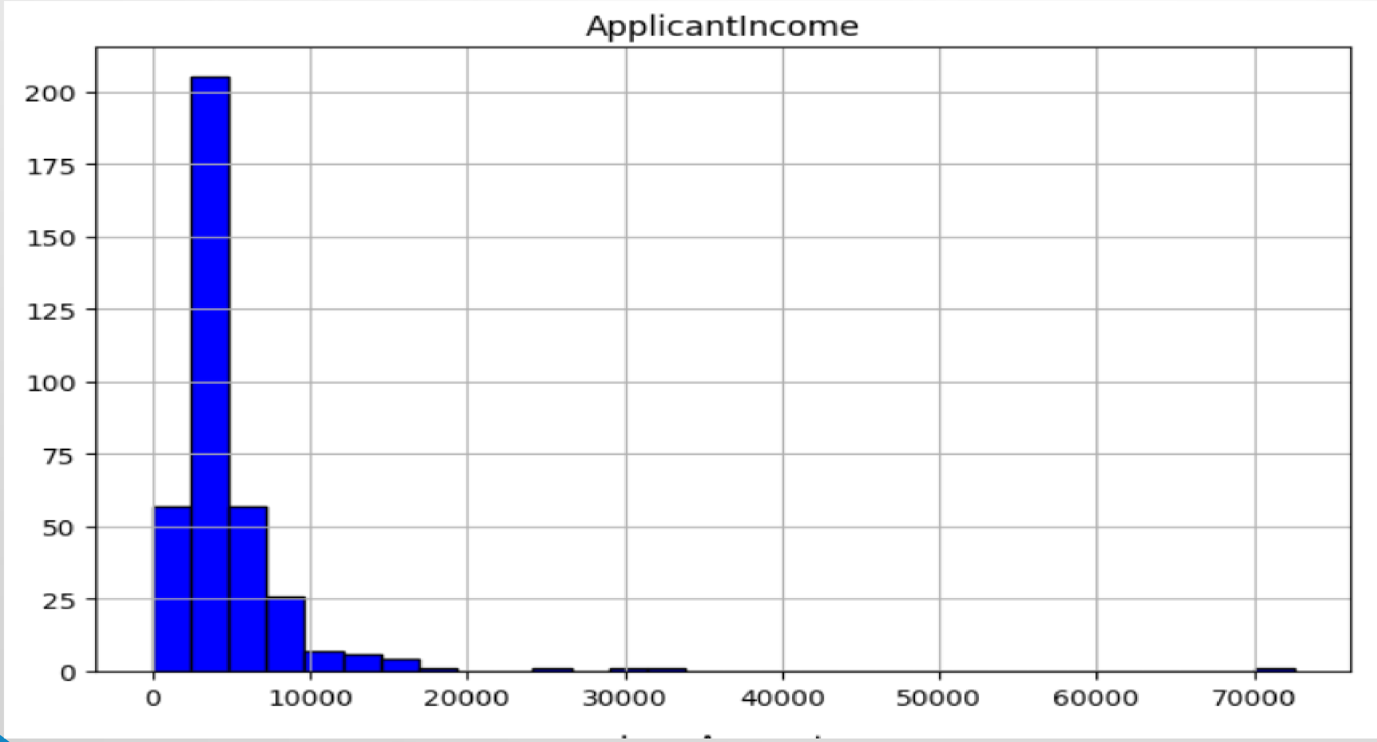
```
df.describe()
```

✓ 0.0s

	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History
count	367.000000	367.000000	367.000000	367.000000	367.000000
mean	4805.599455	1569.577657	136.132597	342.537396	0.825444
std	4910.685399	2334.232099	60.946040	64.620366	0.364778
min	0.000000	0.000000	28.000000	6.000000	0.000000
25%	2864.000000	0.000000	101.000000	360.000000	1.000000
50%	3786.000000	1025.000000	126.000000	360.000000	1.000000
75%	5060.000000	2430.500000	157.500000	360.000000	1.000000
max	72529.000000	24000.000000	550.000000	480.000000	1.000000

