

```
In [1]: import pandas as pd
import numpy as np
```

```
In [2]: import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [3]: #from sklearn.preprocessing import OrdinalEncoder
from sklearn.preprocessing import LabelEncoder
#from sklearn.preprocessing import OneHotEncoder
```

```
In [4]: file_path='Loan_Sanction_DataSet.csv'
df=pd.DataFrame(pd.read_csv(file_path)).drop('Loan_ID',axis=1)
```

```
In [5]: df.head()
```

	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	Coapplicant
0	Male	No	0	Graduate	No	5849	
1	Male	Yes	1	Graduate	No	4583	
2	Male	Yes	0	Graduate	Yes	3000	
3	Male	Yes	0	Not Graduate	No	2583	
4	Male	No	0	Graduate	No	6000	

◀ ▶

```
In [6]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 614 entries, 0 to 613
Data columns (total 12 columns):
 #   Column            Non-Null Count  Dtype  
--- 
 0   Gender             601 non-null    object  
 1   Married            611 non-null    object  
 2   Dependents         599 non-null    object  
 3   Education          614 non-null    object  
 4   Self_Employed      582 non-null    object  
 5   ApplicantIncome    614 non-null    int64  
 6   CoapplicantIncome  614 non-null    float64 
 7   LoanAmount         592 non-null    float64 
 8   Loan_Amount_Term  600 non-null    float64 
 9   Credit_History     564 non-null    float64 
 10  Property_Area      614 non-null    object  
 11  Loan_Status        614 non-null    object  
dtypes: float64(4), int64(1), object(7)
memory usage: 57.7+ KB
```

```
In [7]: df.columns
```

```
Out[7]: Index(['Gender', 'Married', 'Dependents', 'Education', 'Self_Employed',
   'ApplicantIncome', 'CoapplicantIncome', 'LoanAmount',
   'Loan_Amount_Term', 'Credit_History', 'Property_Area', 'Loan_Status'],
  dtype='object')
```

```
In [8]: numerical_variables = df.select_dtypes(include=['int', 'float']).columns
categorical_variables=df.select_dtypes(include=['object']).columns
```

```
In [9]: print('Numerical Data are :',numerical_variables)
print('Categorical Data are :',categorical_variables)
```

```
Numerical Data are : Index(['ApplicantIncome', 'CoapplicantIncome', 'LoanAmount',
   'Loan_Amount_Term', 'Credit_History'],
  dtype='object')
Categorical Data are : Index(['Gender', 'Married', 'Dependents', 'Education', 'Self_Employed',
   'Property_Area', 'Loan_Status'],
  dtype='object')
```

```
In [10]: # Create an instance of LabelEncoder
encoder = LabelEncoder()
#df[categorical_columns]=encoder.fit_transform(categorical_columns)

for i in categorical_variables:
    if(df[i].dtype =='object'):
        #df[i]=encoder.fit_transform(df[i])
        pass
    else:
        print(i +" is already Numerial Type")
        pass
```

```
In [11]: df.head()
```

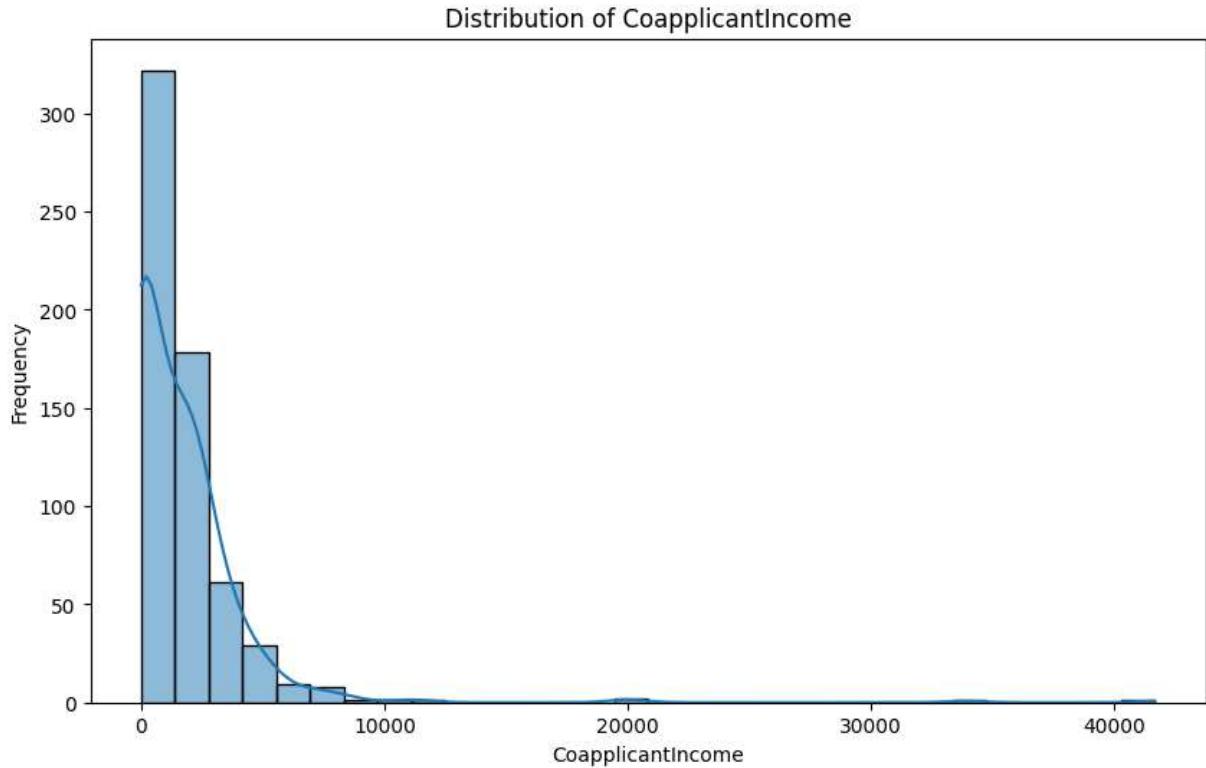
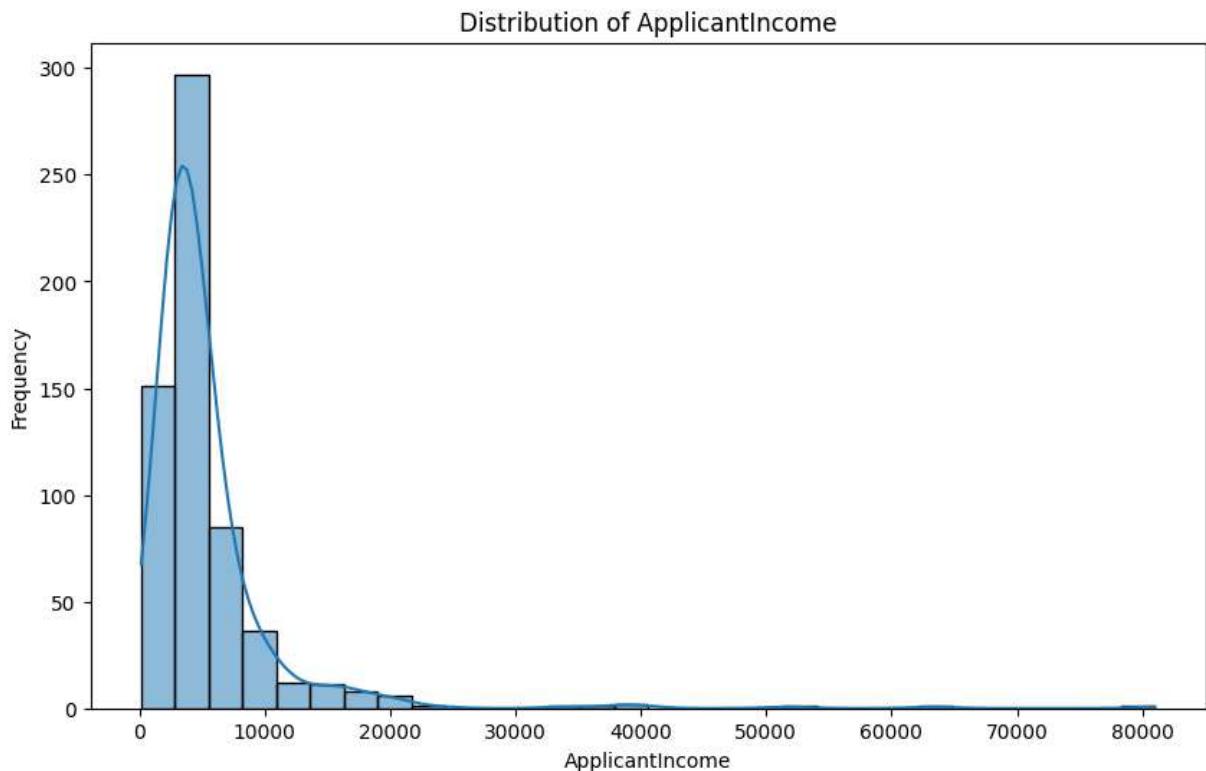
	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	Coapplicant
0	Male	No	0	Graduate	No	5849	
1	Male	Yes	1	Graduate	No	4583	
2	Male	Yes	0	Graduate	Yes	3000	
3	Male	Yes	0	Not Graduate	No	2583	
4	Male	No	0	Graduate	No	6000	



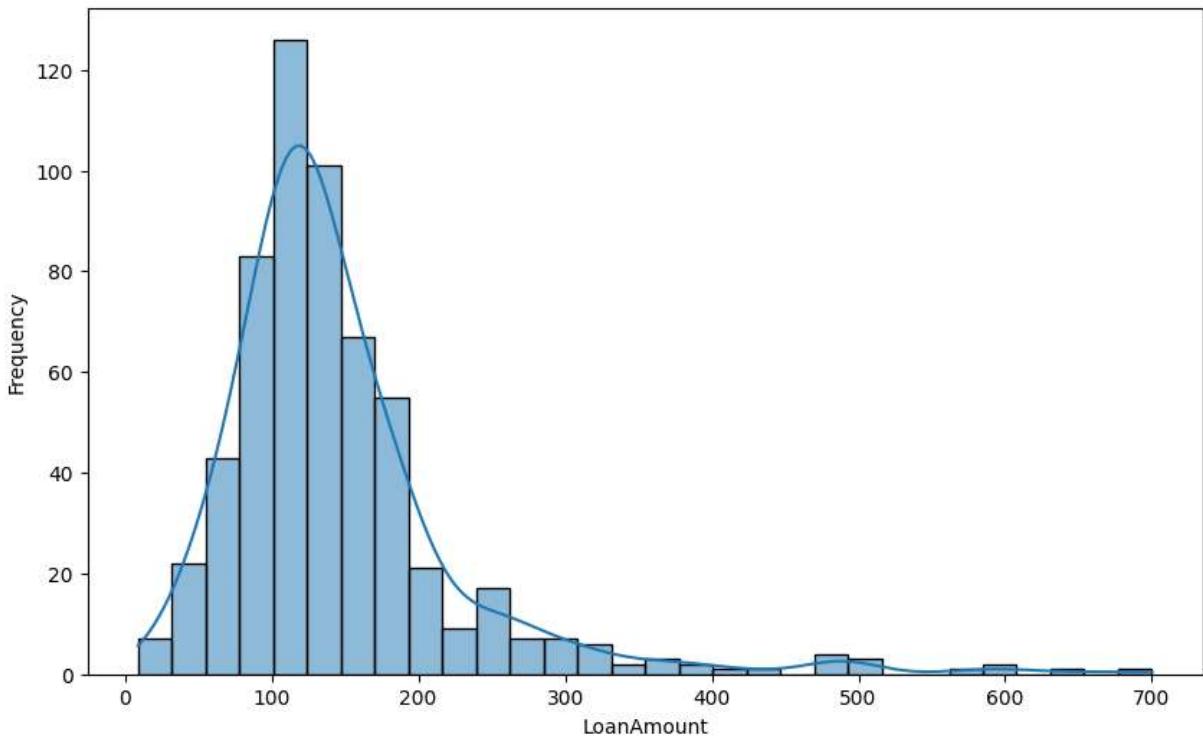
```
In [12]: #Univariate analysis
graph_types=['histPlot-Numerical','countPlot-Categorical','Box Plot-Numerical','Grou
```

```
In [13]: # Univariate analysis for each numerical variable
for variable in numerical_variables:
    plt.figure(figsize=(10, 6))
    sns.histplot(df[variable].dropna(), bins=30, kde=True)
```

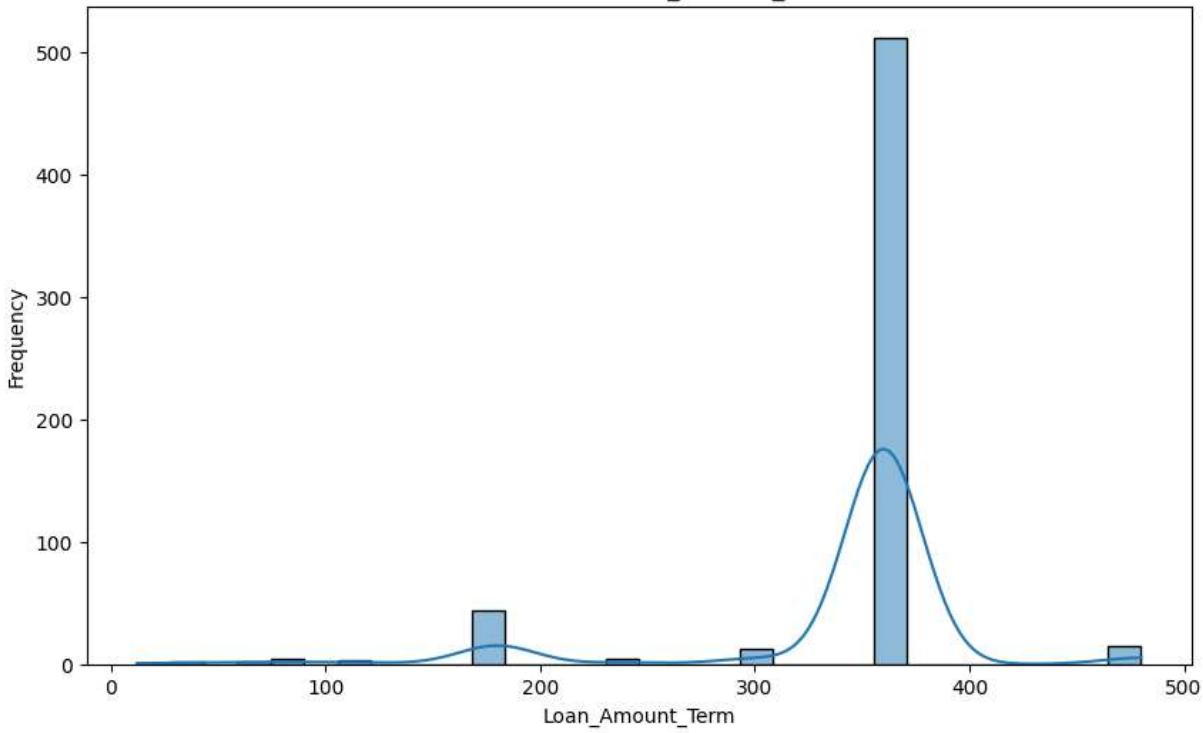
```
plt.title(f'Distribution of {variable}')
plt.xlabel(variable)
plt.ylabel('Frequency')
plt.show()
```