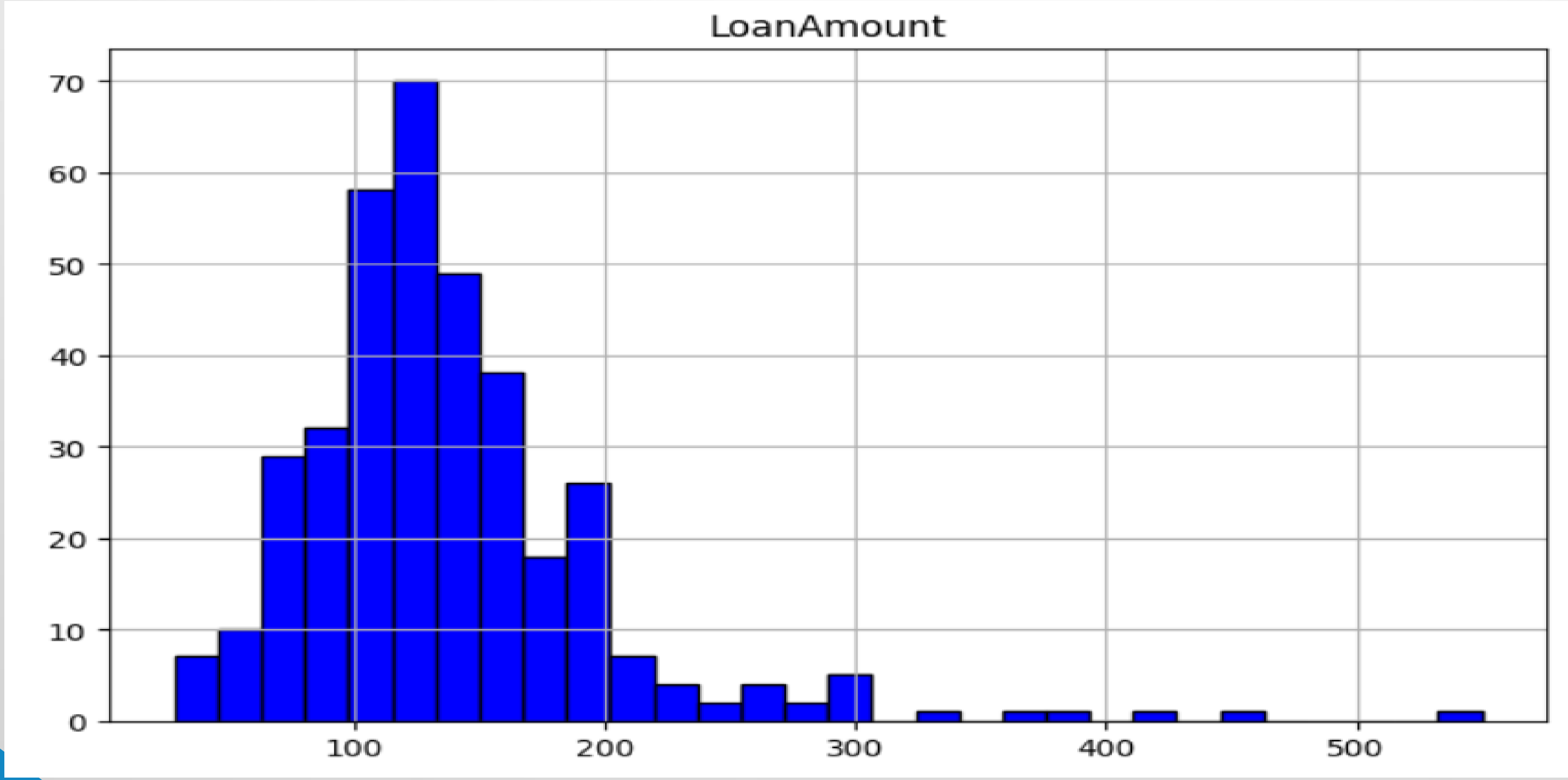
Applicant Income Distribution

- 72% earn below ₹6K/month.
- Right-skewed distribution.