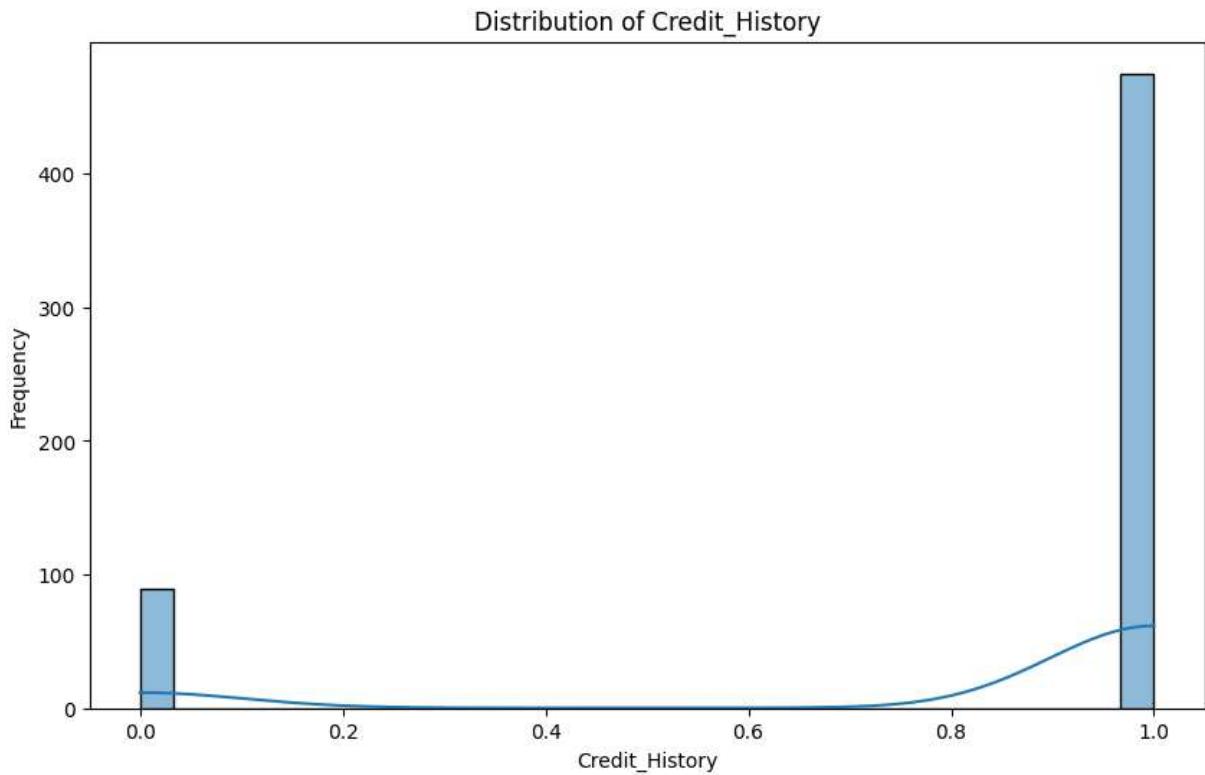


Distribution of LoanAmount

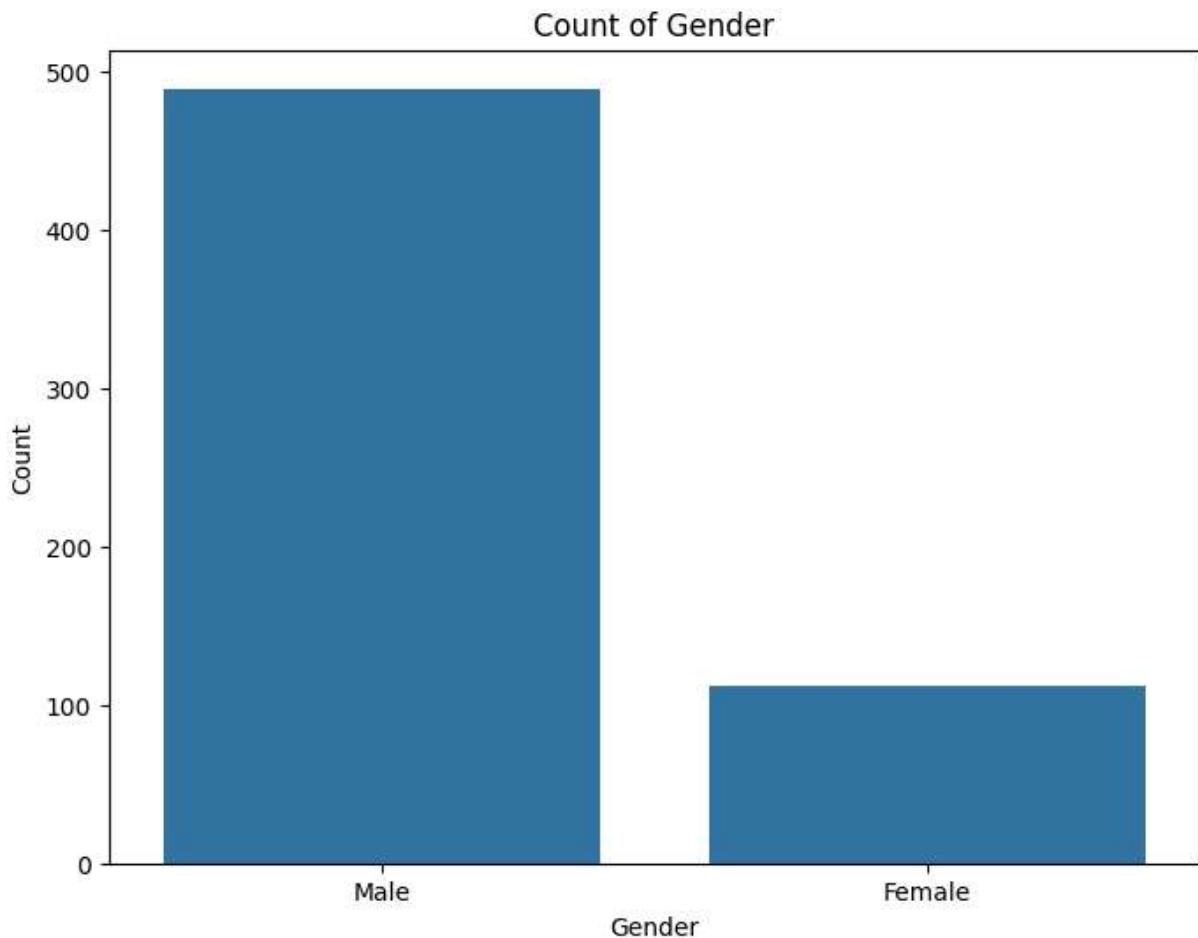


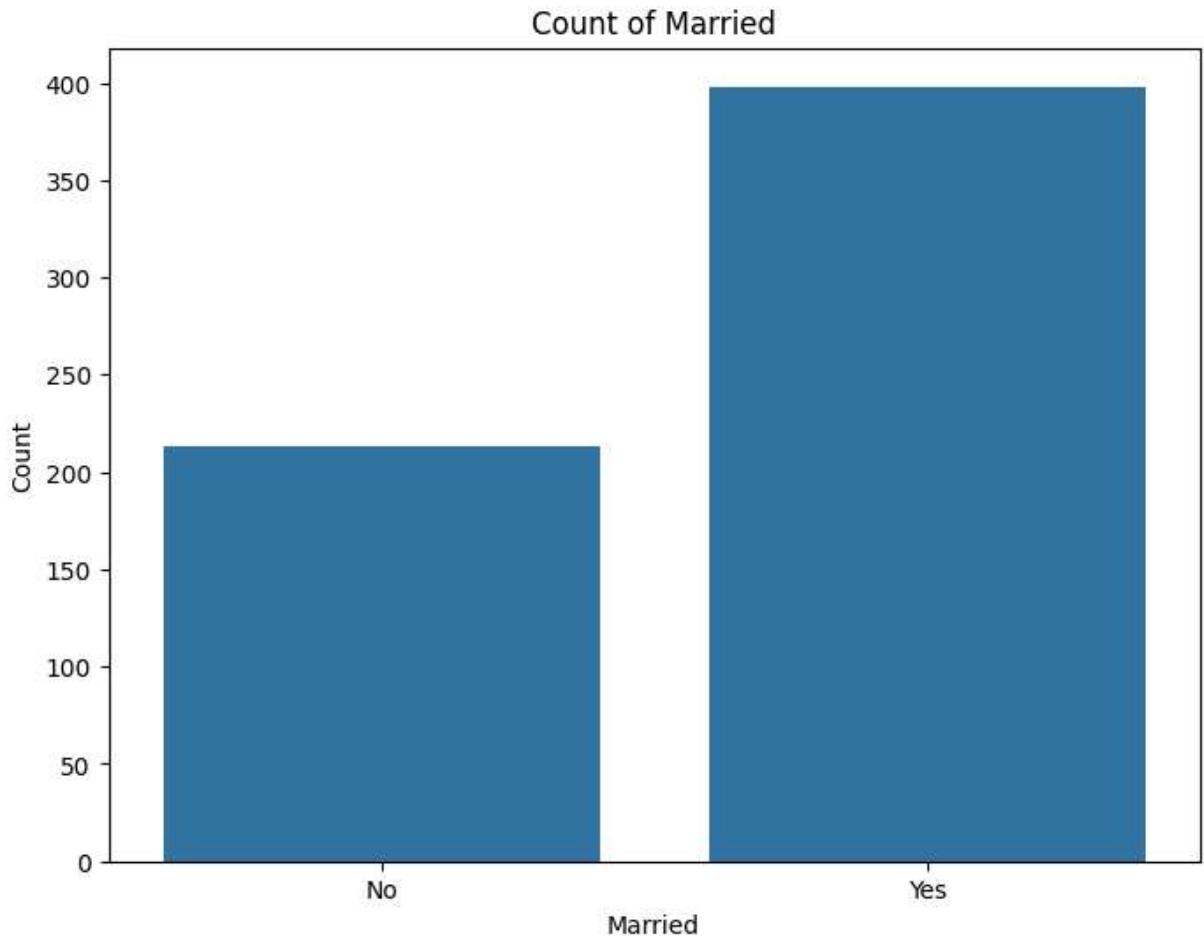
Distribution of Loan_Amount_Term

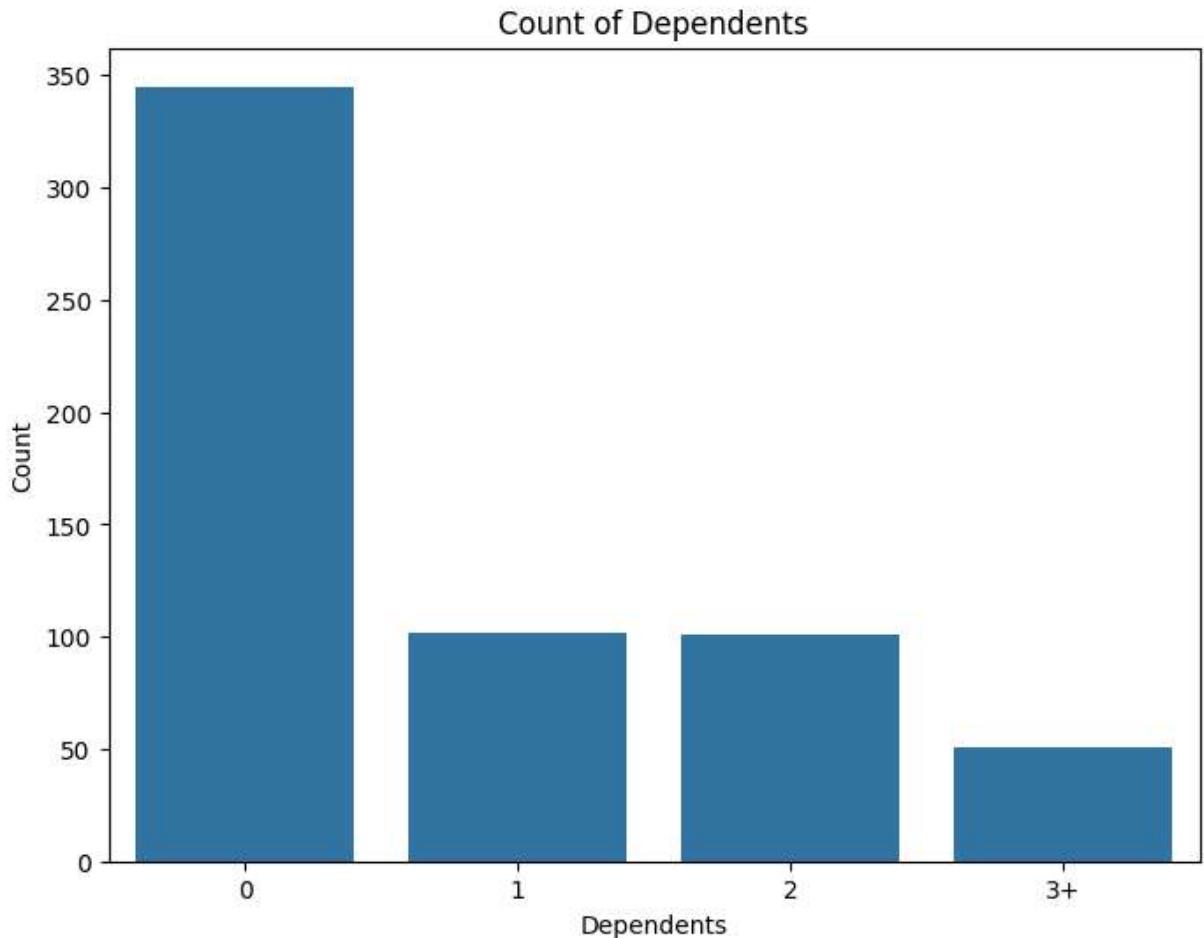


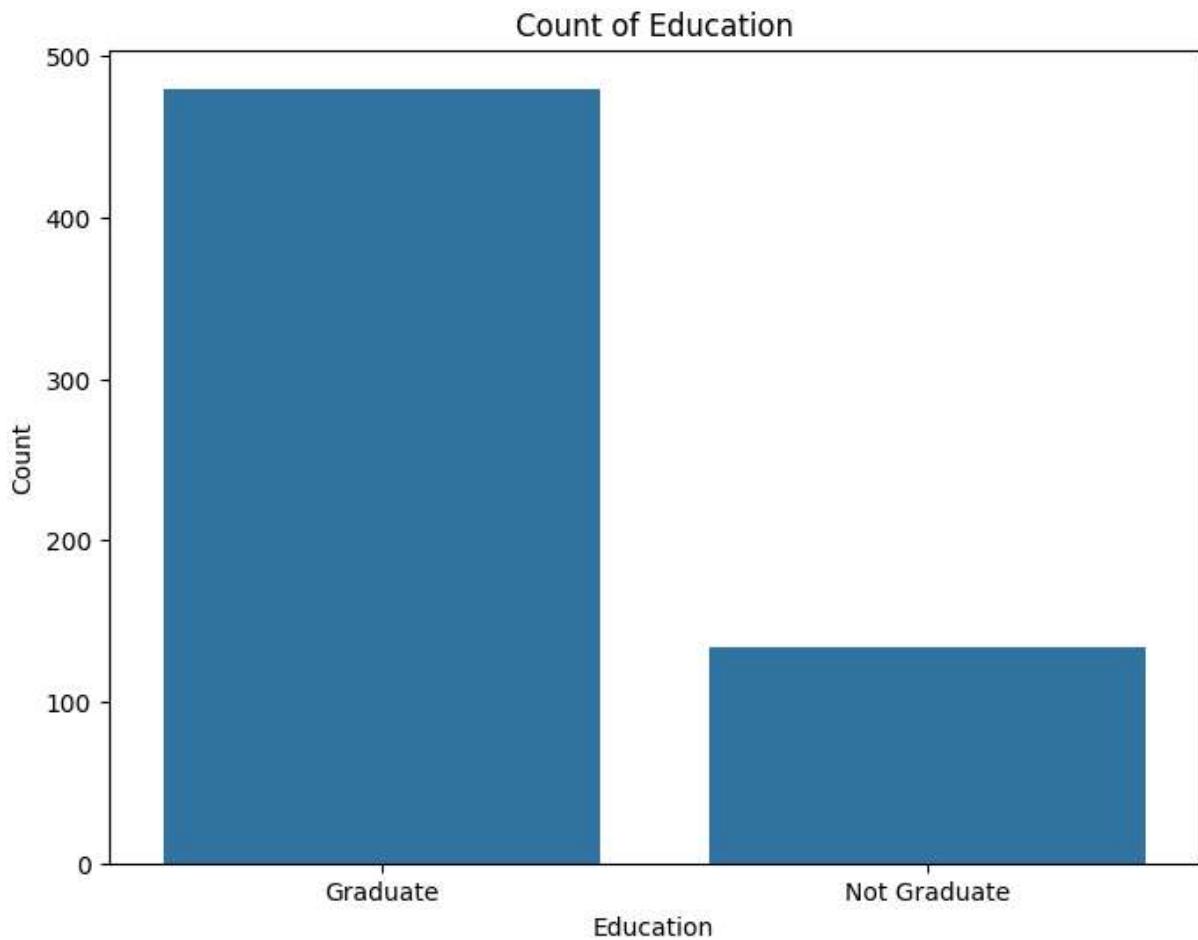


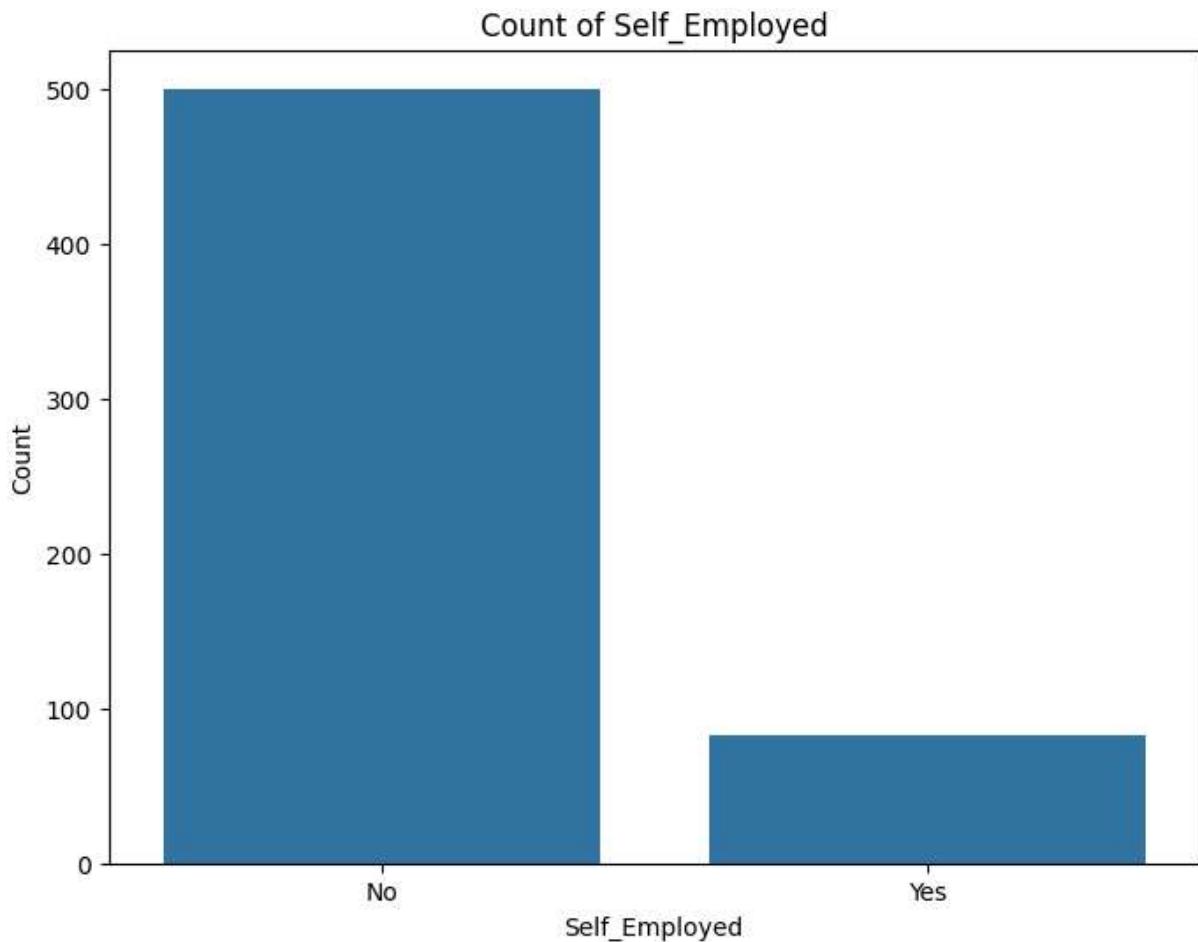
```
In [14]: # Univariate analysis for each categorical variable
for variable in categorical_variables:
    plt.figure(figsize=(8, 6))
    sns.countplot(x=variable, data=df)
    plt.title(f'Count of {variable}')
    plt.xlabel(variable)
    plt.ylabel('Count')
    plt.show()
```

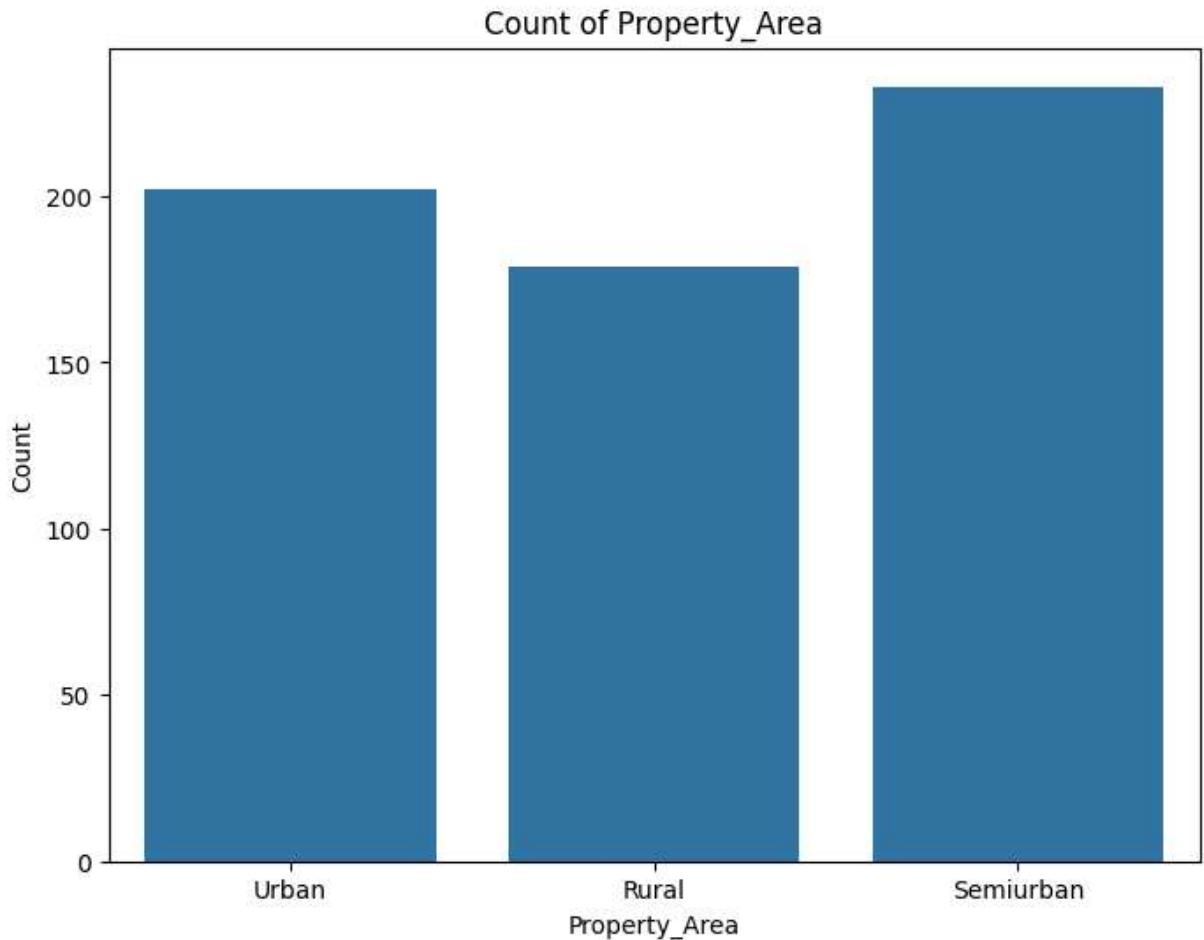


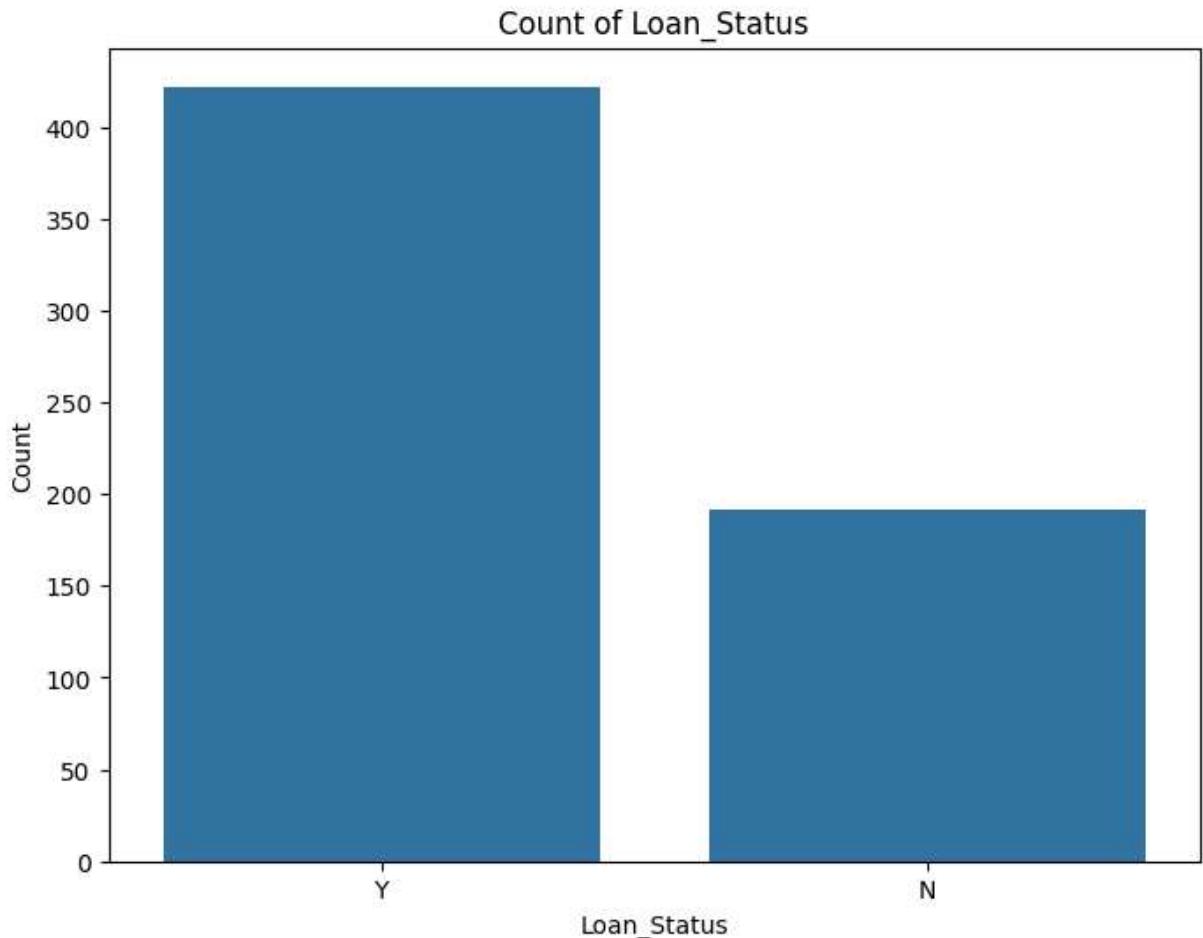






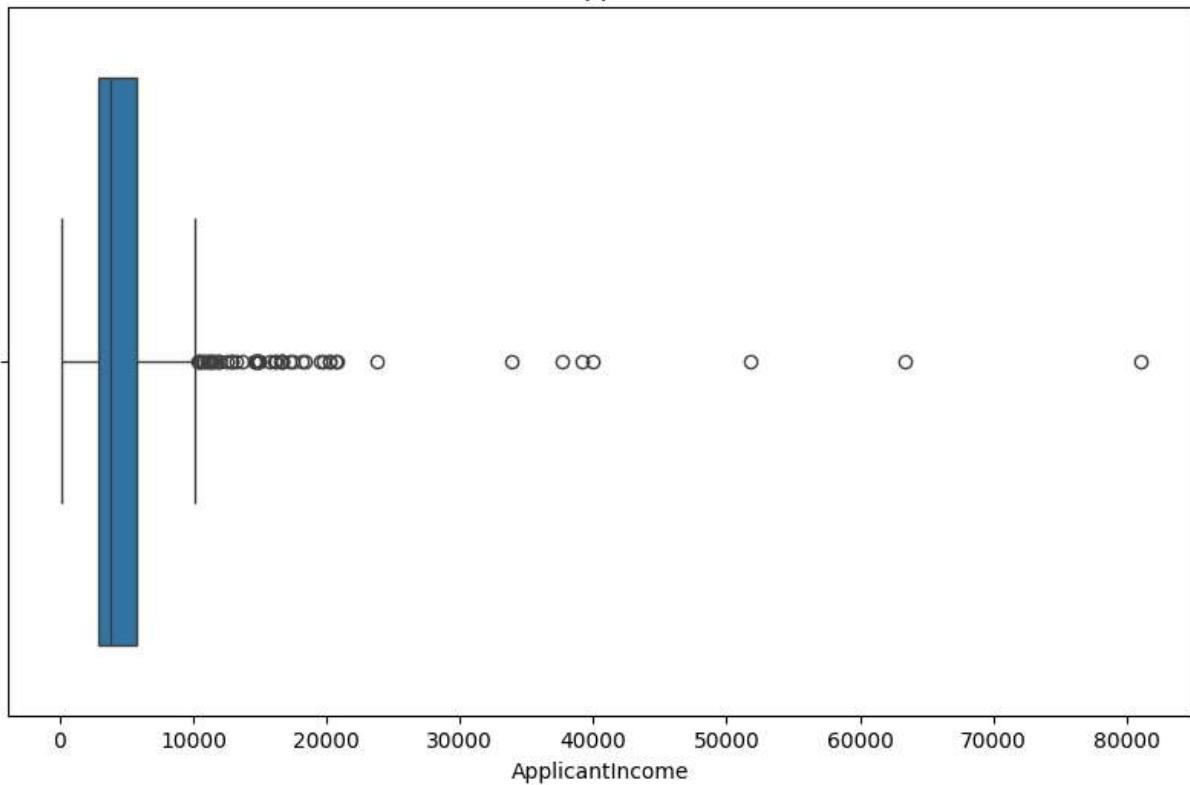




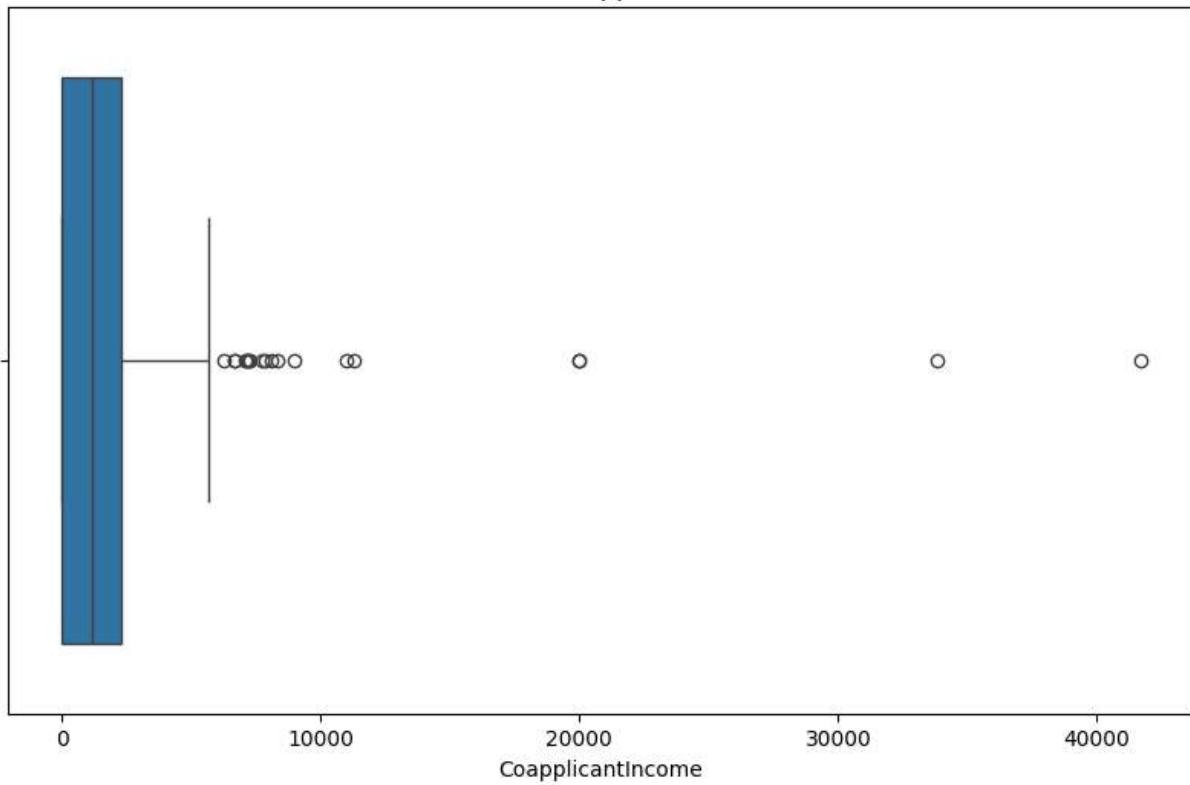


```
In [15]: # Box plot
for variables in numerical_variables:
    plt.figure(figsize=(10, 6))
    sns.boxplot(x=variables, data=df)
    plt.title(f'Box Plot of { variables}')
    plt.show()
```

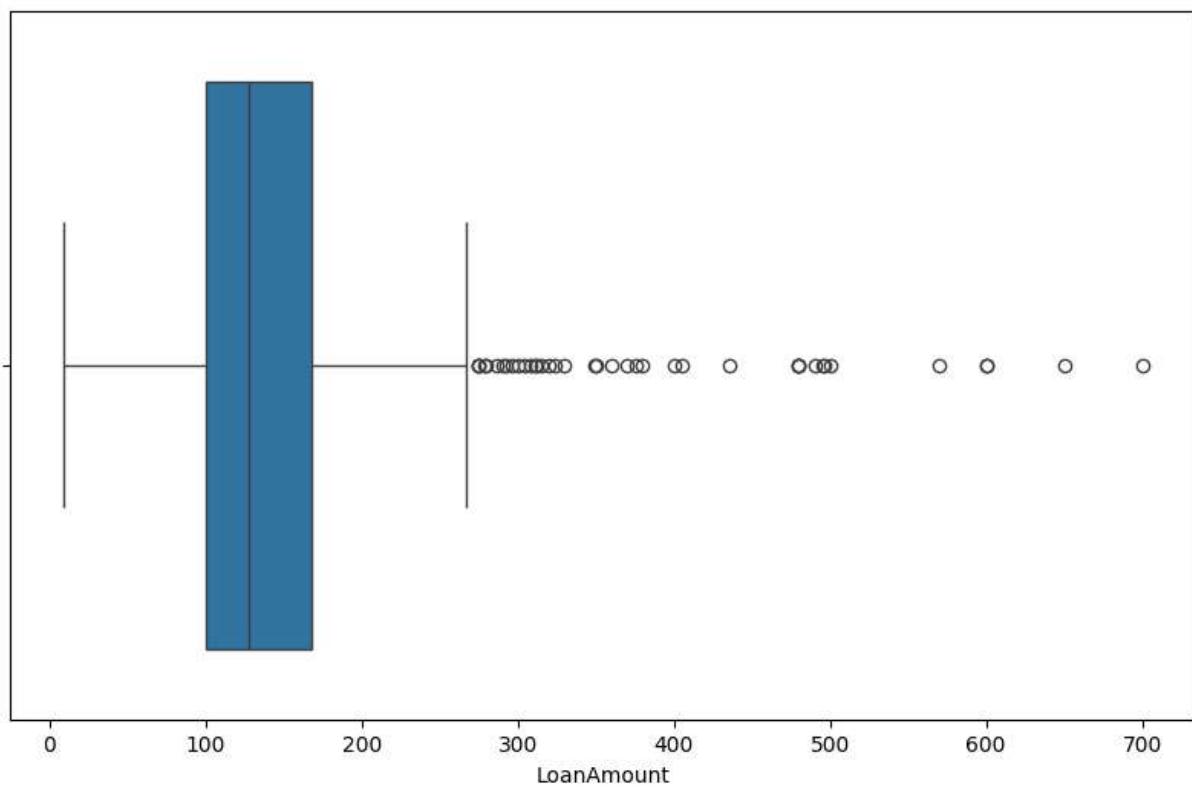
Box Plot of ApplicantIncome



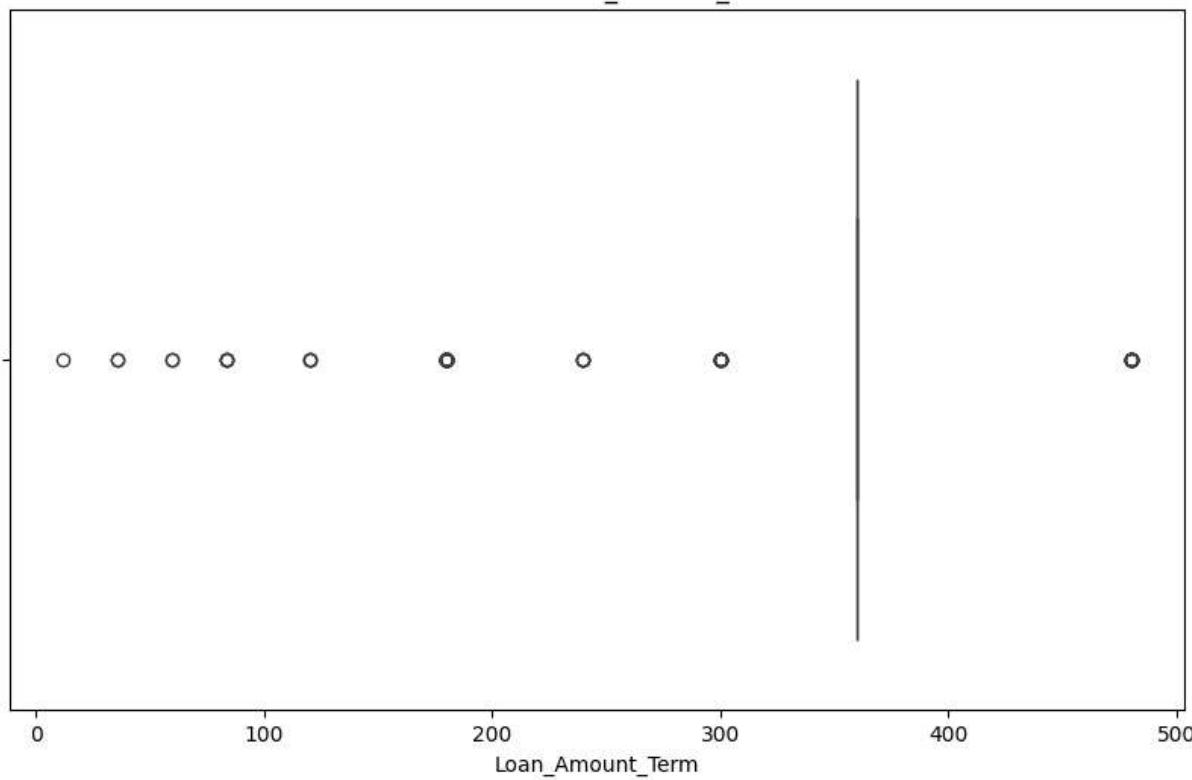
Box Plot of CoapplicantIncome



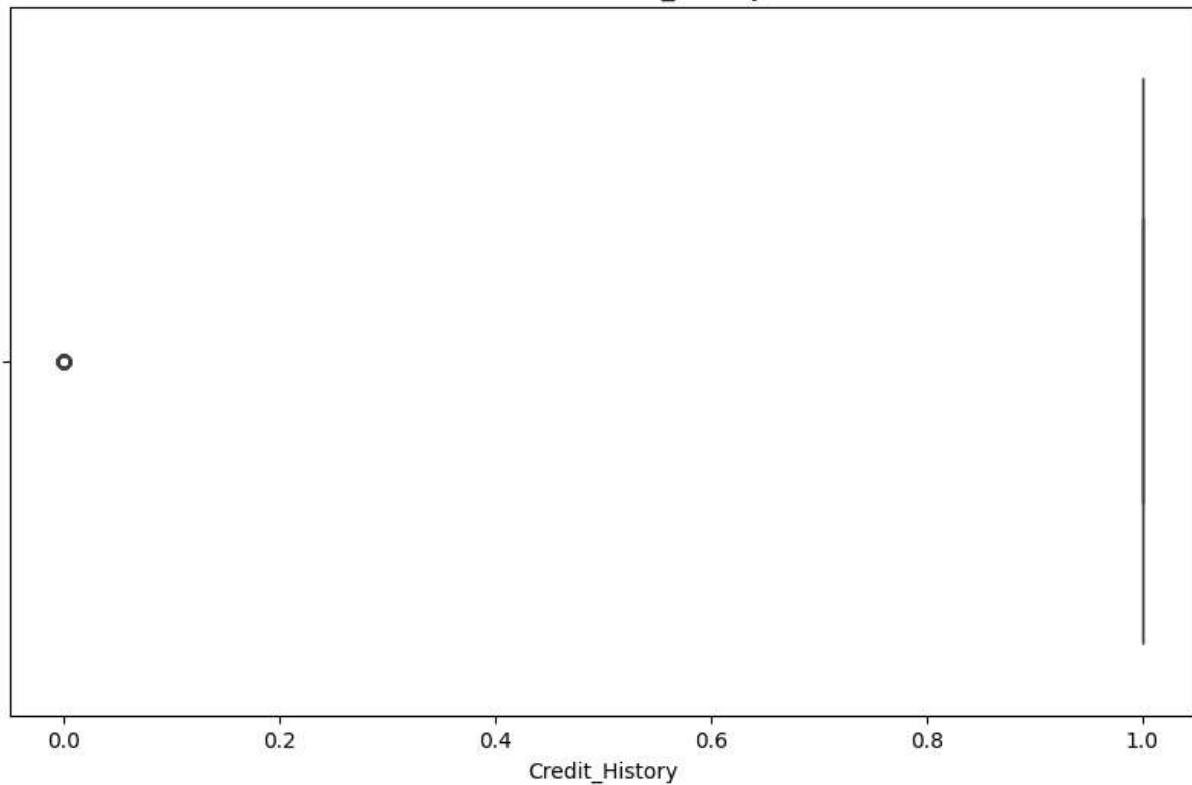
Box Plot of LoanAmount



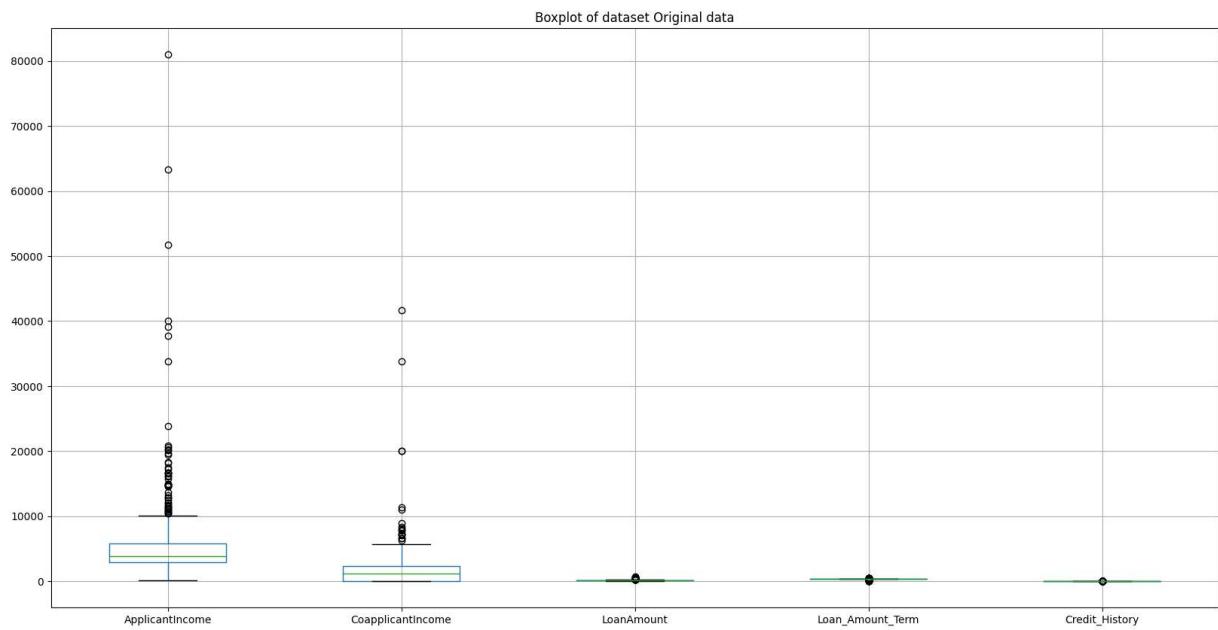
Box Plot of Loan_Amount_Term



Box Plot of Credit_History



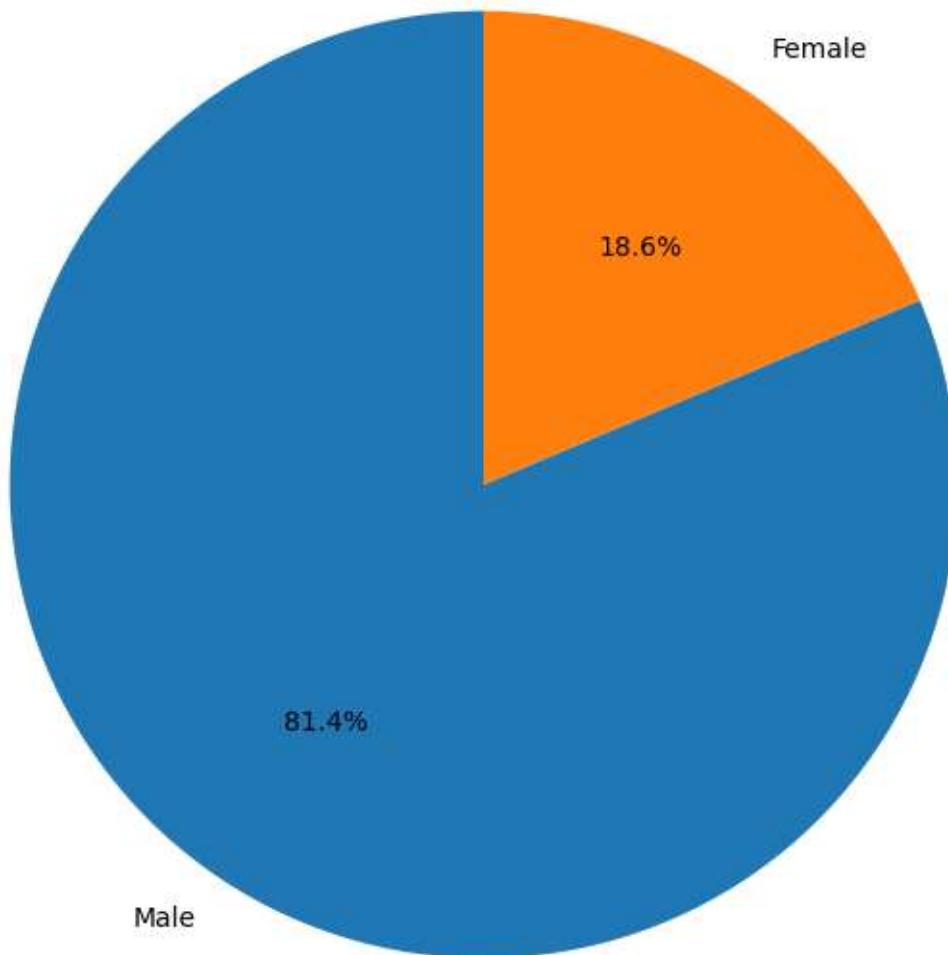
```
In [16]: #Grouped Box Plot
df.boxplot(figsize=(20, 10))
plt.title("Boxplot of dataset Original data")
plt.show()
```



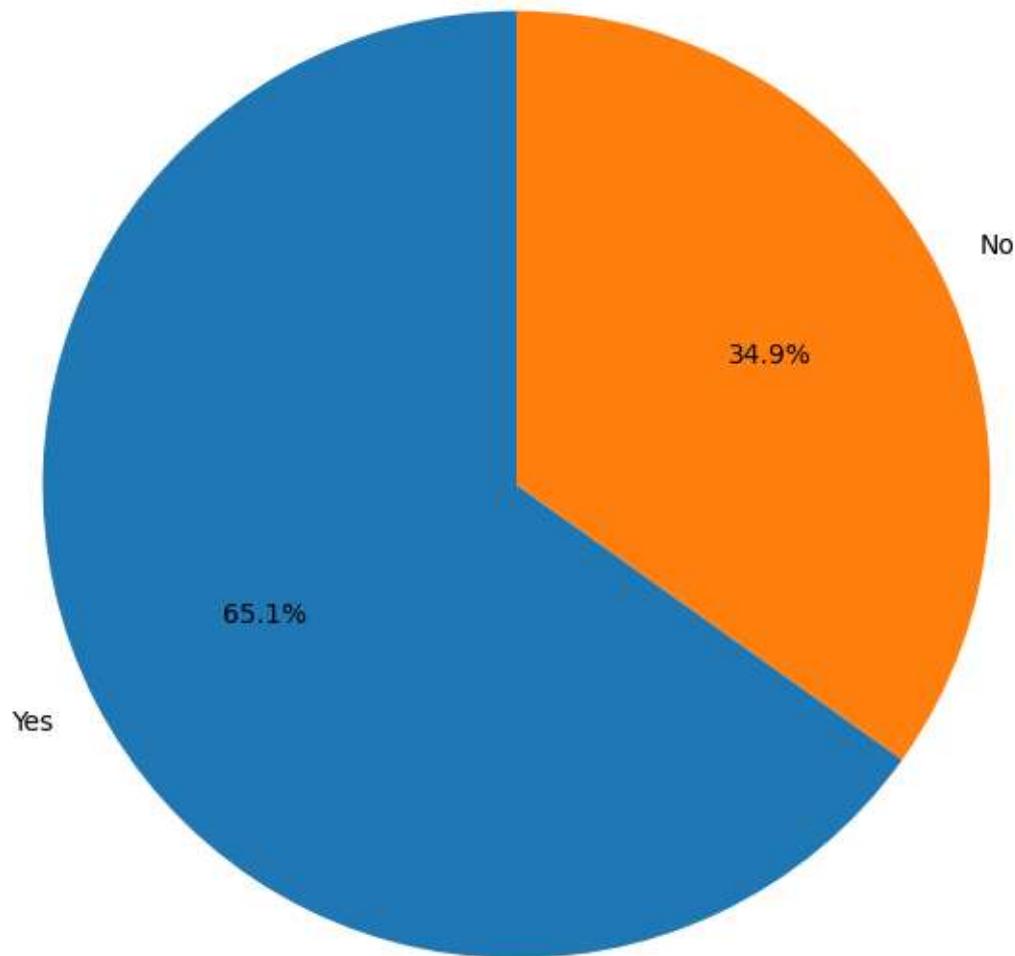
```
In [17]: # Pie chart
for variables in categorical_variables:
    status_distribution = df[variables].value_counts()
    plt.figure(figsize=(8,8))
    plt.pie(status_distribution, labels=status_distribution.index, autopct='%1.1f%%')
```

```
plt.title(f'{variables}' 'Distribution')  
plt.show()
```