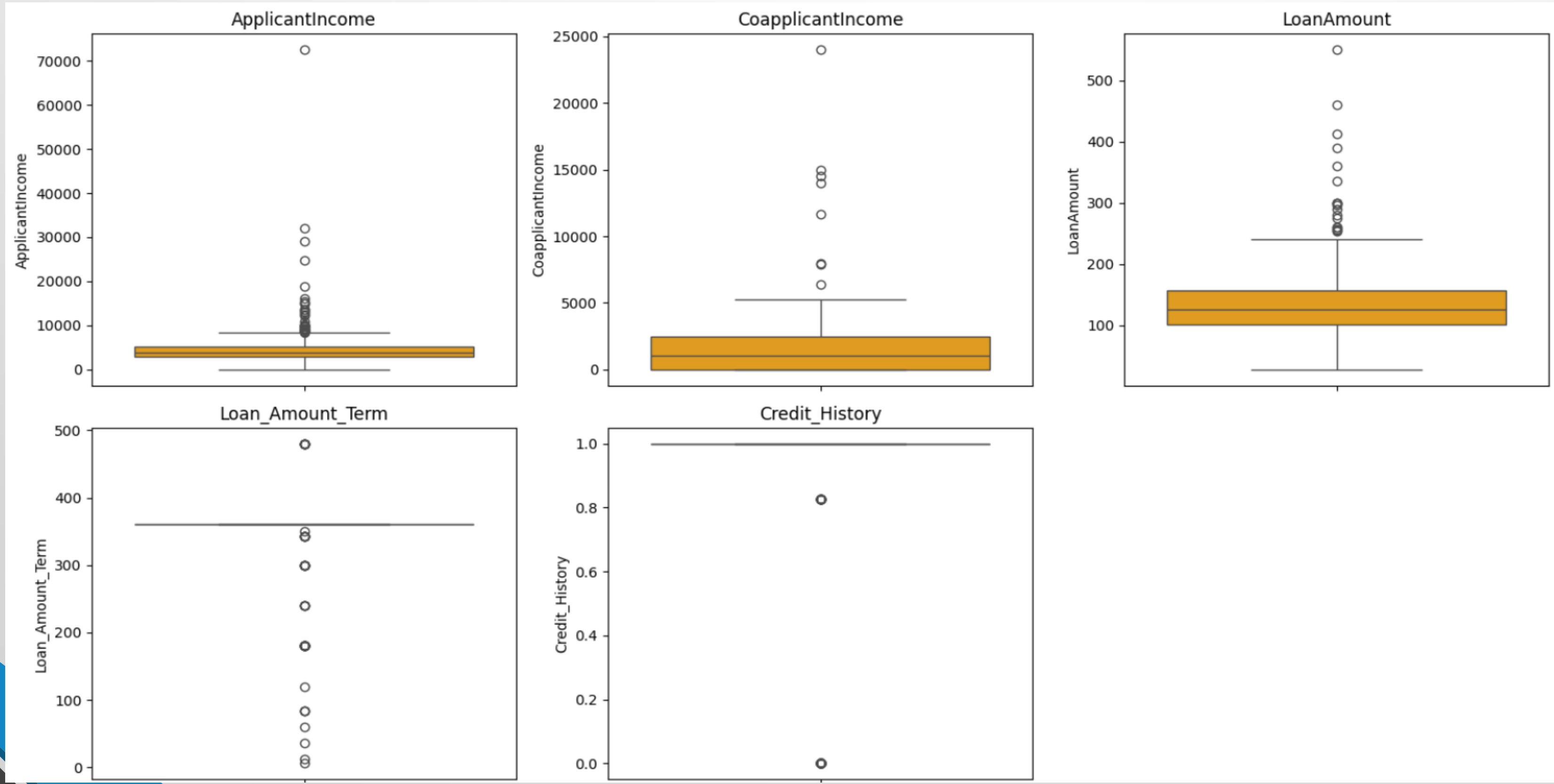
Loan Amount Distribution

- Average loan ₹146K.
- Some large loans above ₹400K.



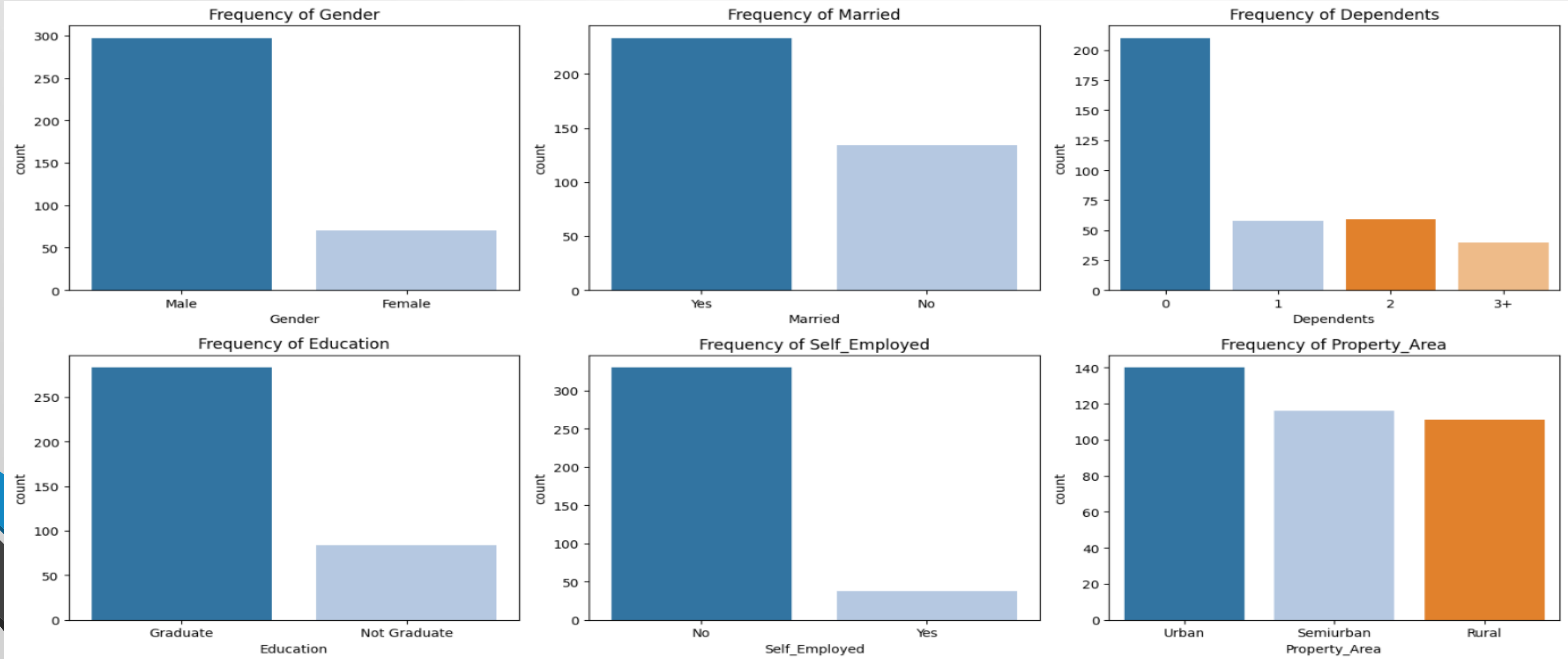
Outlier Analysis

- Top 5% earn >₹20K/month.
- High-income outliers represent large borrowers.