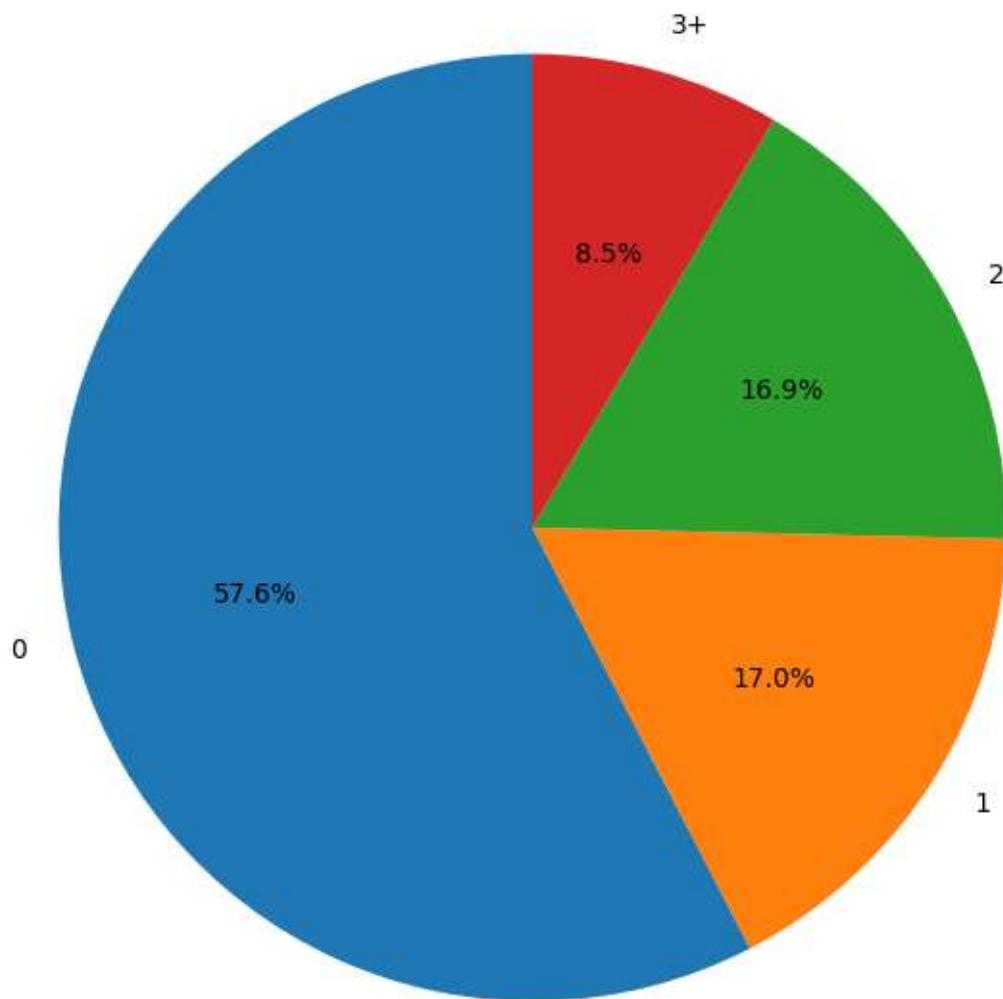
GenderDistribution



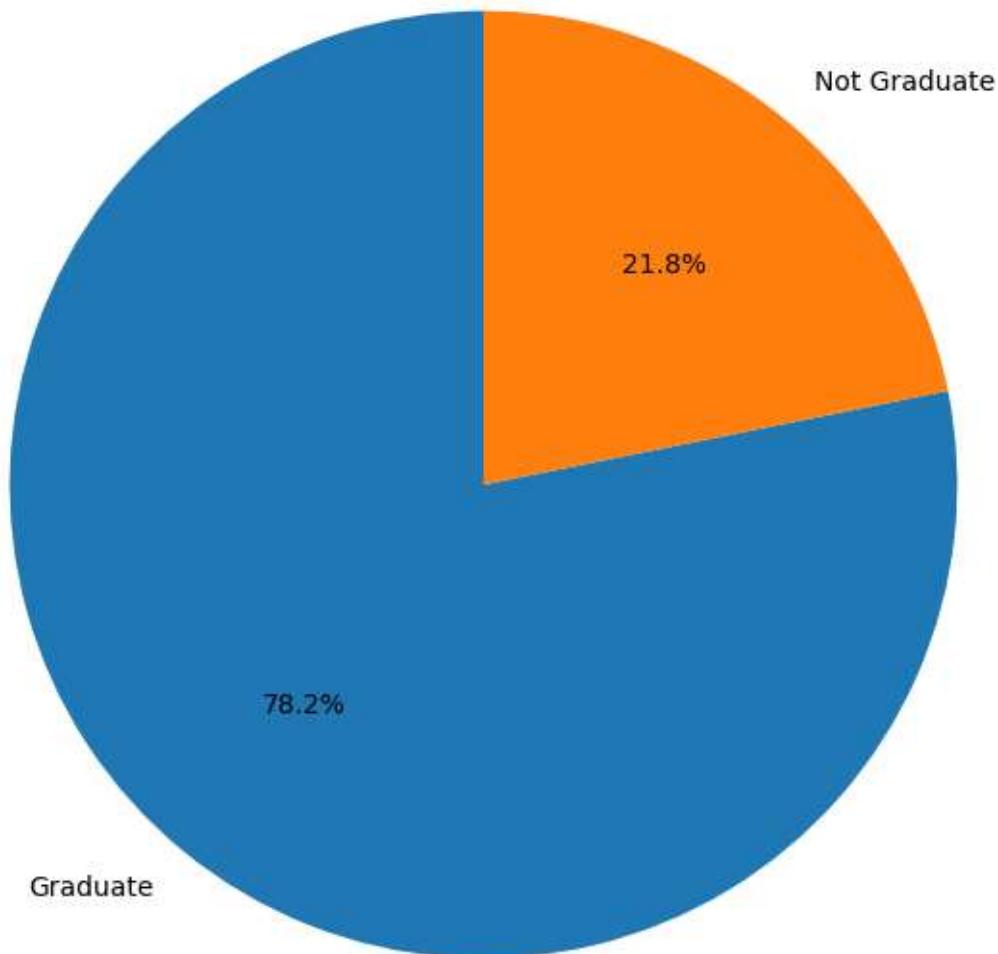
Married Distribution



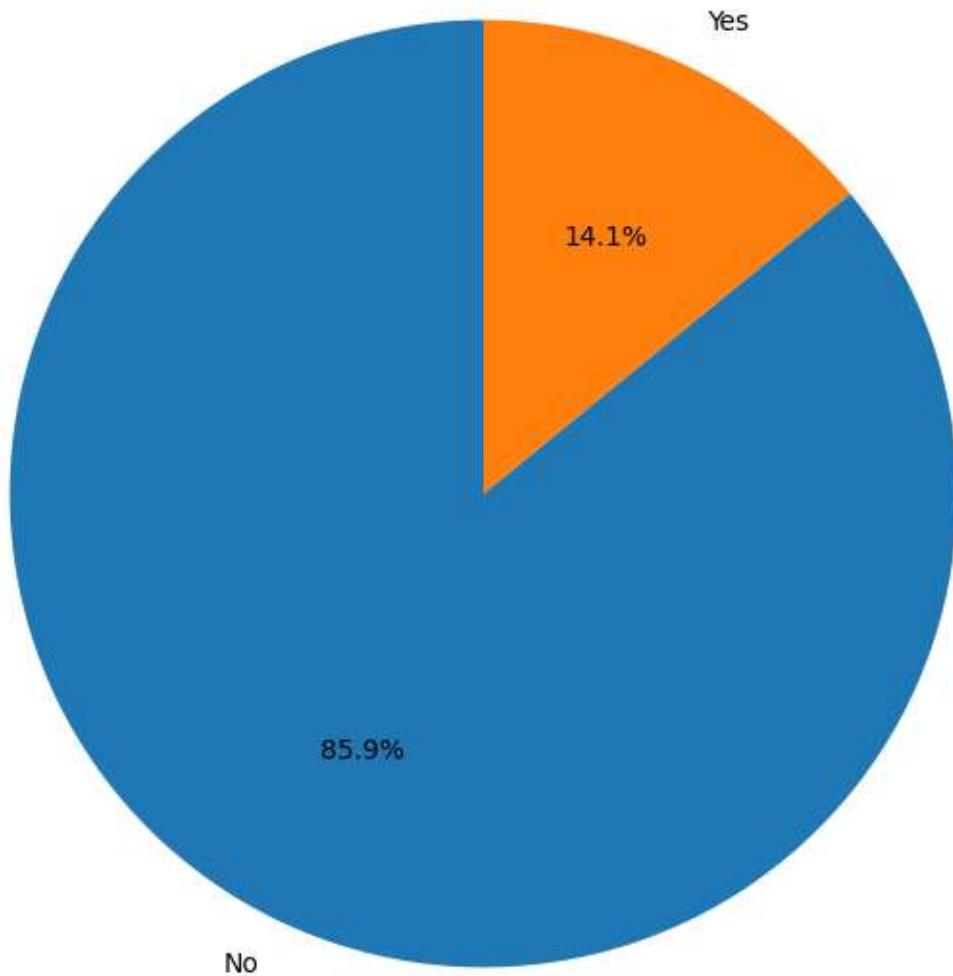
Dependents Distribution



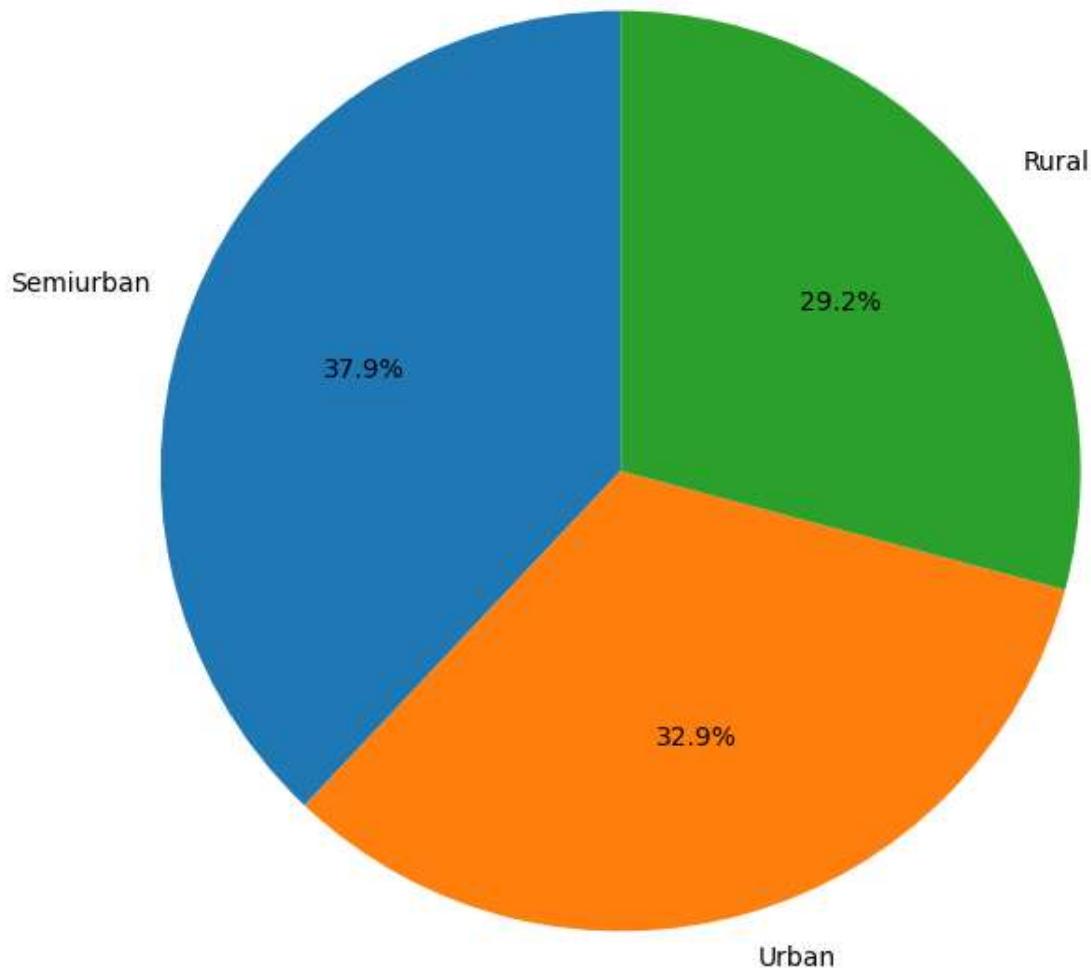
Education Distribution



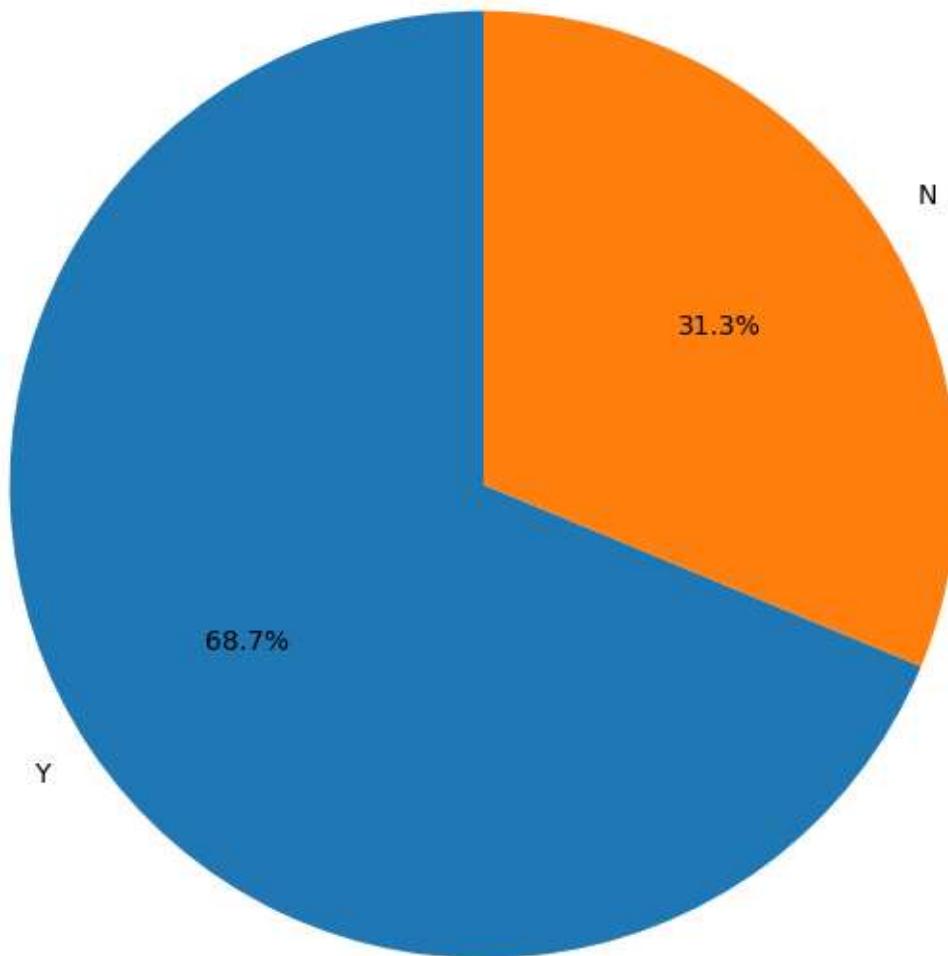
Self_EmployedDistribution



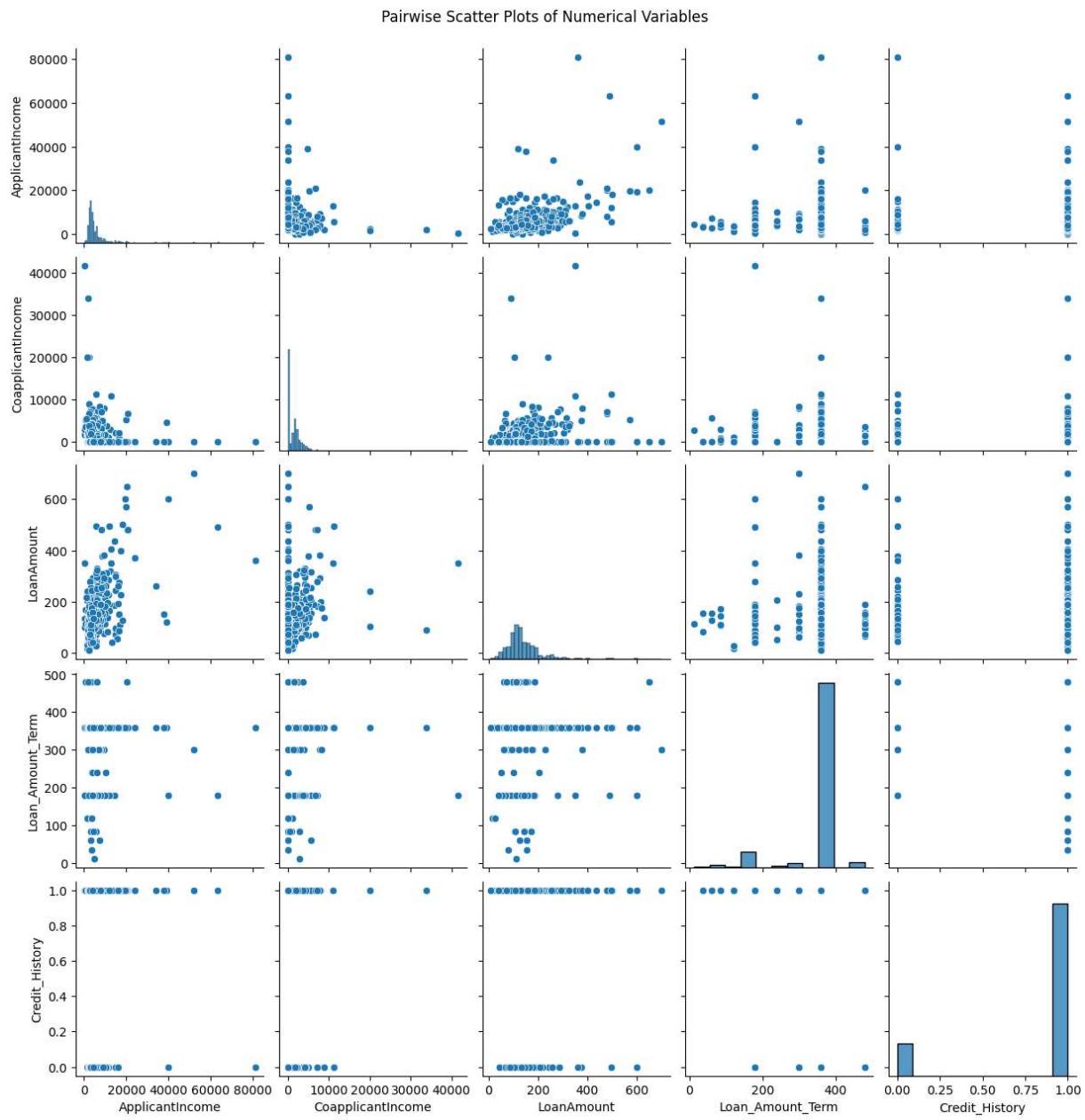
Property_AreaDistribution



Loan_StatusDistribution

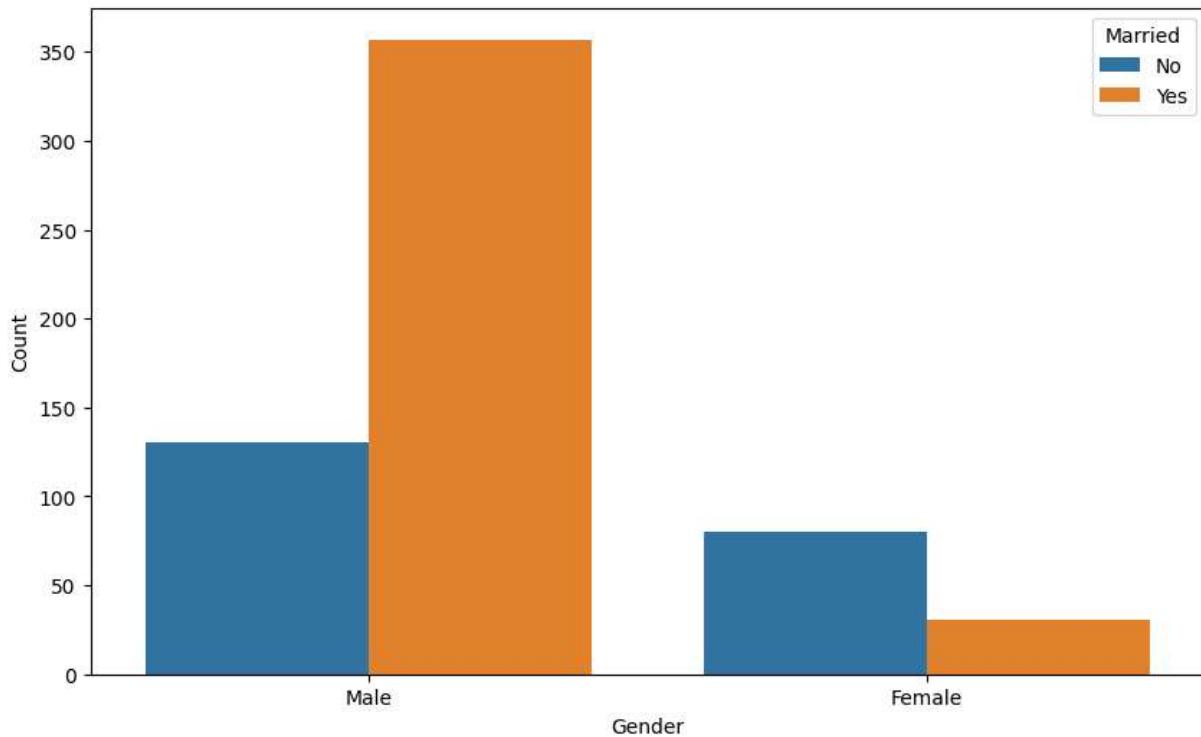


```
In [18]: # Bivariate analysis for each pair of numerical variables  
# Pairwise scatter plots for numerical variables  
sns.pairplot(df[numerical_variables])  
plt.suptitle('Pairwise Scatter Plots of Numerical Variables', y=1.02)  
plt.show()
```

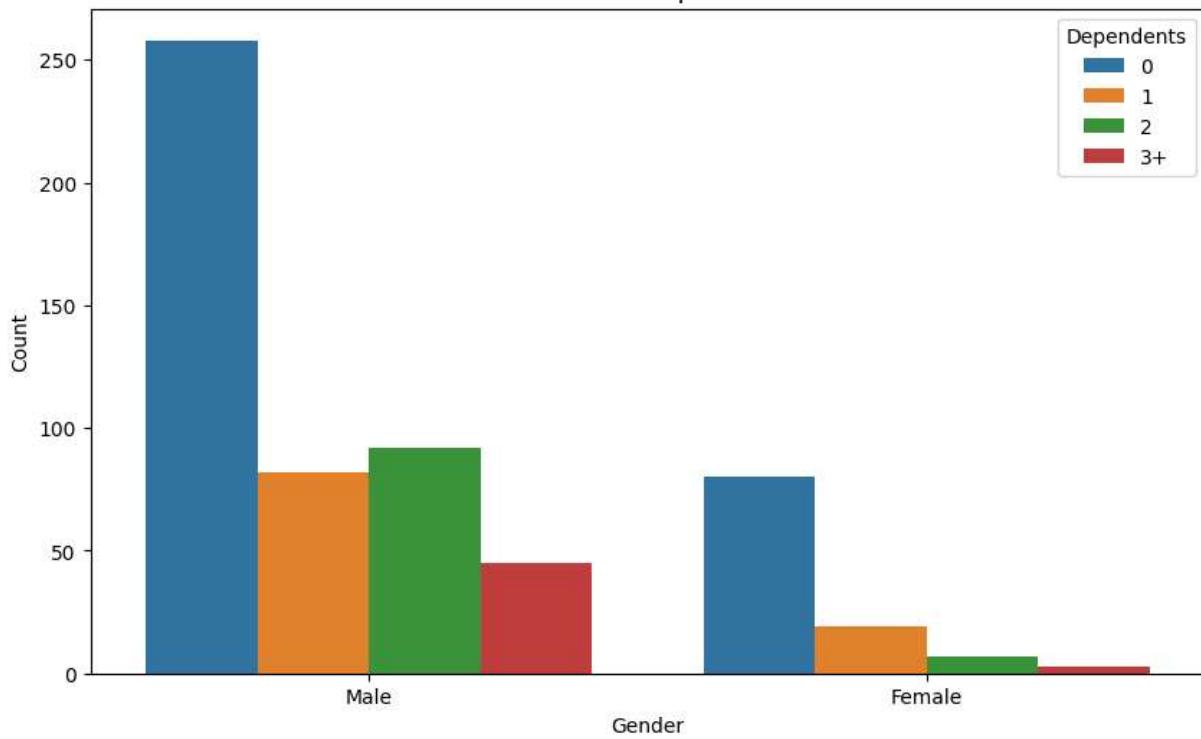


```
In [19]: # Bivariate analysis for each pair of categorical variables
for i in range(len(categorical_variables) - 1):
    for j in range(i + 1, len(categorical_variables)):
        plt.figure(figsize=(10, 6))
        sns.countplot(x=categorical_variables[i], hue=categorical_variables[j], data=loan)
        plt.title(f'{categorical_variables[i]} vs. {categorical_variables[j]}')
        plt.xlabel(categorical_variables[i])
        plt.ylabel('Count')
        plt.show()
```