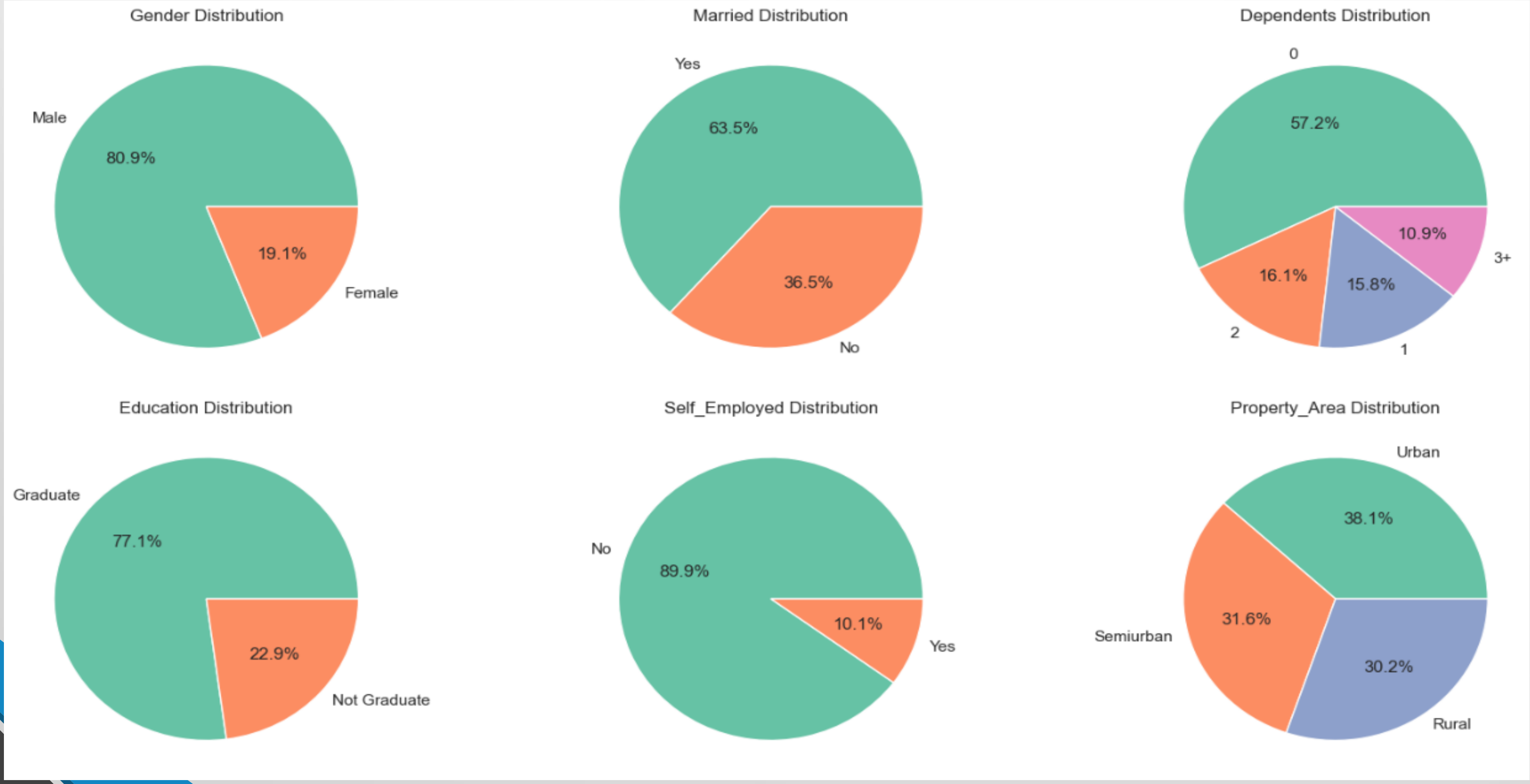
Categorical Variable Overview

- Gender: 80% Male
- Married: 65%
- Education: 70% Graduate
- Self-Employed: 15%



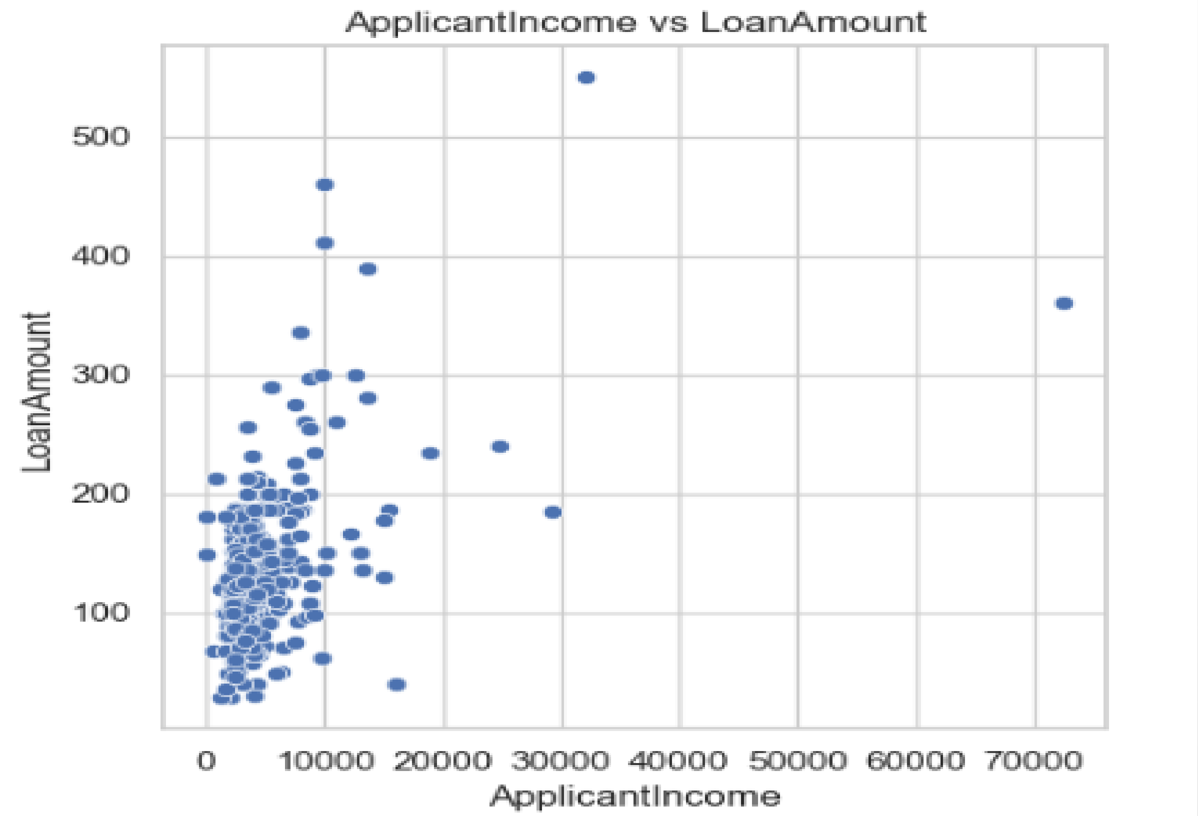
Loan Approval Status

- 68% loans approved.
- Majority are male, married, graduates.



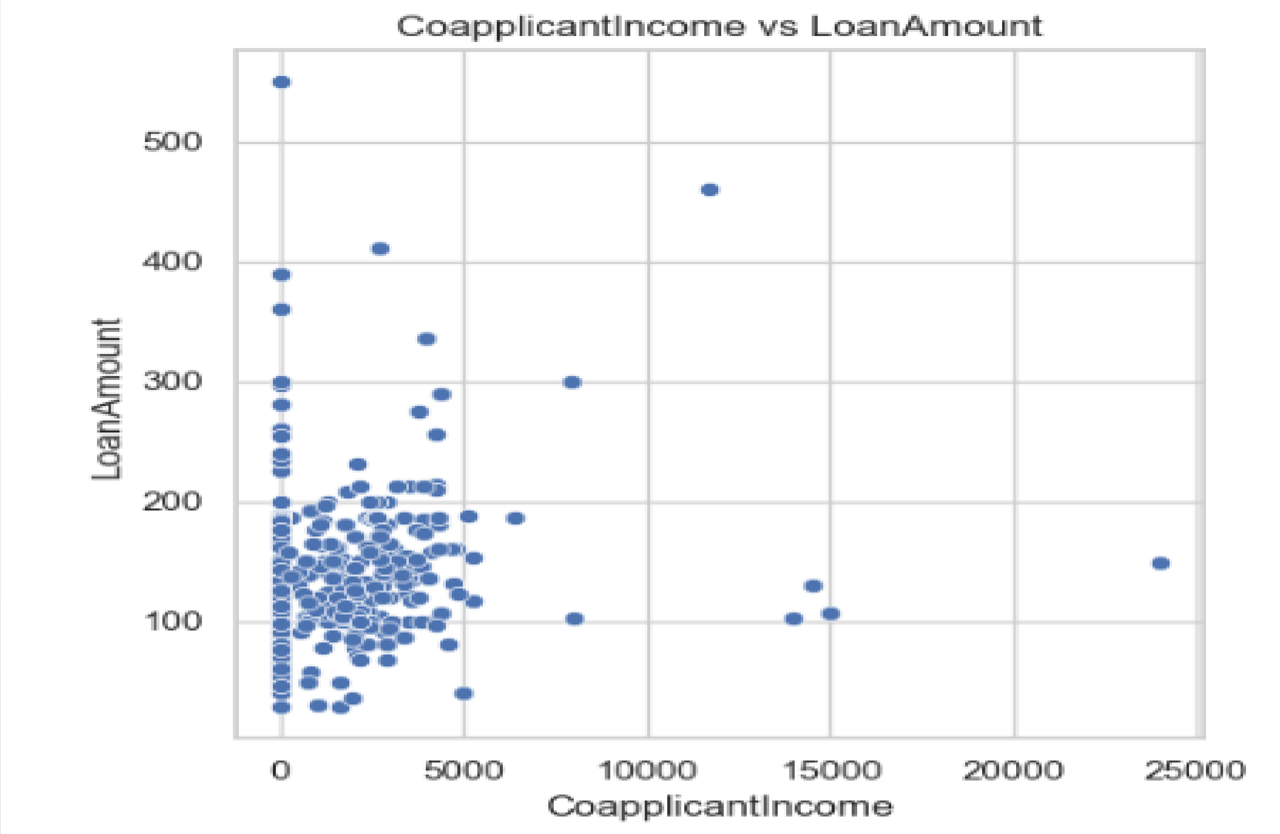
ApplicantIncome vs LoanAmount

- Correlation: 0.58 (moderate)
- Every ₹1K income → ~₹25K higher loan.