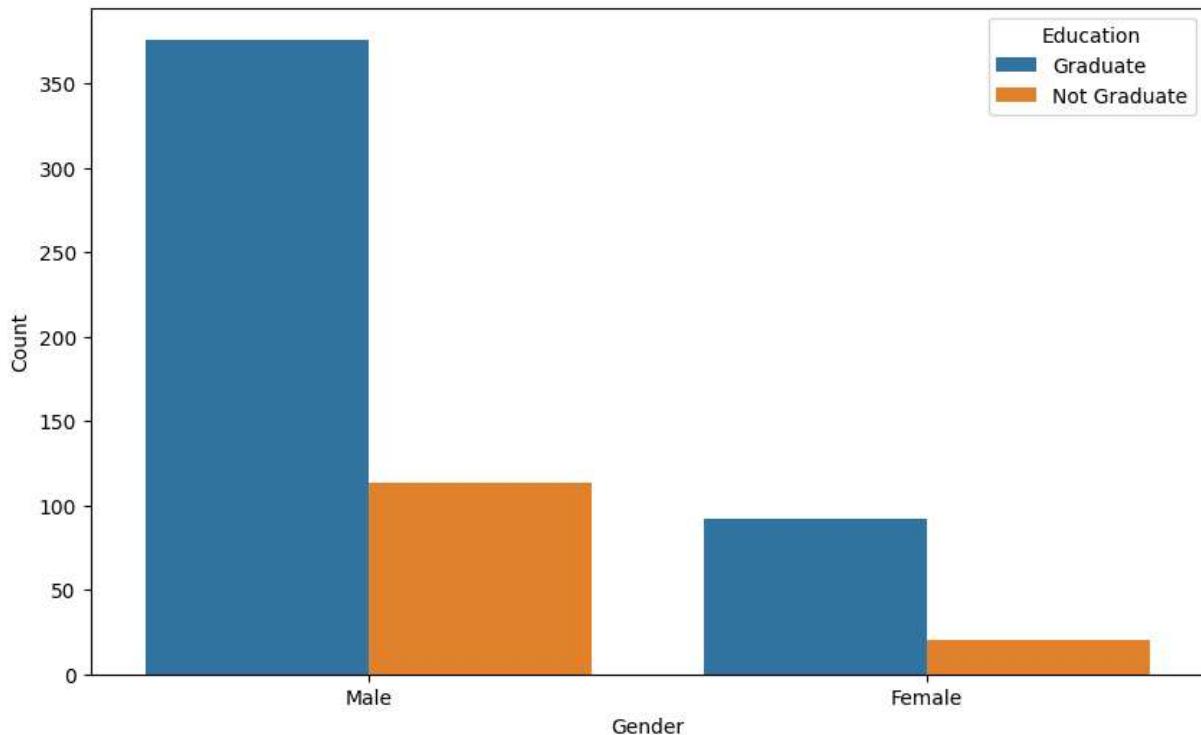
Gender vs. Married



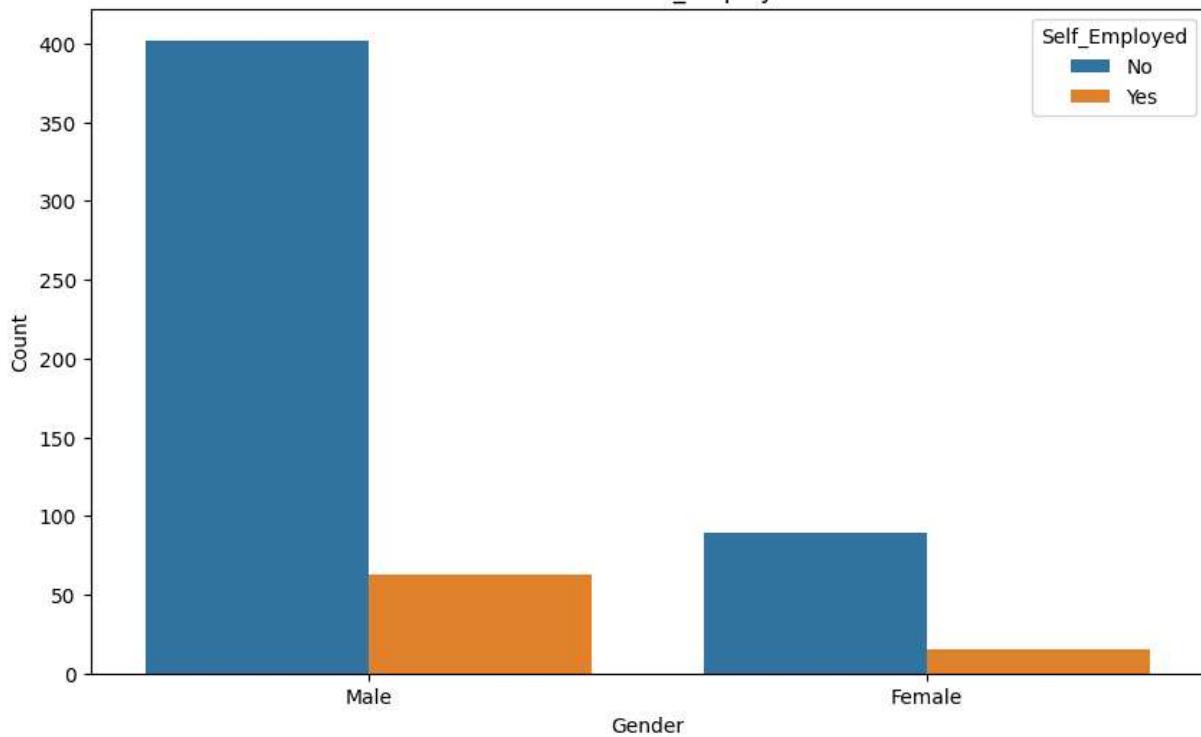
Gender vs. Dependents

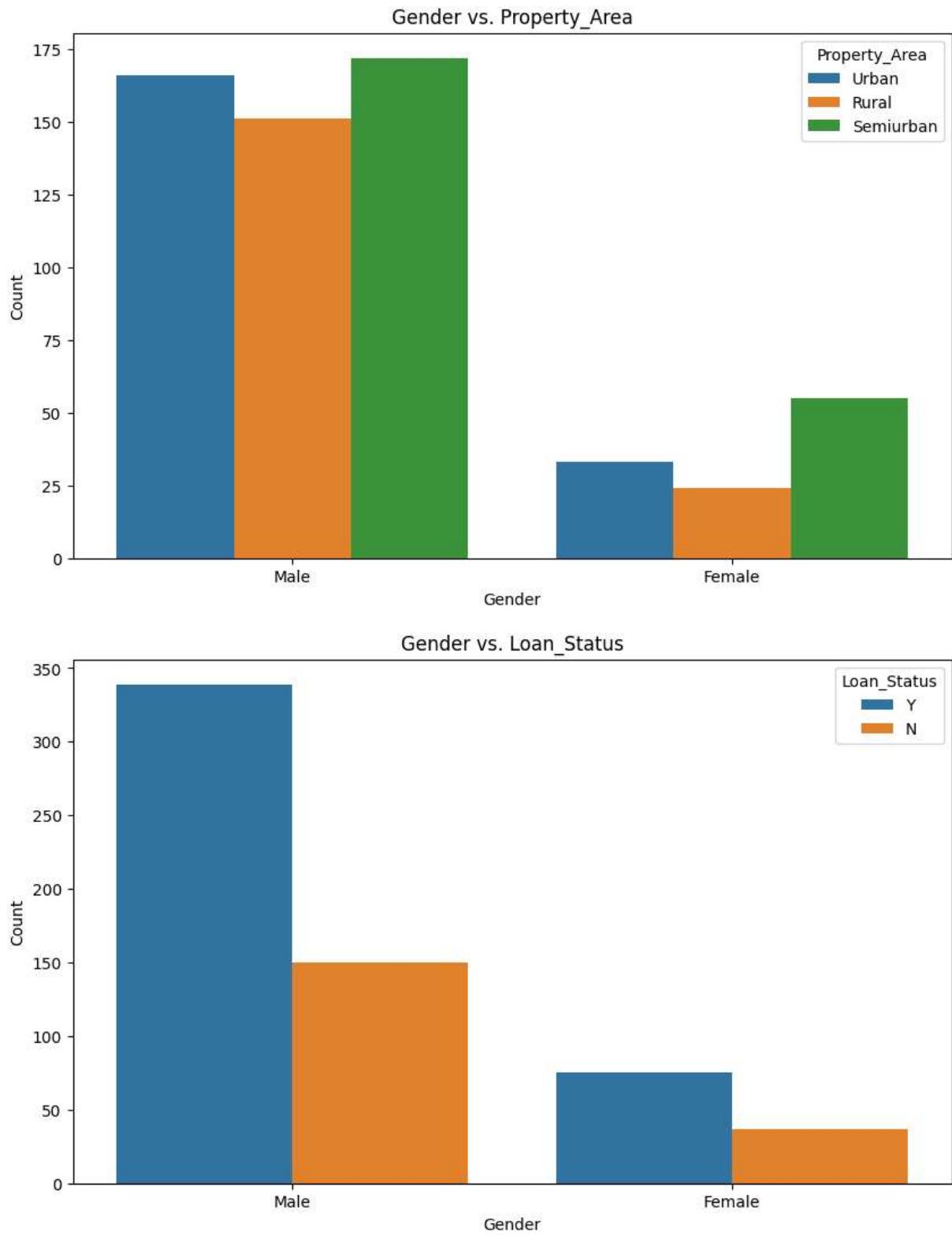


Gender vs. Education

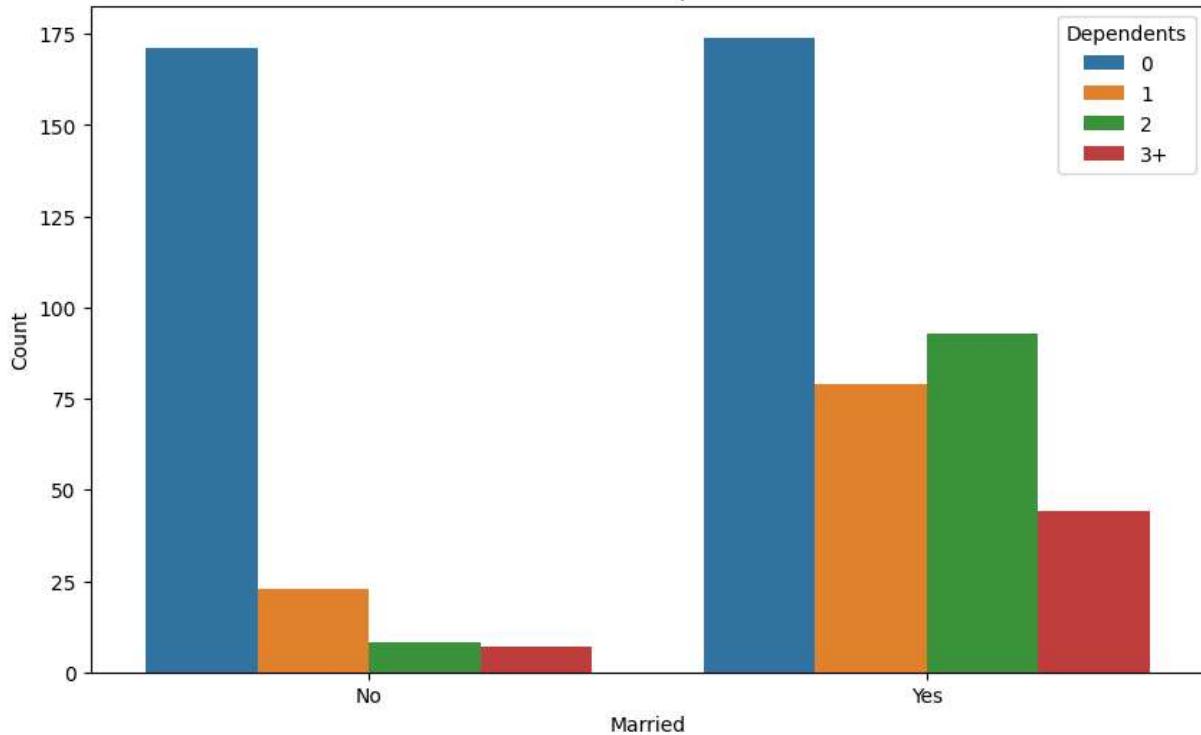


Gender vs. Self_Employed

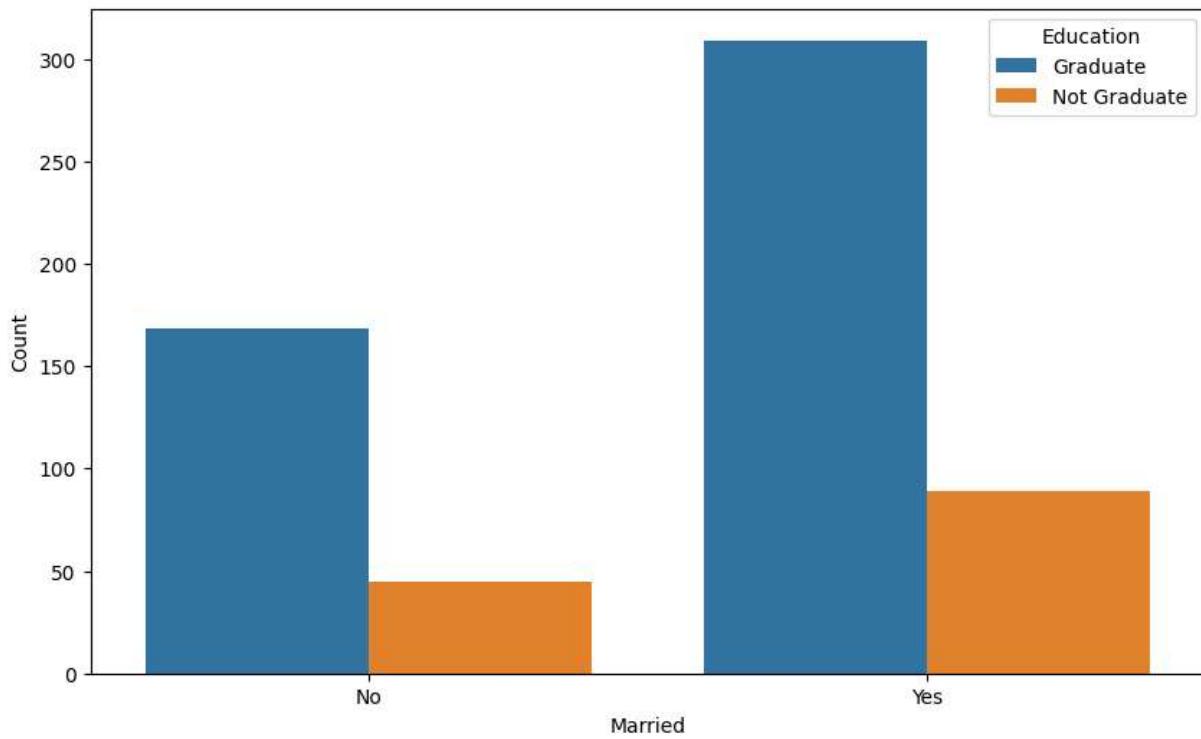




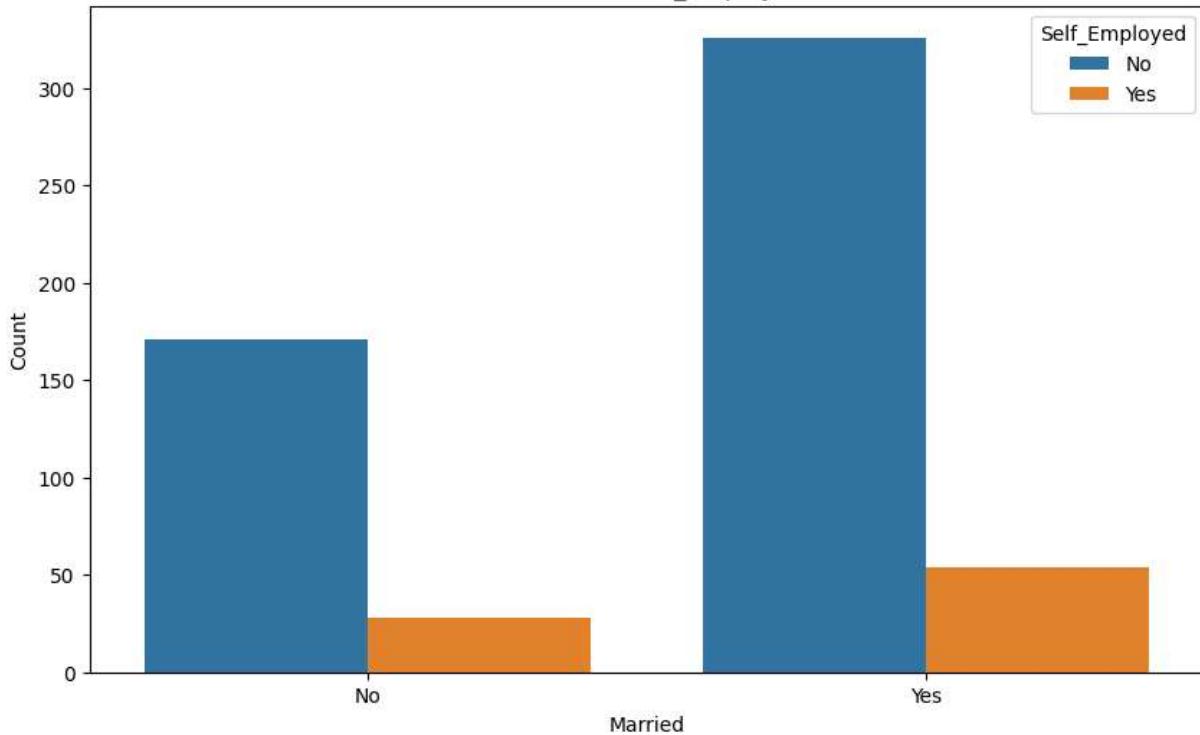
Married vs. Dependents



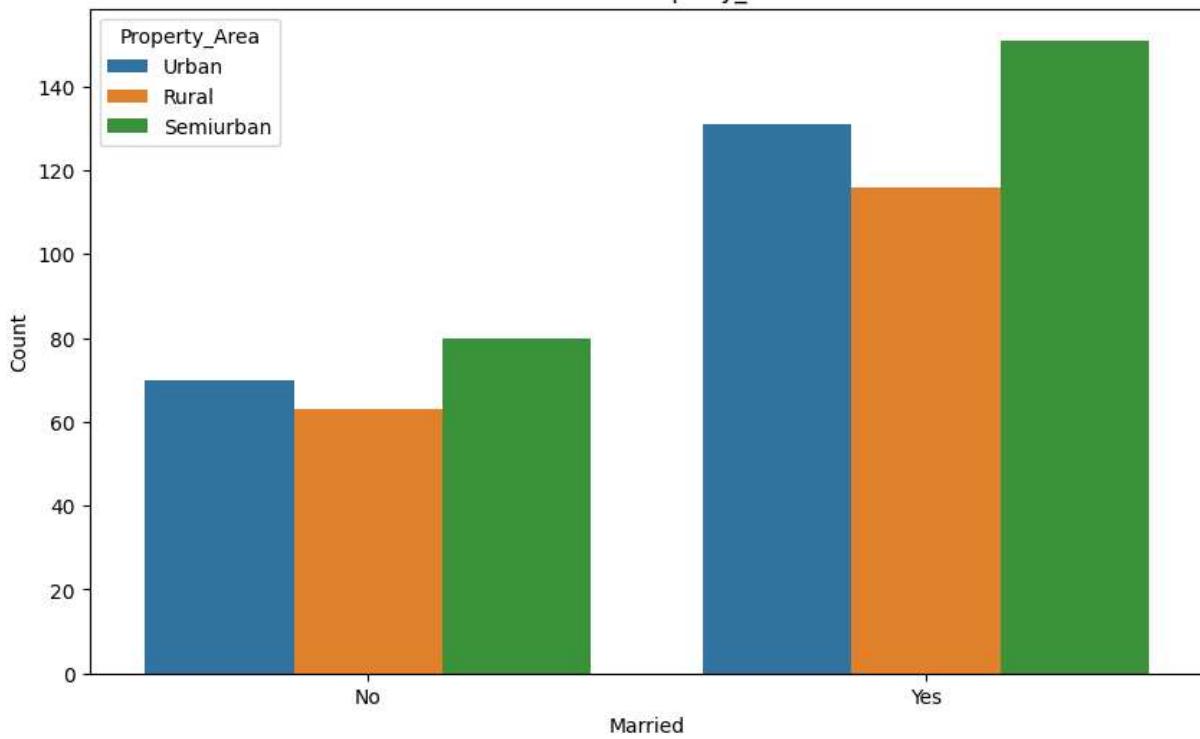
Married vs. Education



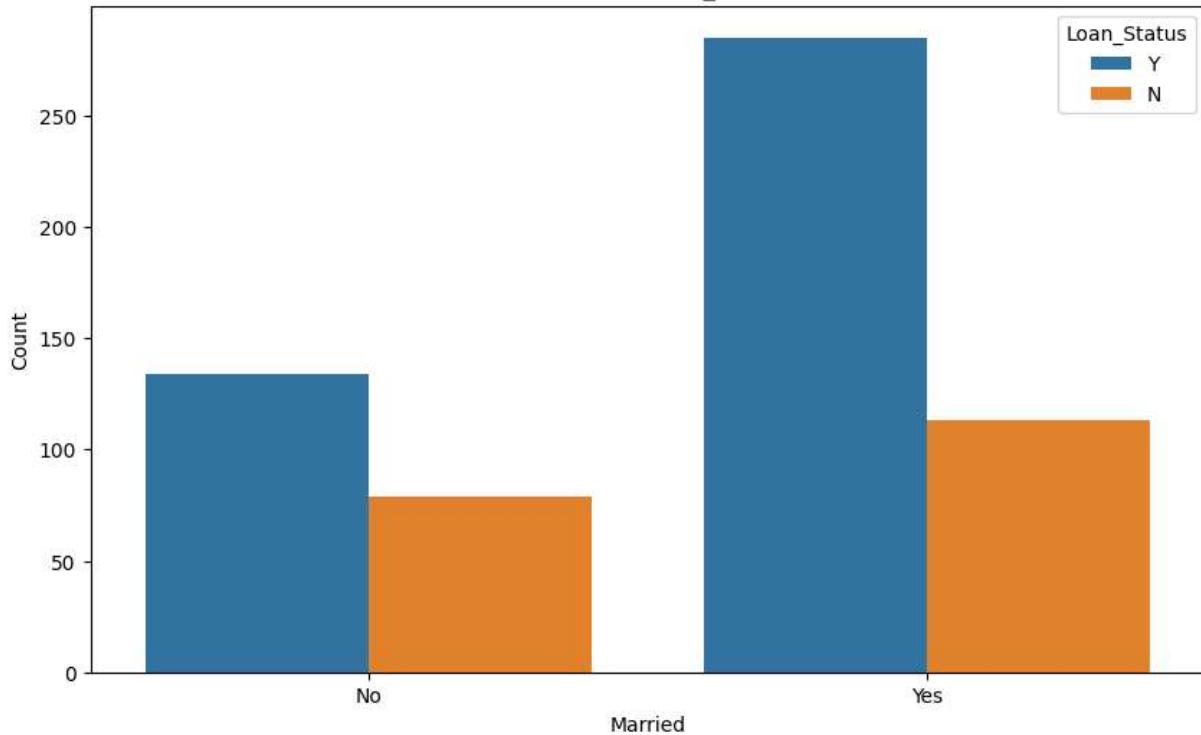
Married vs. Self_Employed



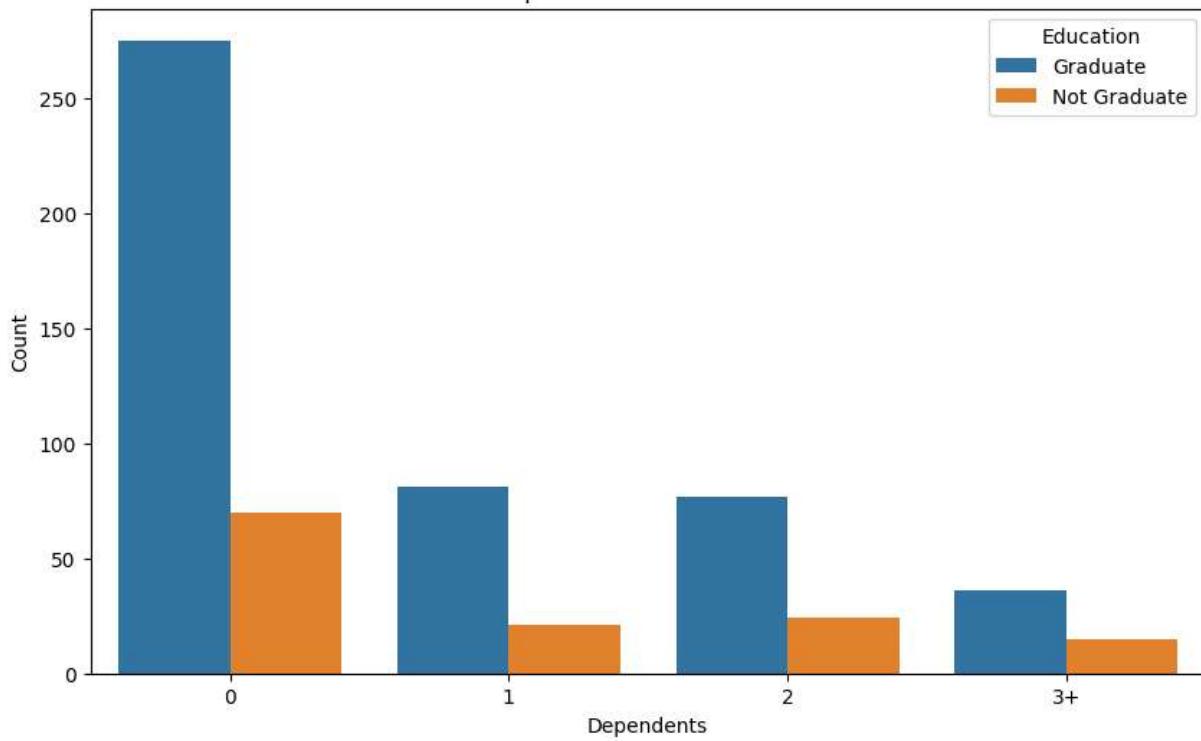
Married vs. Property_Area



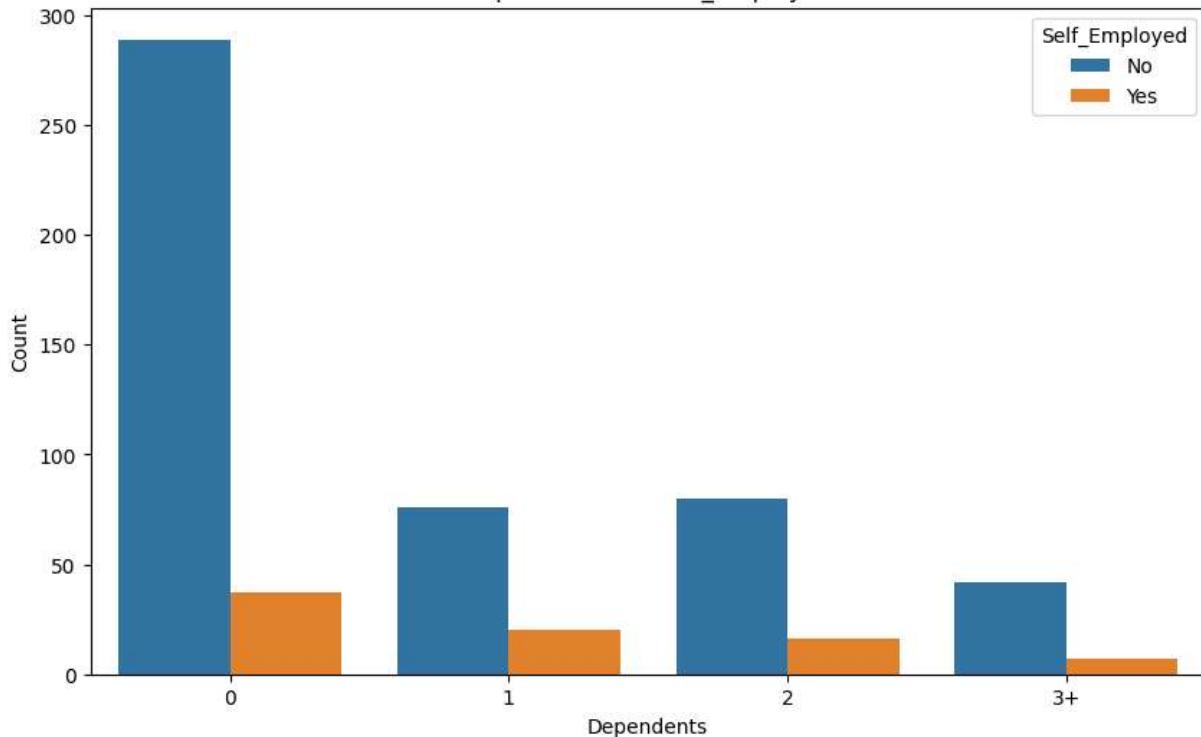
Married vs. Loan_Status



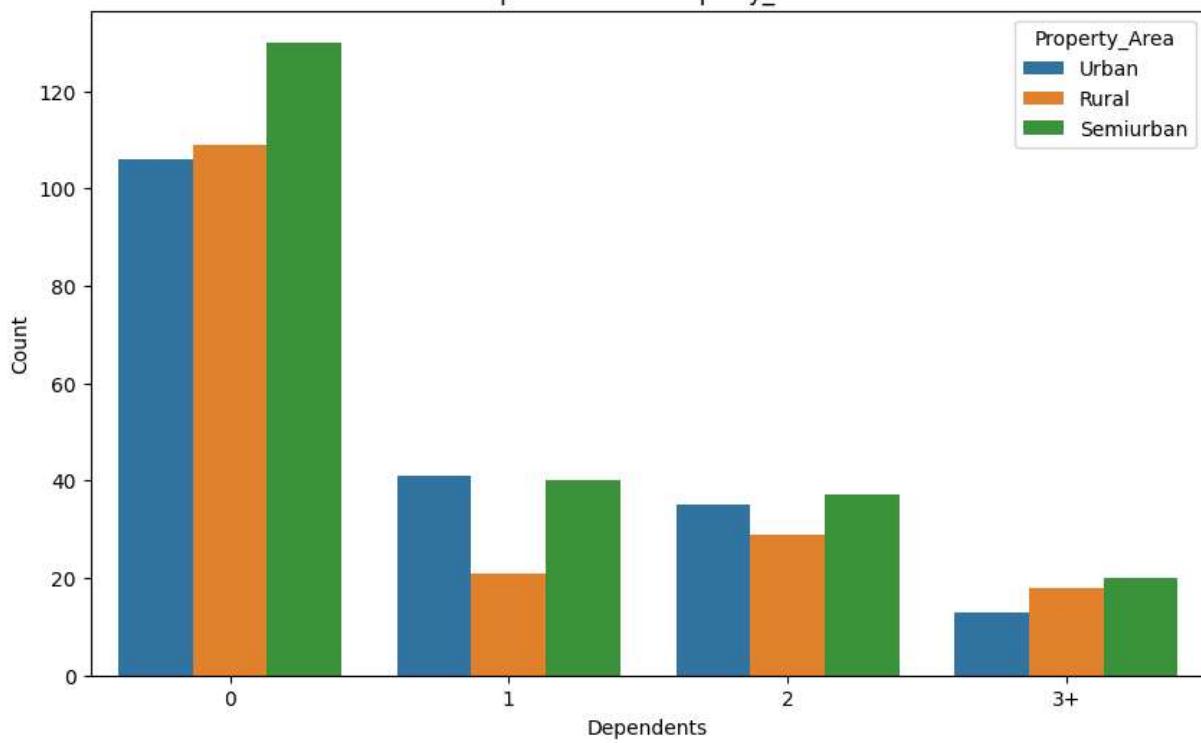
Dependents vs. Education



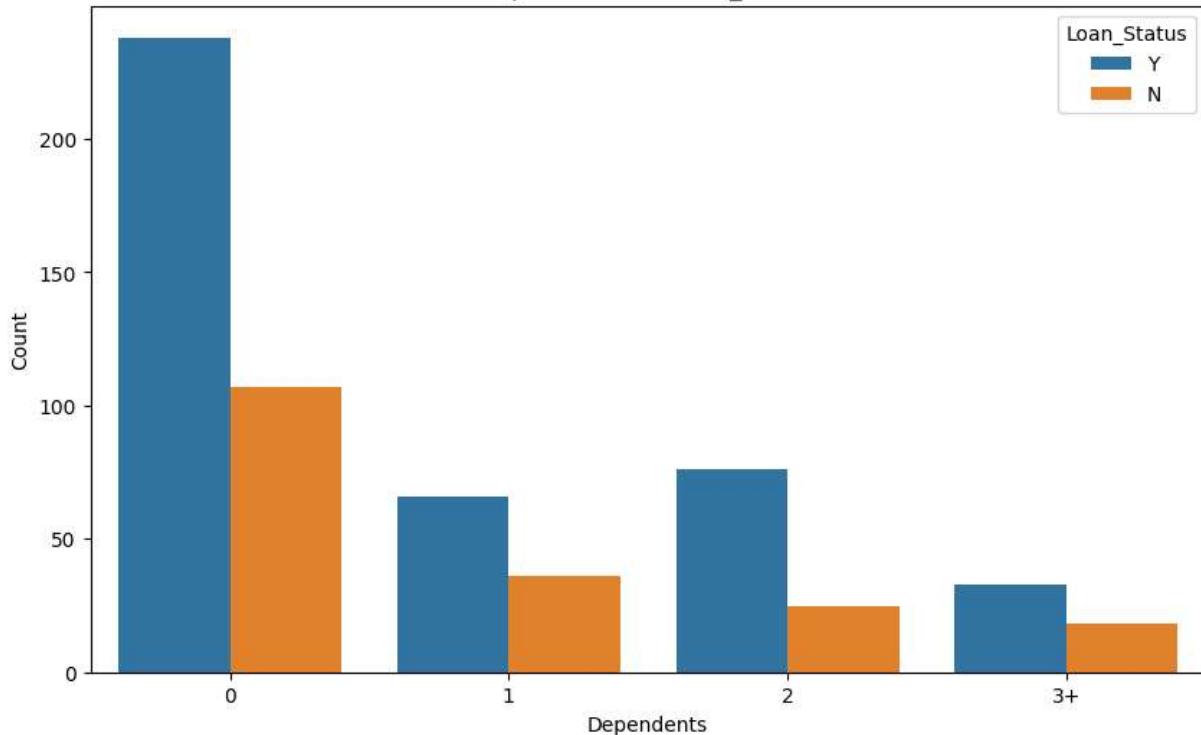
Dependents vs. Self_Employed



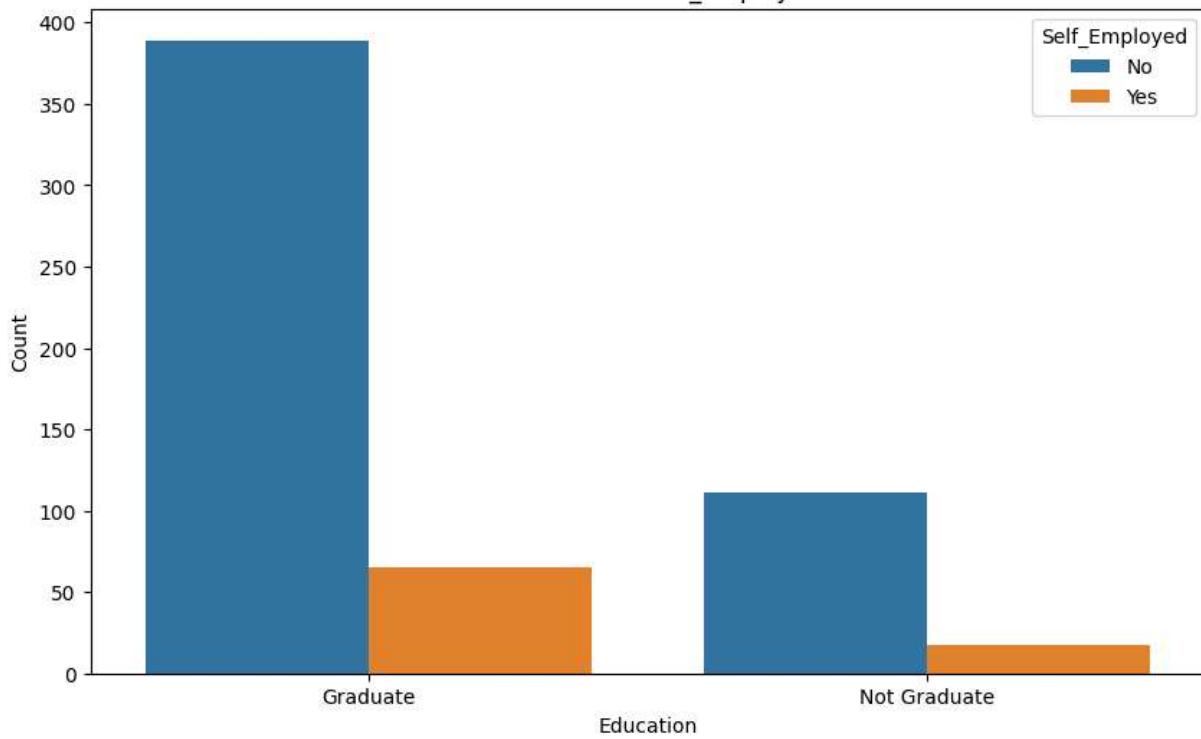
Dependents vs. Property_Area



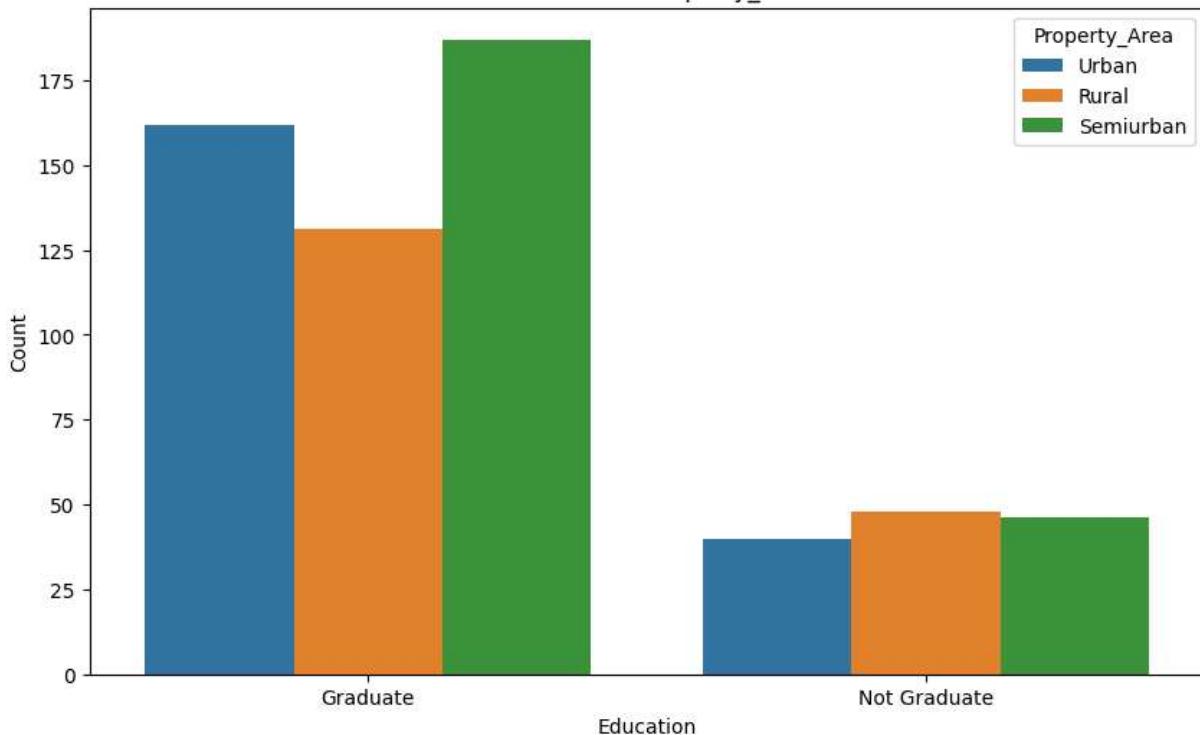
Dependents vs. Loan_Status



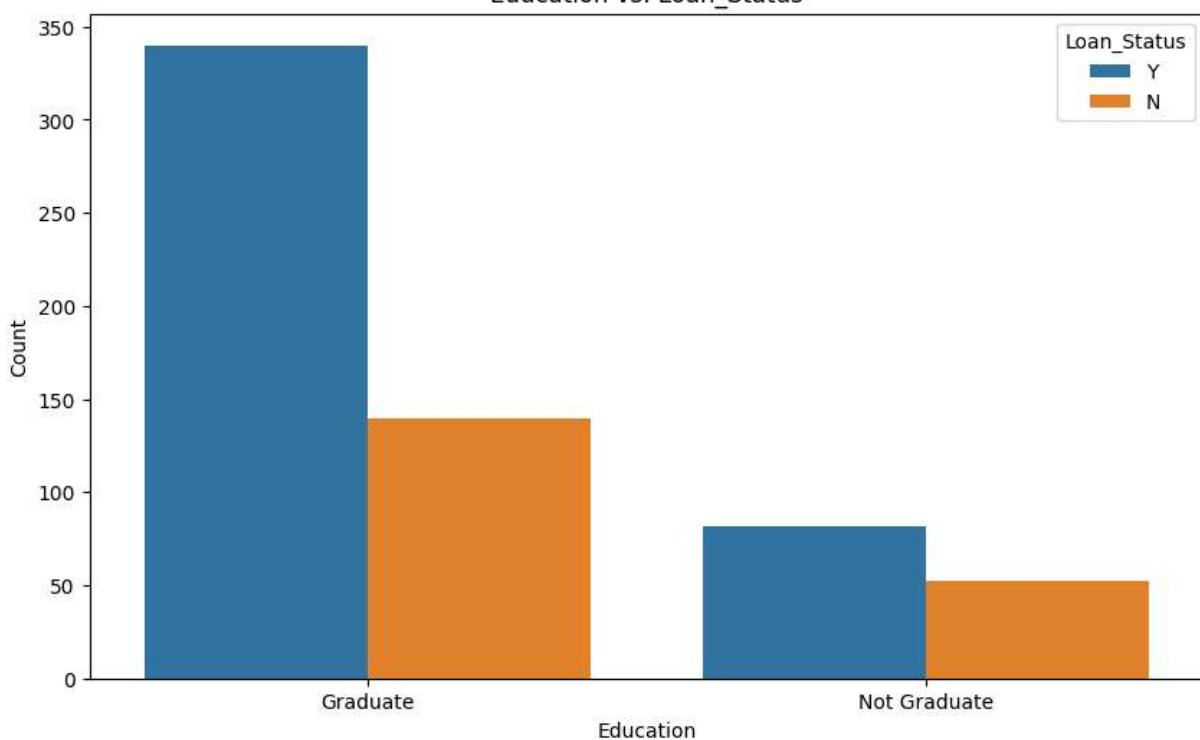
Education vs. Self_Employed

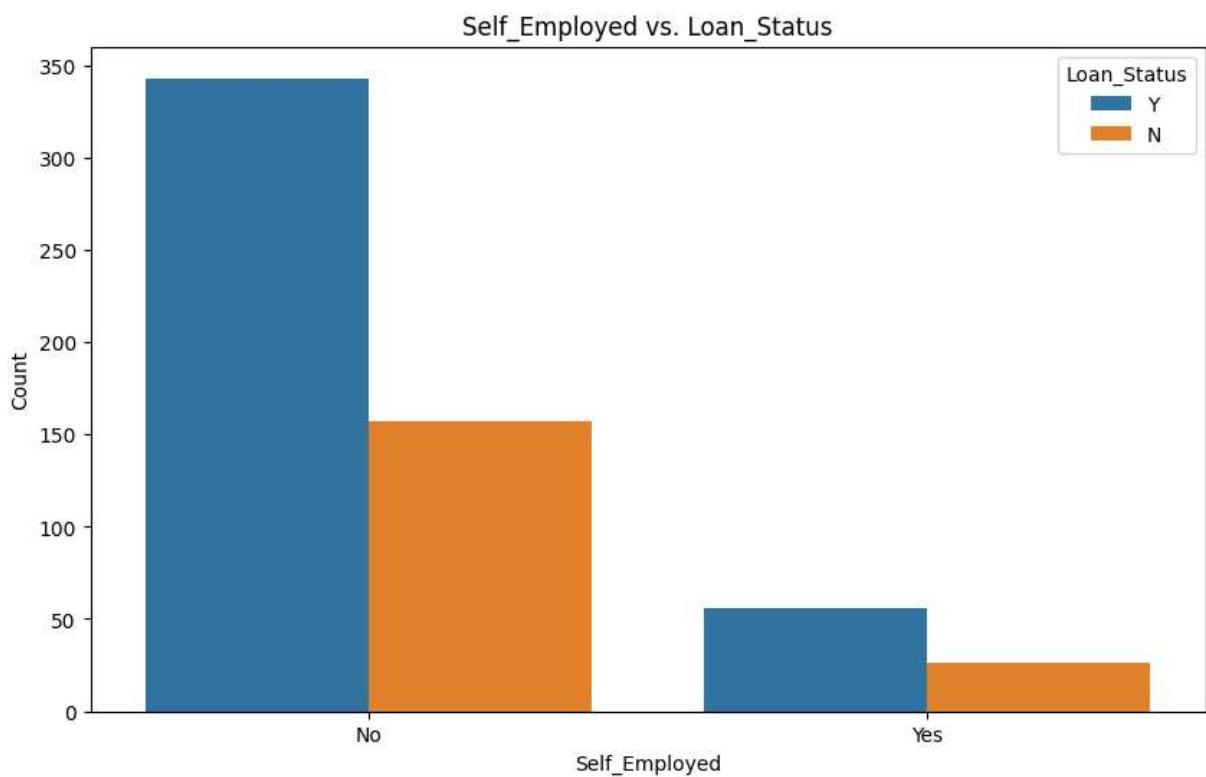
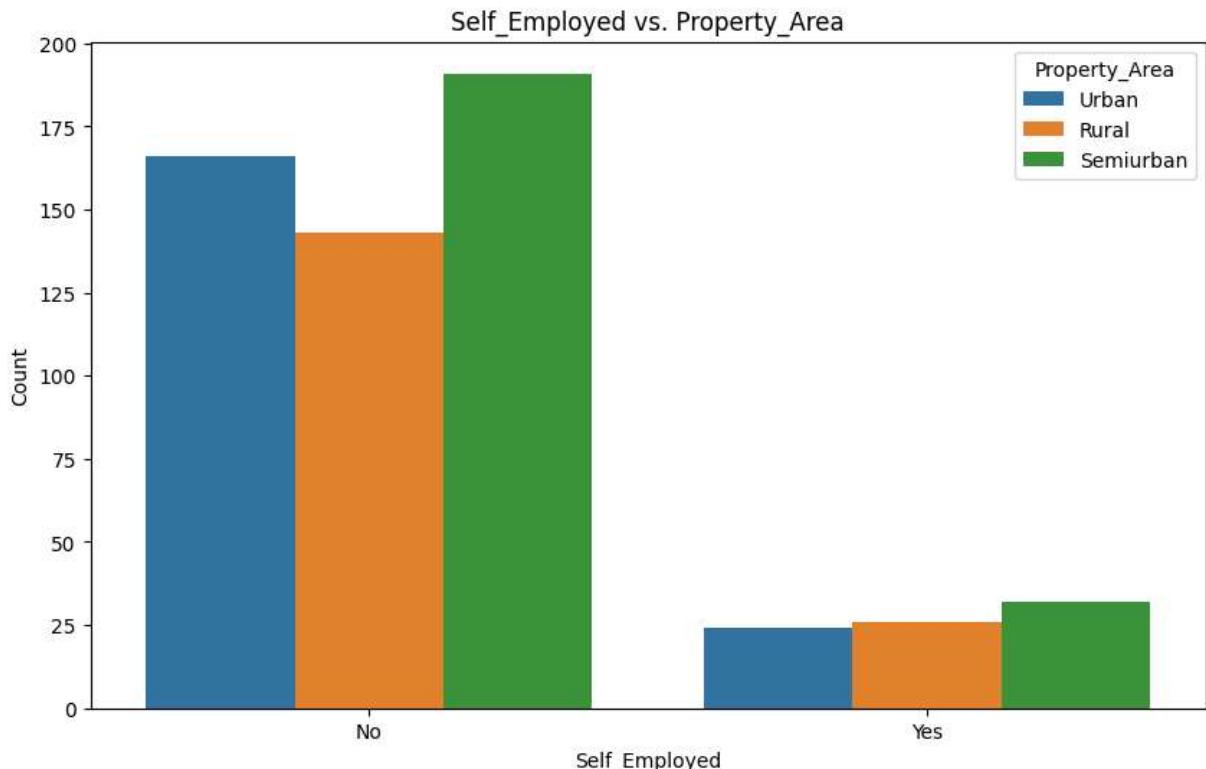