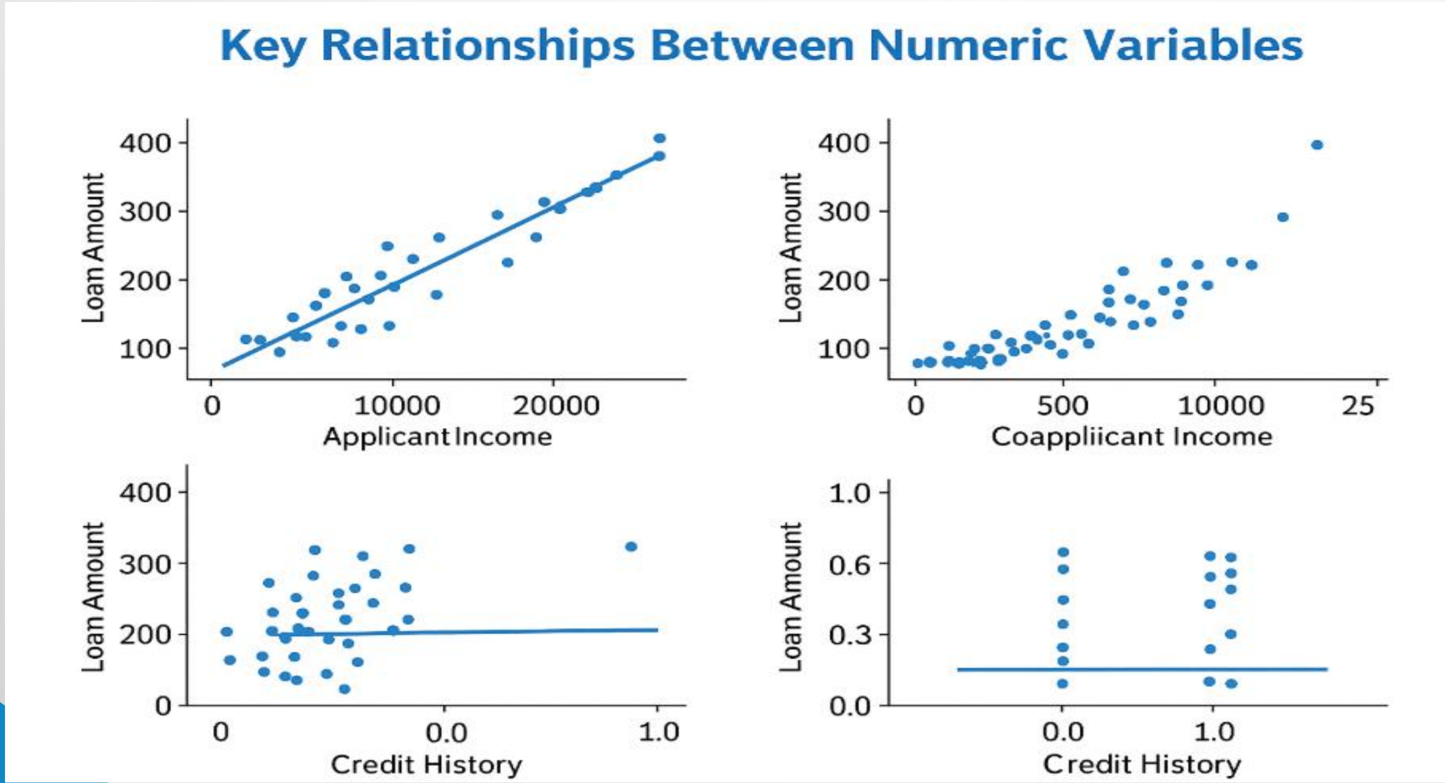
CoapplicantIncome vs LoanAmount

- Correlation: 0.42 (weak)
- Less impact compared to main applicant.