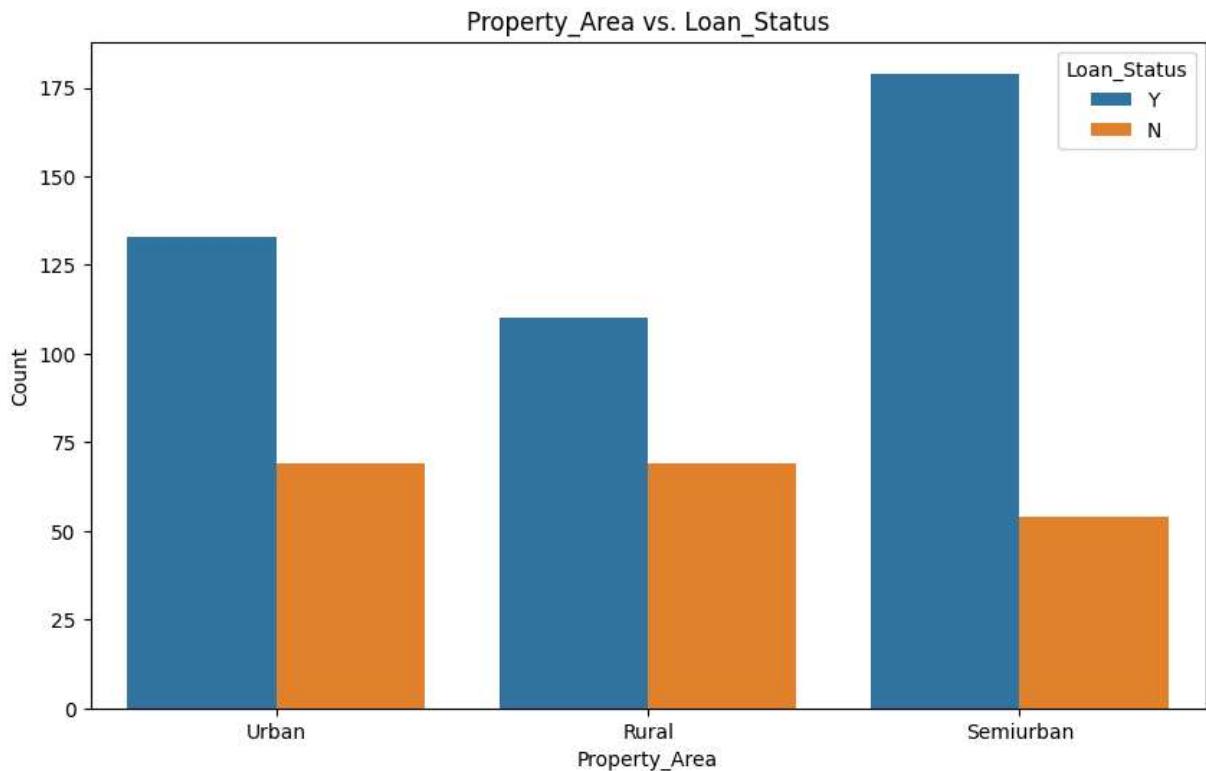
Education vs. Property_Area



Education vs. Loan_Status

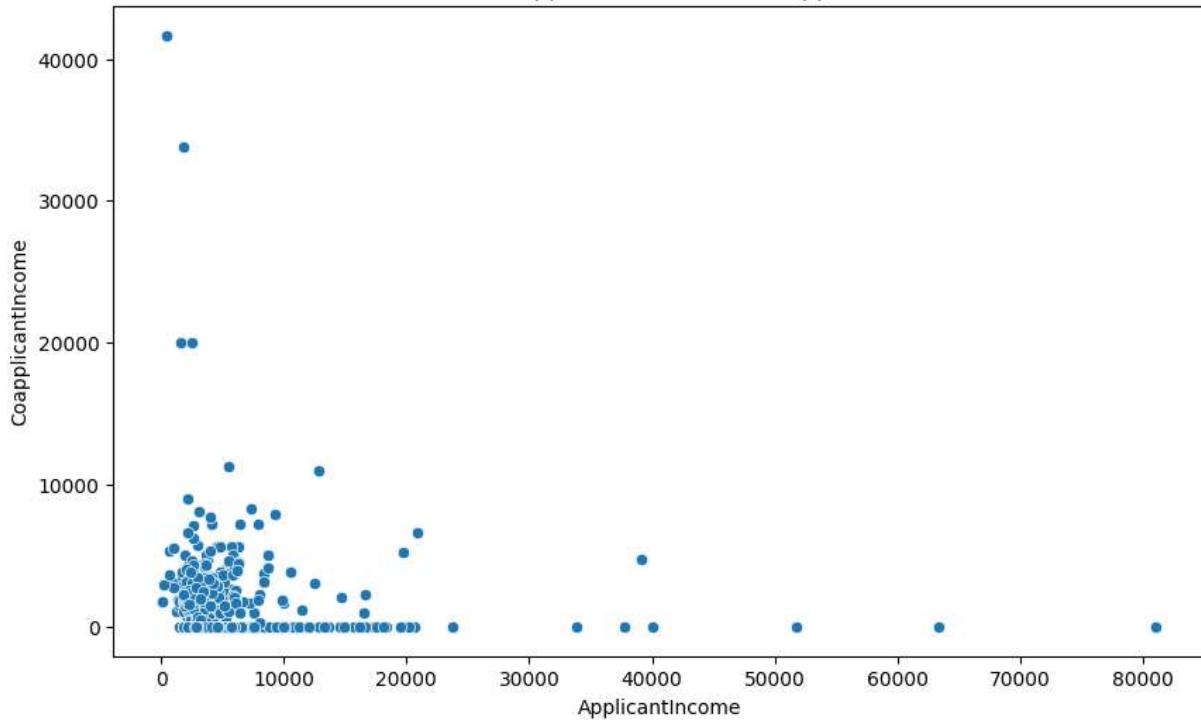




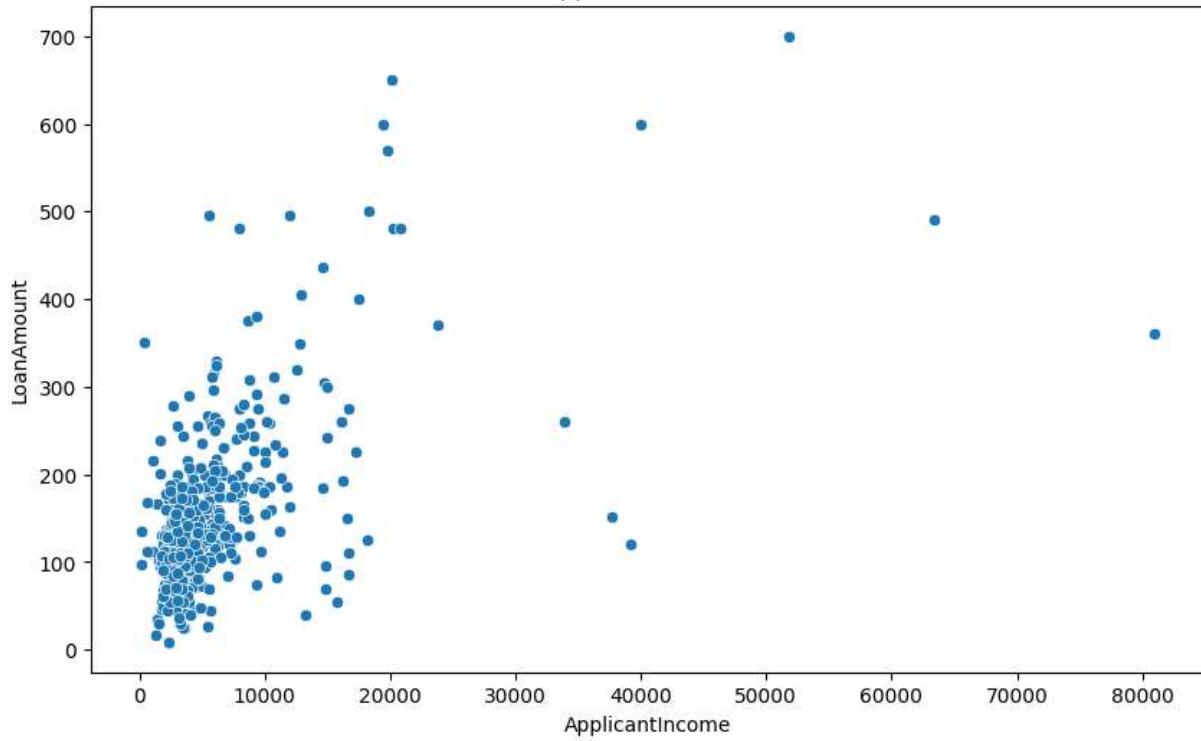


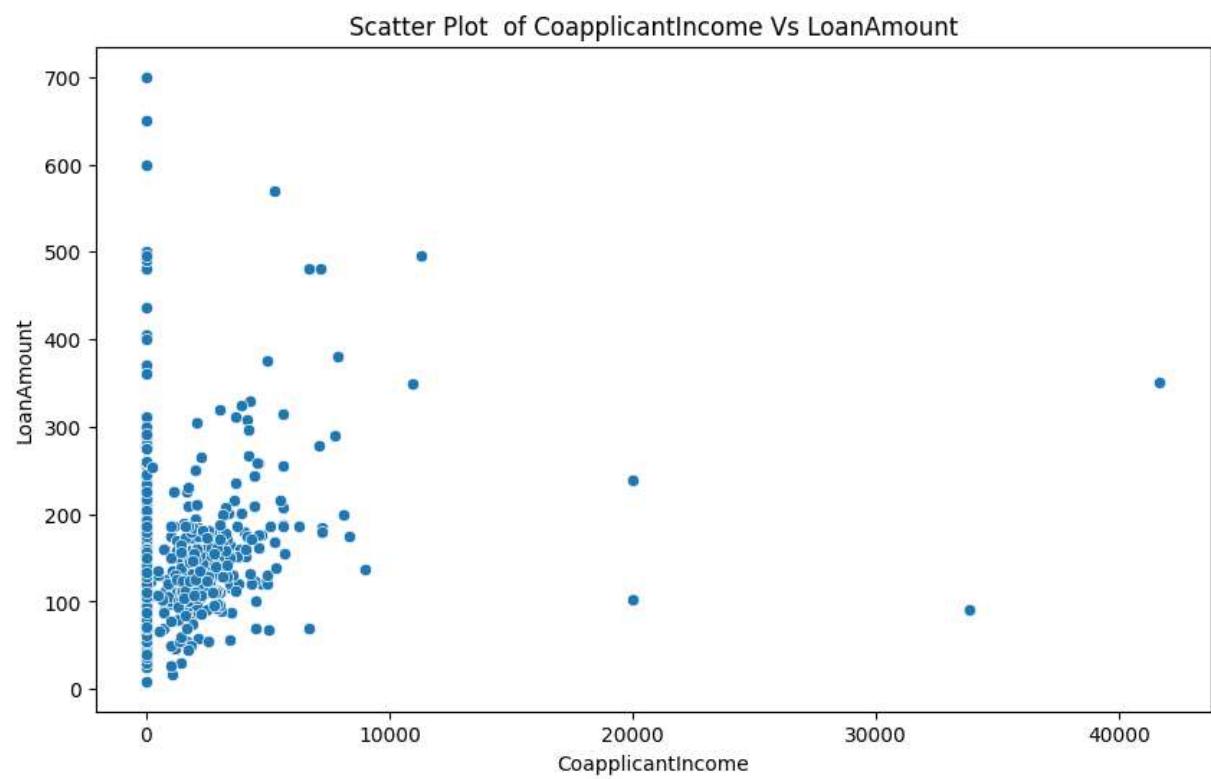
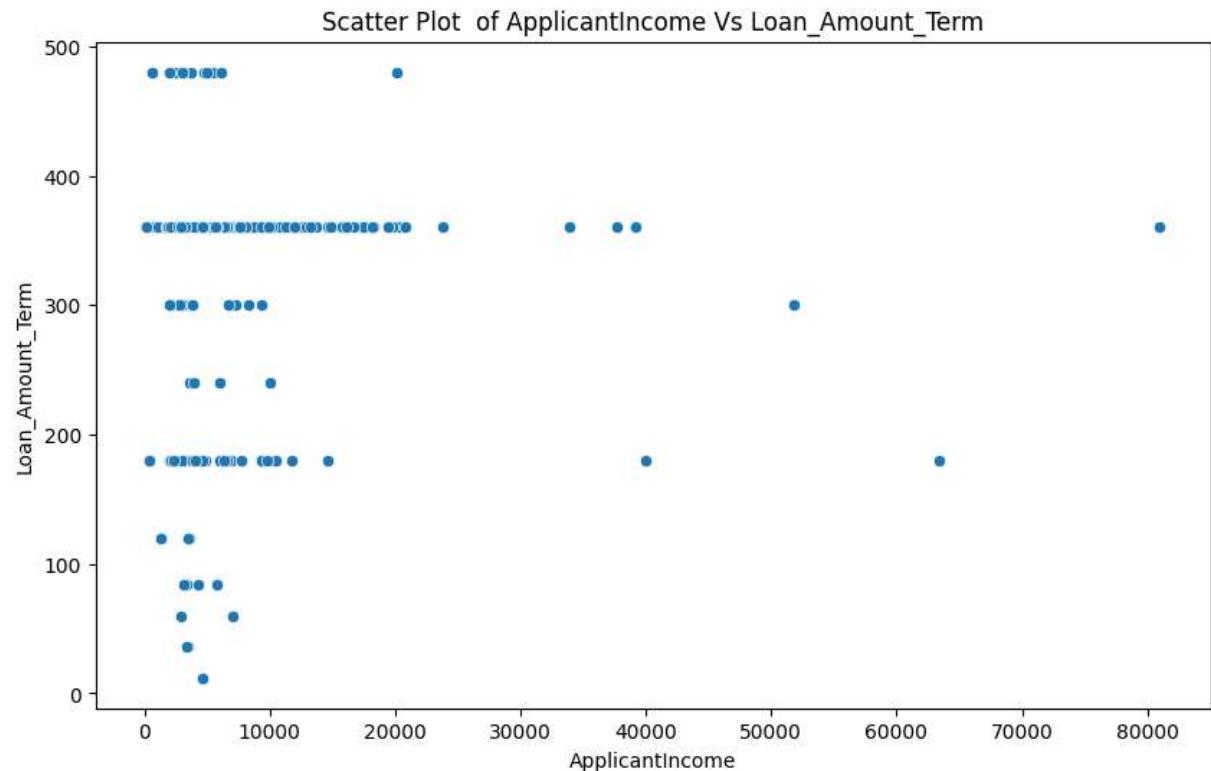
```
In [20]: #Scatter plot for Numerical Columns
for i in range(len(numerical_variables)-1):
    for j in range(i+1,len(numerical_variables)-1):
        plt.figure(figsize=(10,6))
        sns.scatterplot(x=numerical_variables[i],y=numerical_variables[j],data=df)
        plt.title(f'Scatter Plot of {numerical_variables[i]} Vs {numerical_variables[j]}')
        plt.xlabel(numerical_variables[i])
        plt.ylabel(numerical_variables[j])
        plt.show()
```

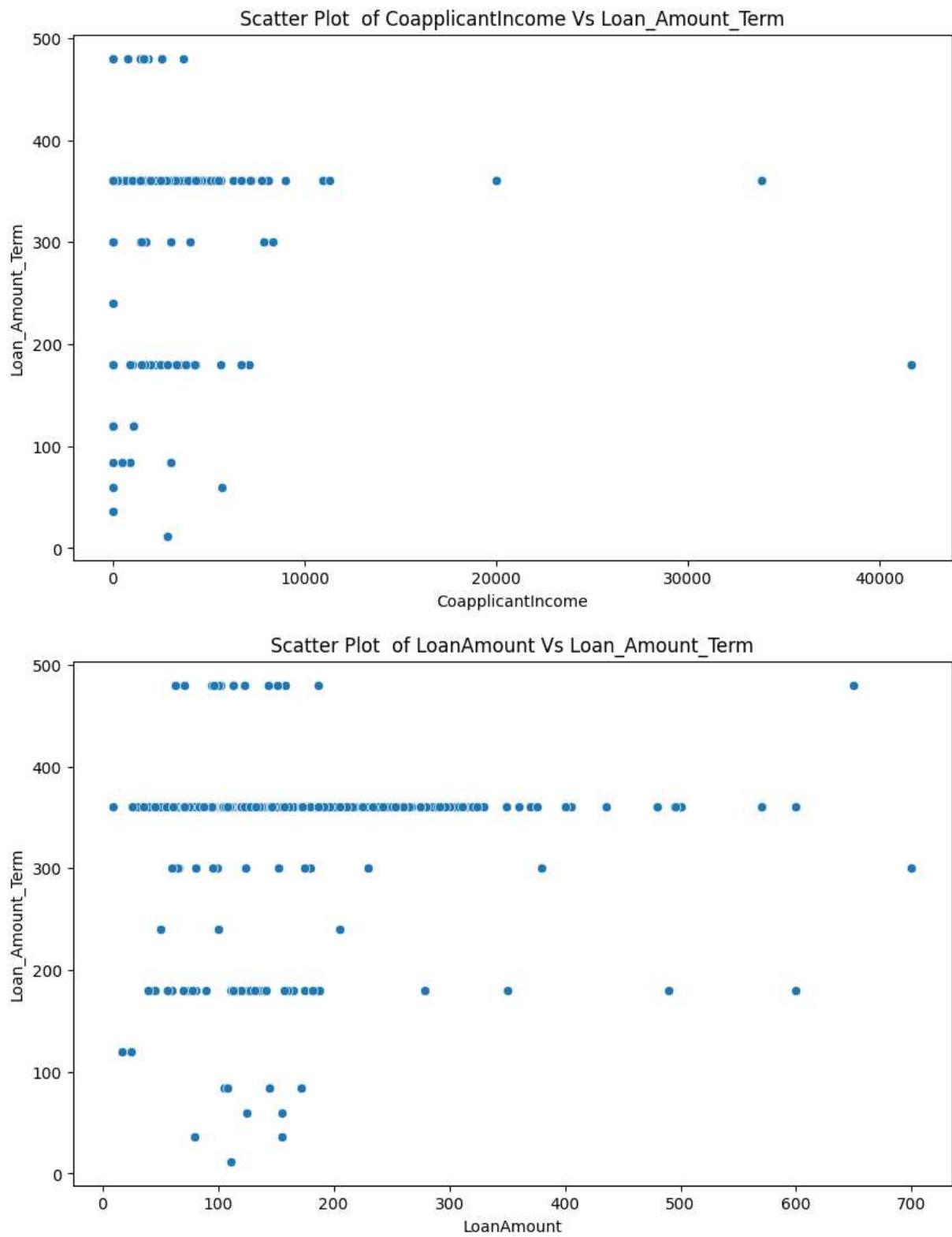
Scatter Plot of ApplicantIncome Vs CoapplicantIncome



Scatter Plot of ApplicantIncome Vs LoanAmount

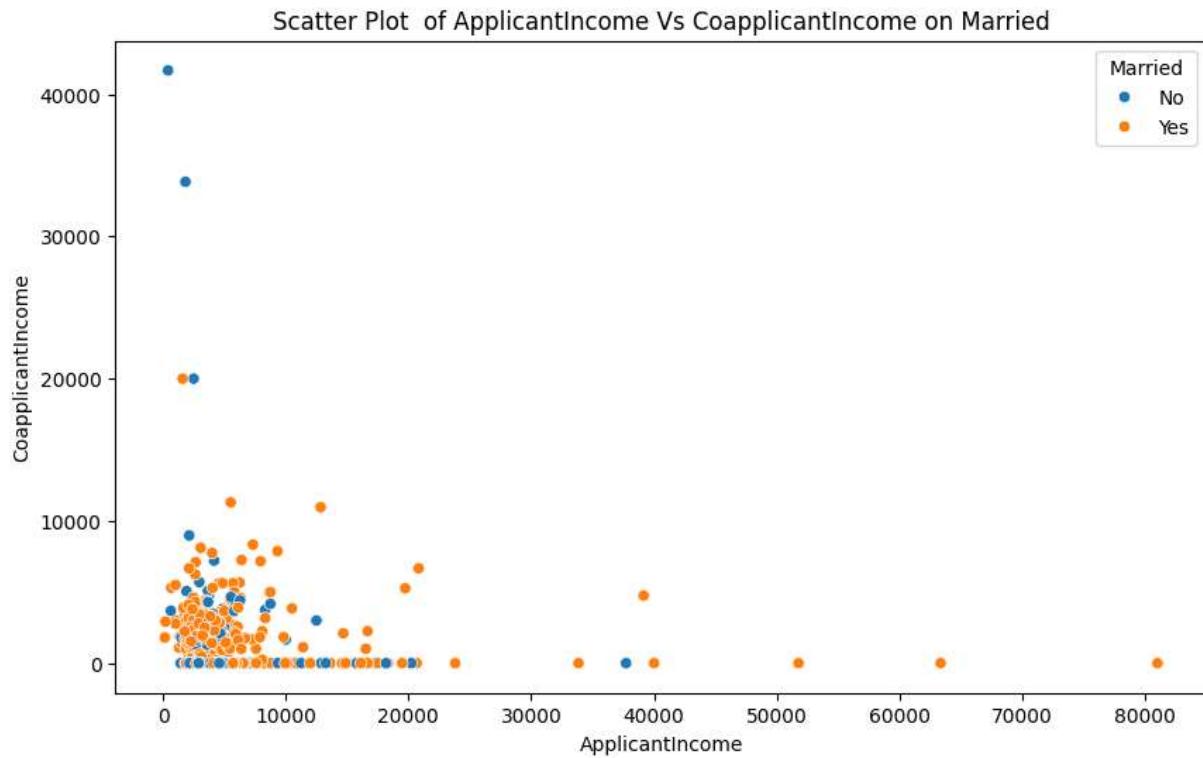
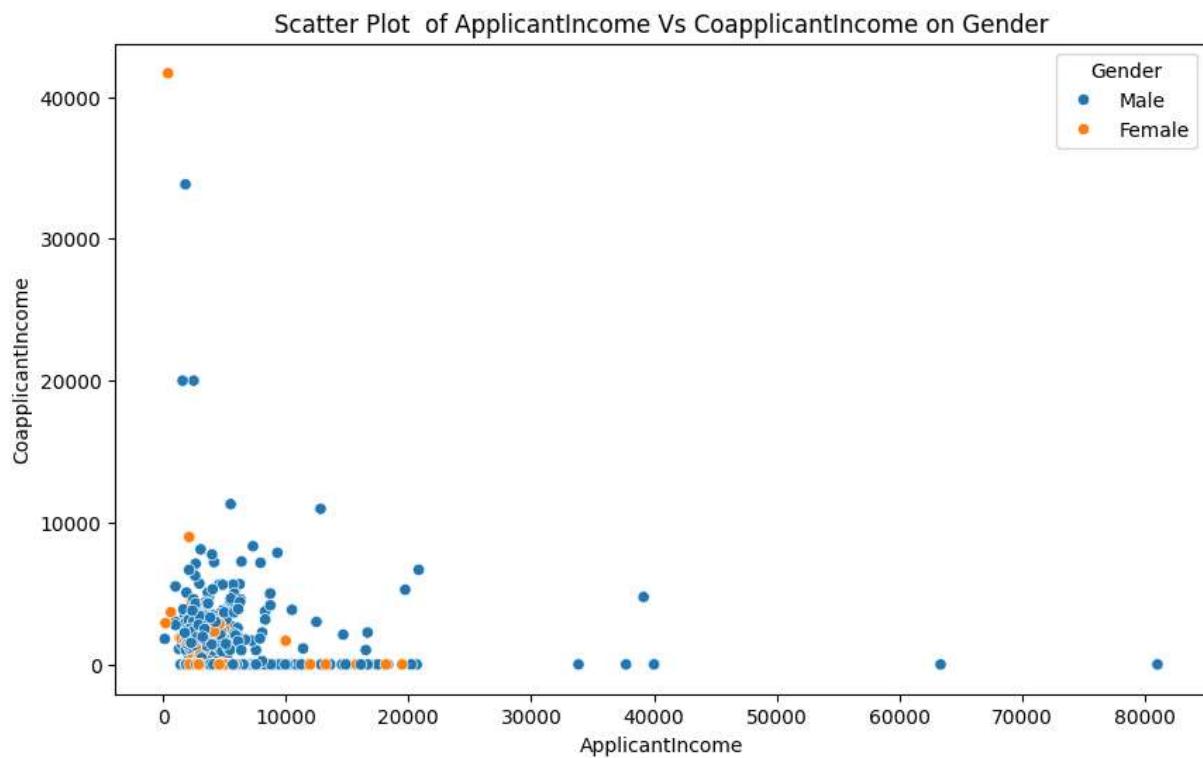