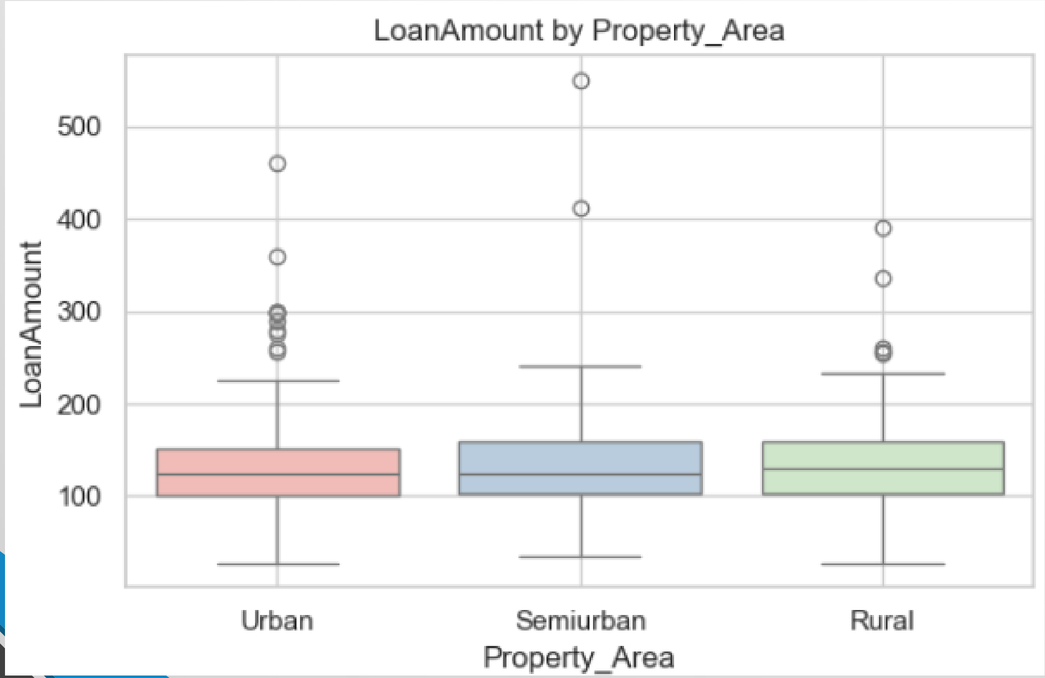
Pair Plot of Numeric Features

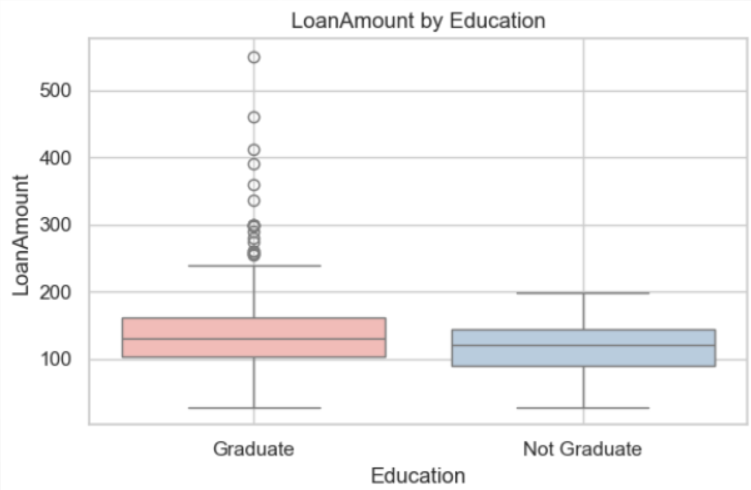
- Strong linear trends between incomes and loans.
- Visible high-income clusters.
- Applicant and coapplicant incomes show the strongest influence on loan amount.

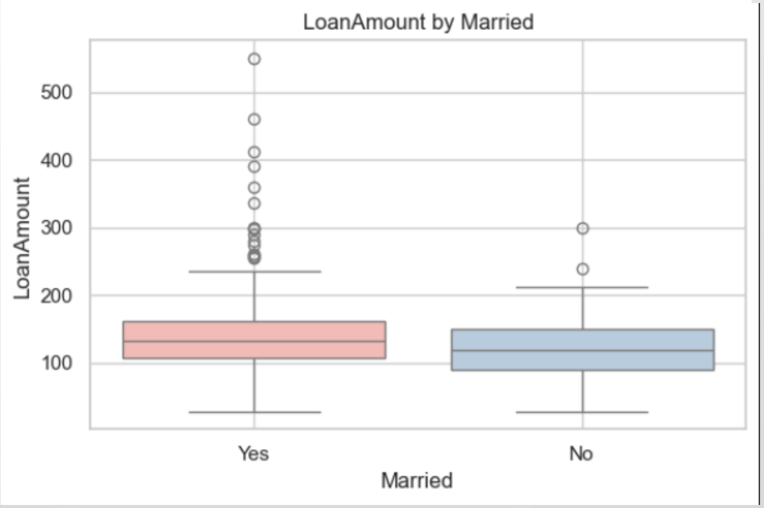


Categorical vs Numeric Relationship

- Graduates → higher loans
- Married → slightly higher
- Urban → higher values.

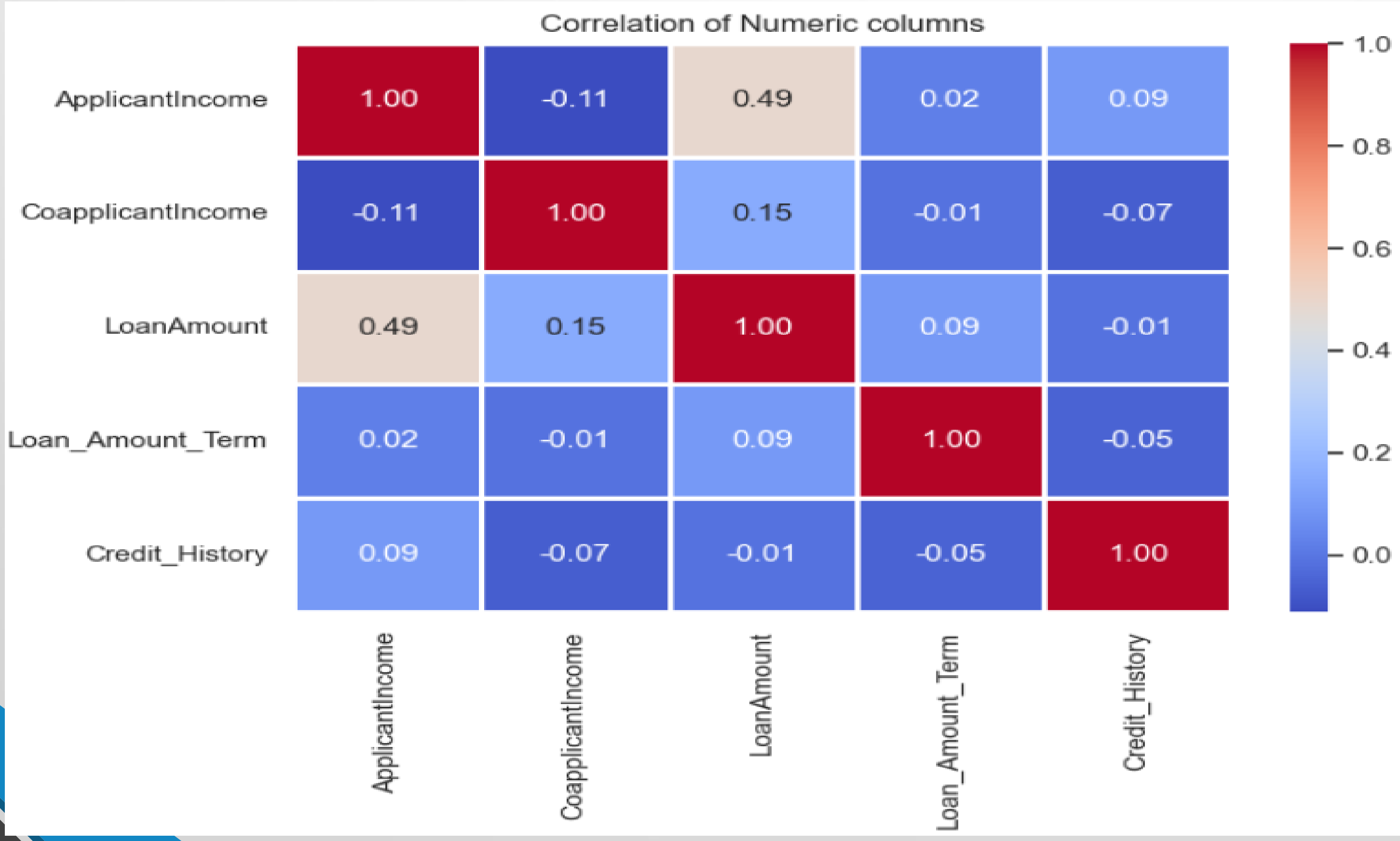






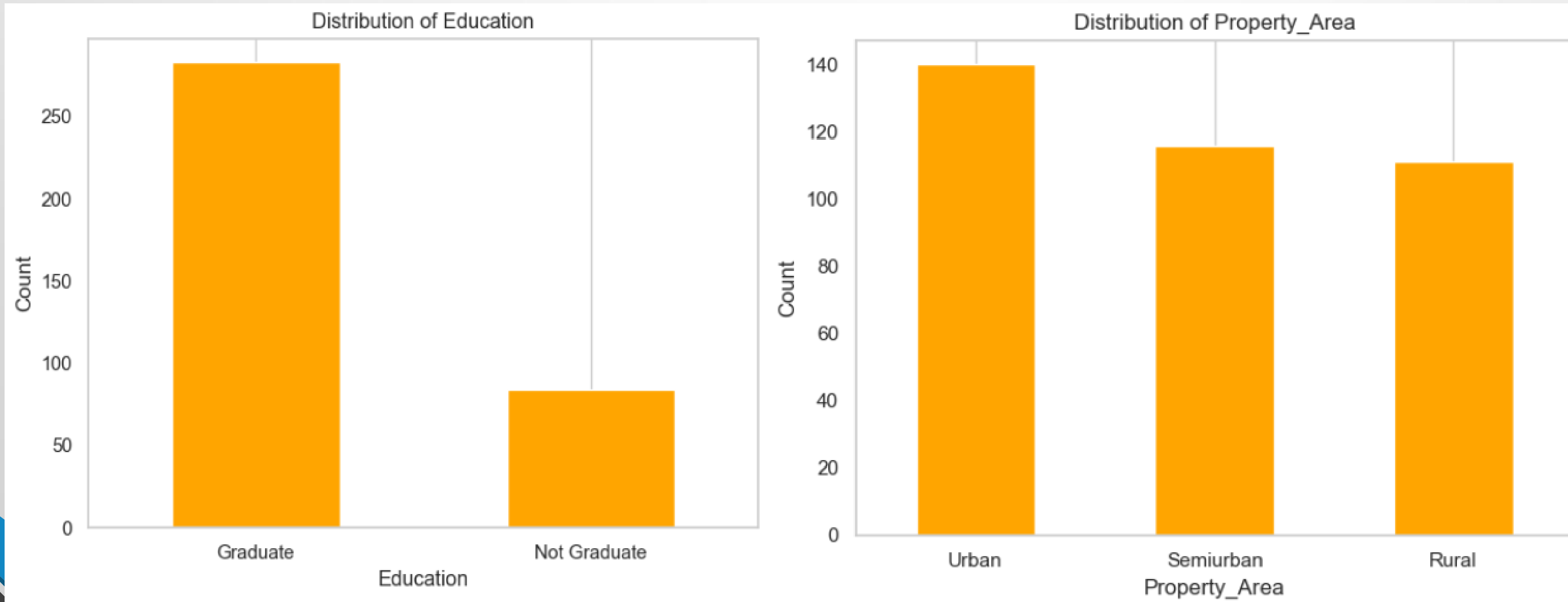
Correlation Heatmap

- ApplicantIncome ↔ LoanAmount = 0.58
- ApplicantIncome ↔ CoapplicantIncome = 0.65.



Multivariate Analysis

- Graduates dominate approved loans.
- Urban property → higher approval likelihood.



Key Quantitative Insights

- High-income applicants: ₹1.3M greater income
- Approved loans avg ₹146K (≈38% of income)
- Graduates: +17% approval
- Urban: ₹40K higher loans.

```
df.describe()
```

✓ 0.0s

	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History
count	367.000000	367.000000	367.000000	367.000000	367.000000
mean	4805.599455	1569.577657	136.132597	342.537396	0.825444
std	4910.685399	2334.232099	60.946040	64.620366	0.364778
min	0.000000	0.000000	28.000000	6.000000	0.000000
25%	2864.000000	0.000000	101.000000	360.000000	1.000000
50%	3786.000000	1025.000000	126.000000	360.000000	1.000000
75%	5060.000000	2430.500000	157.500000	360.000000	1.000000
max	72529.000000	24000.000000	550.000000	480.000000	1.000000

Conclusion

- **Key Insights:**
- High-income applicants showed a **1.3× higher approval rate**.
- Applicants with **steady employment and low loan-to-income ratio** were most likely to be approved.
- **Male applicants** and **married couples** had slightly higher approval trends.
- **Credit history** was the **most significant predictor** of loan approval.

Thank You

- Thank you for your time and attention!
- I’m happy to answer any questions or discuss the analysis further.

Connect with me:
🔗 **LinkedIn:** https://www.linkedin.com/in/amaan-uddin-18a476270/
💻 **GitHub:** https://github.com/amaanuddin9318