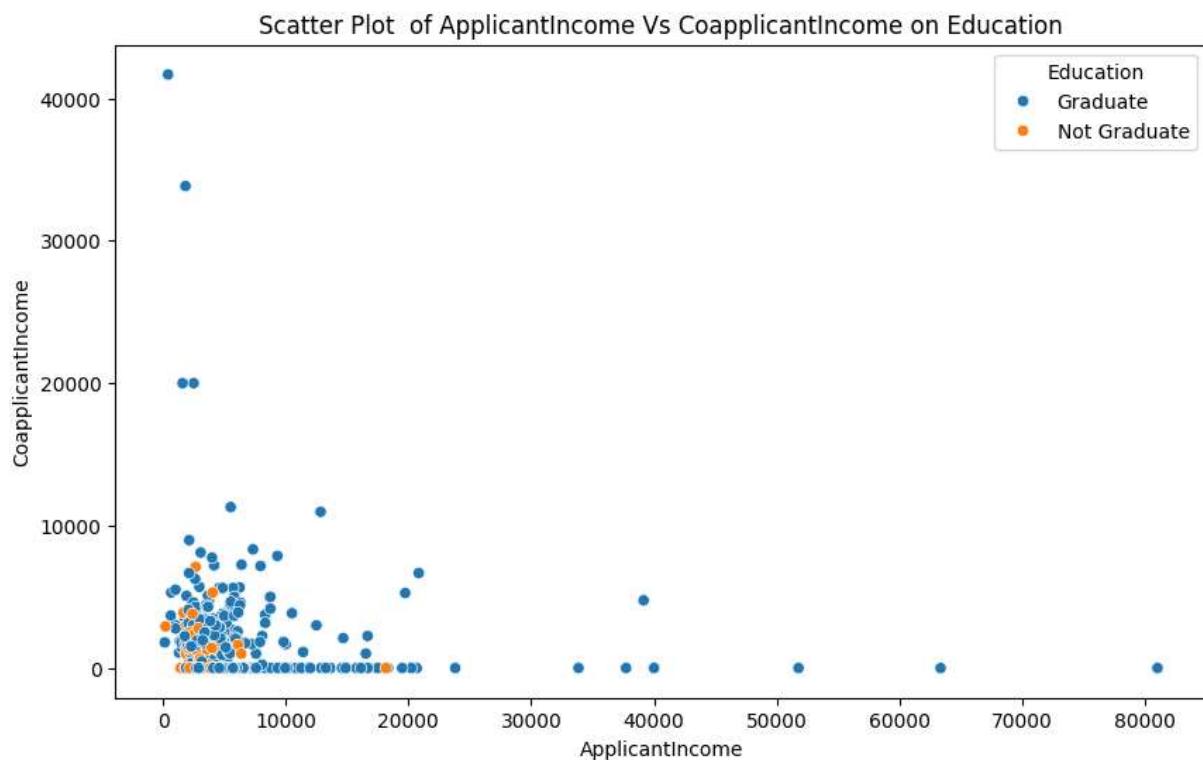
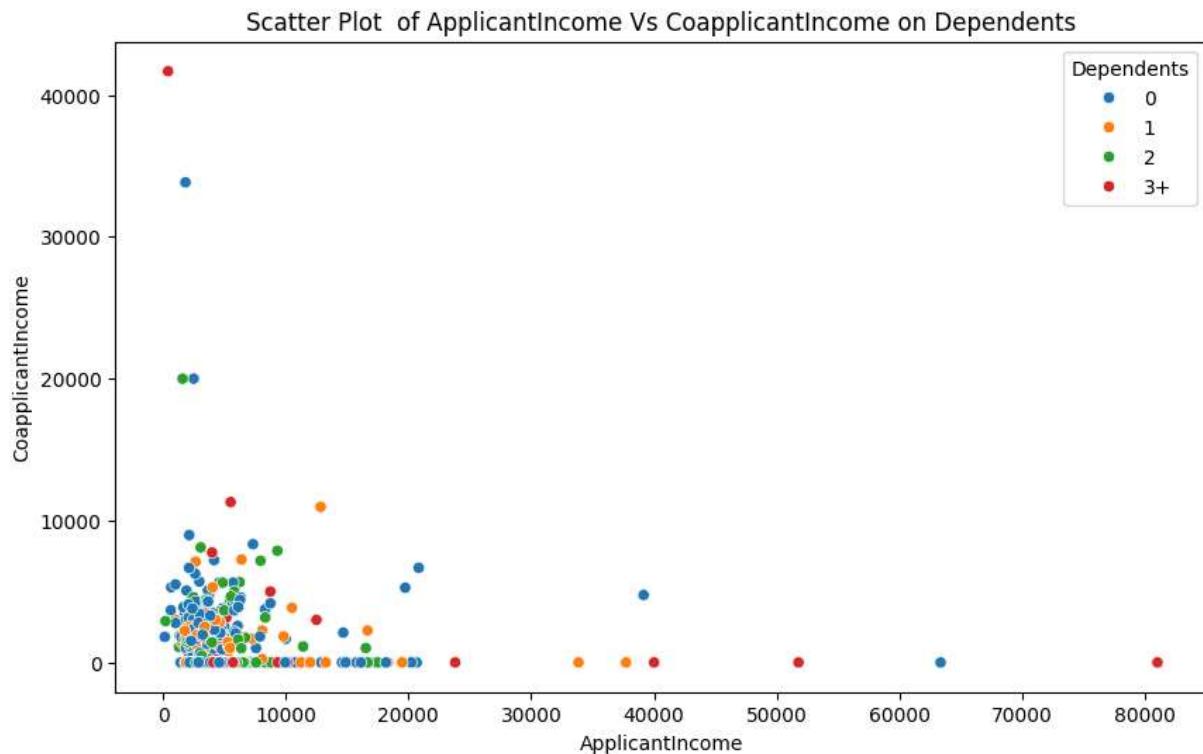


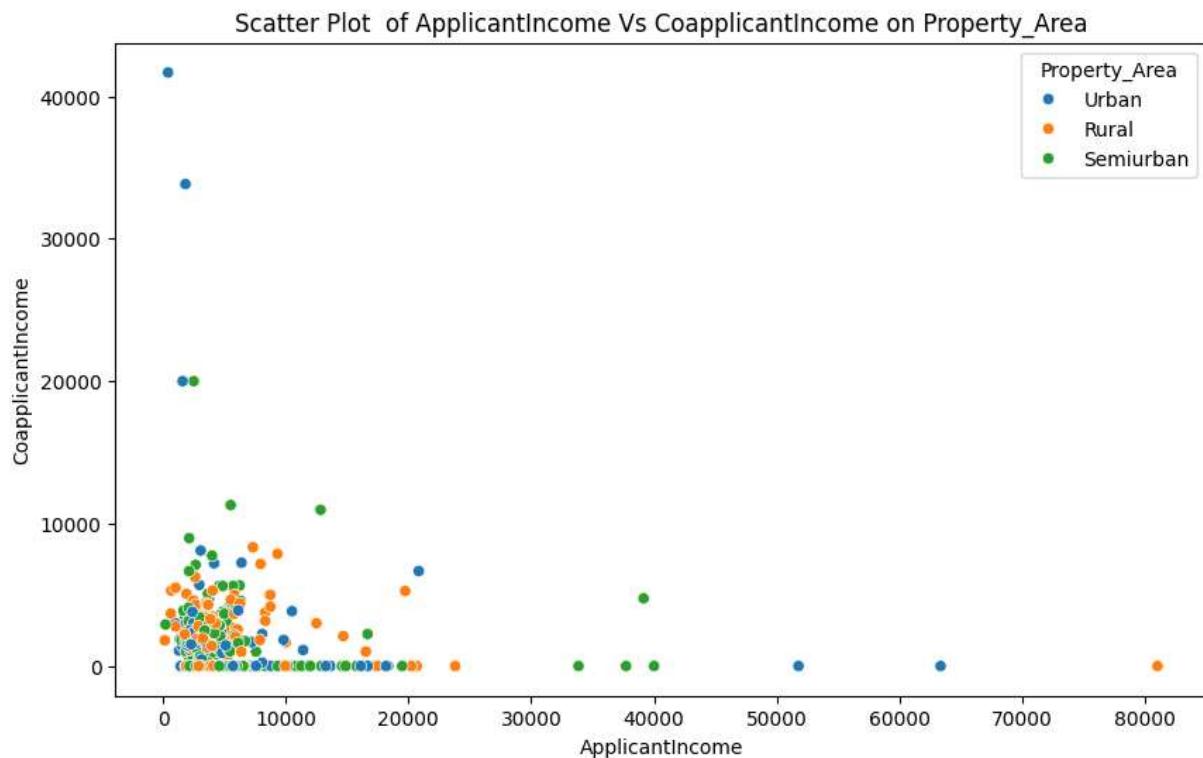
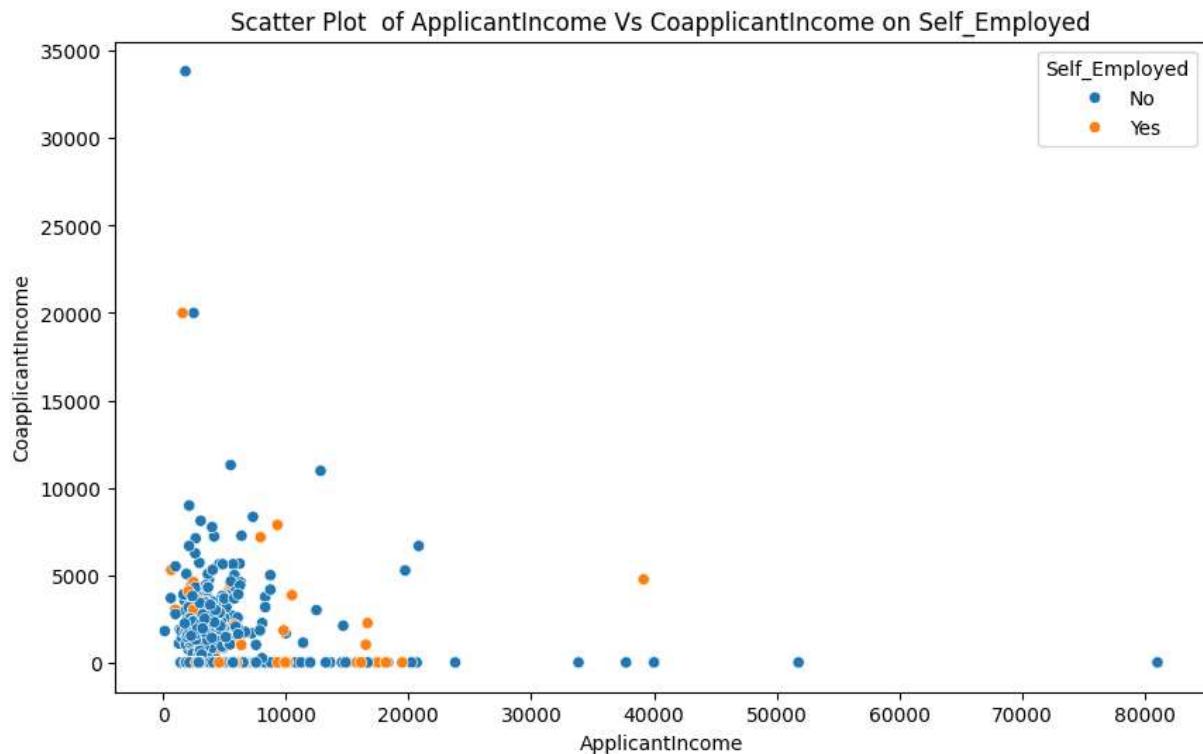


```
In [21]: #Multi Variant Analysis
#Scatter plot for Numerical Columns
for i in range(len(numerical_variables)-1):
    for j in range(i+1,len(numerical_variables)-1):
        for k in range(len(categorical_variables)-1):
            plt.figure(figsize=(10,6))
            sns.scatterplot(x=numerical_variables[i],y=numerical_variables[j],hue=c
            plt.title(f'Scatter Plot of {numerical_variables[i]} Vs {numerical_var
```

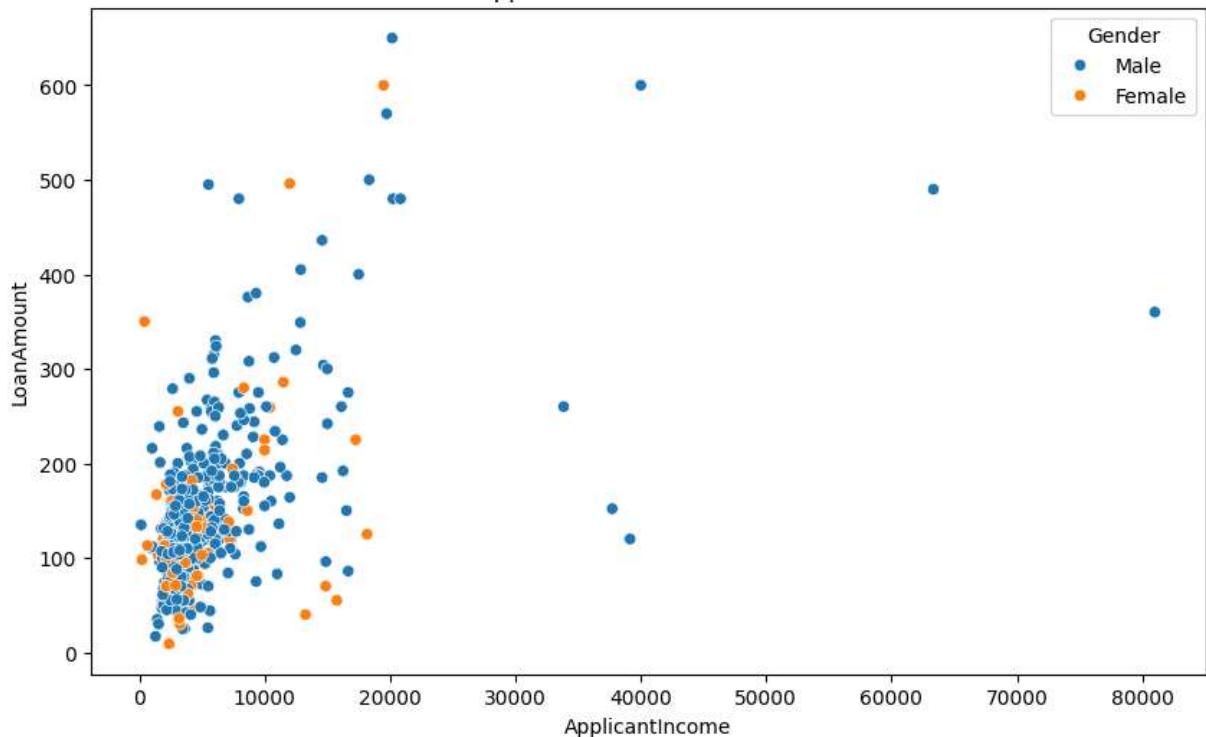
```
plt.xlabel(numerical_variables[i])
plt.ylabel(numerical_variables[j])
plt.show()
```



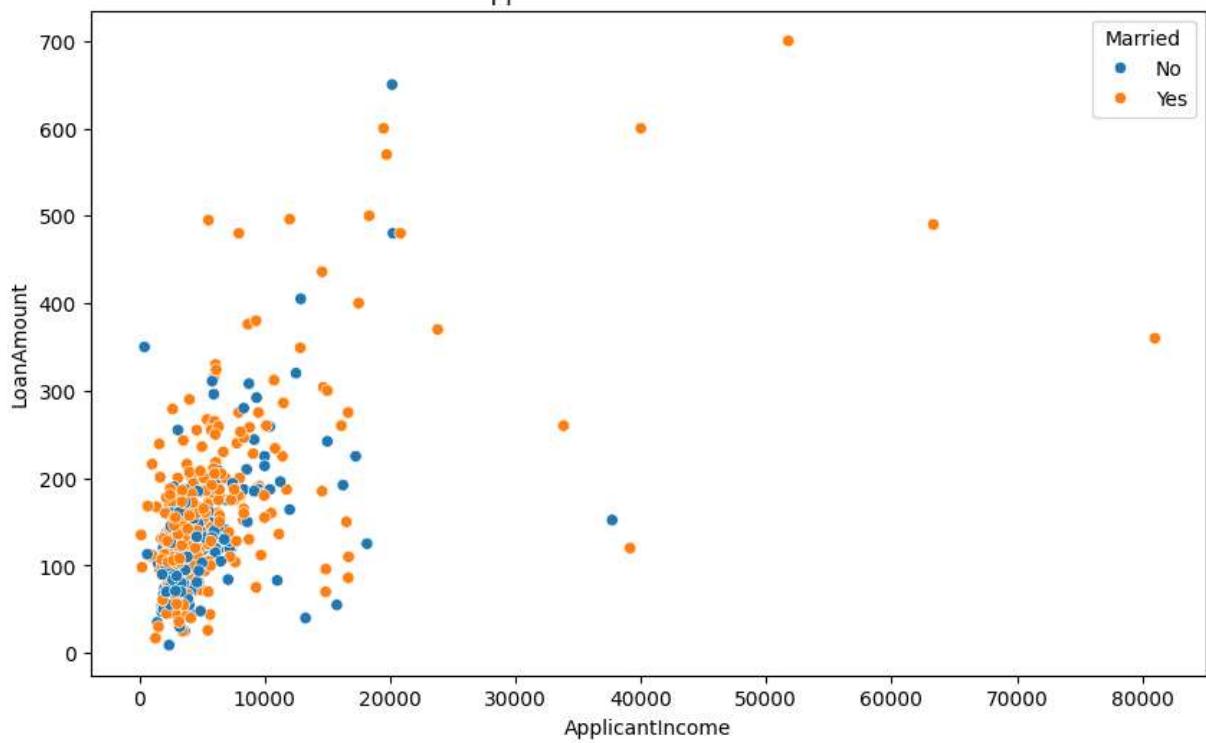




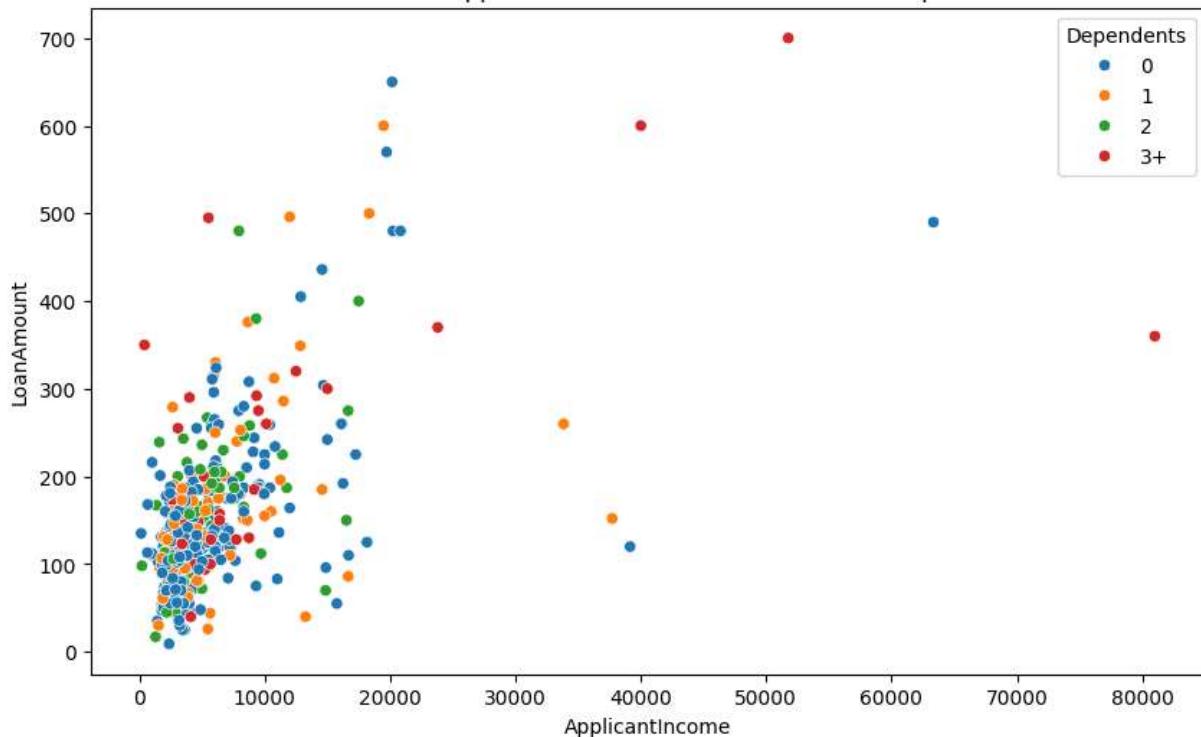
Scatter Plot of ApplicantIncome Vs LoanAmount on Gender



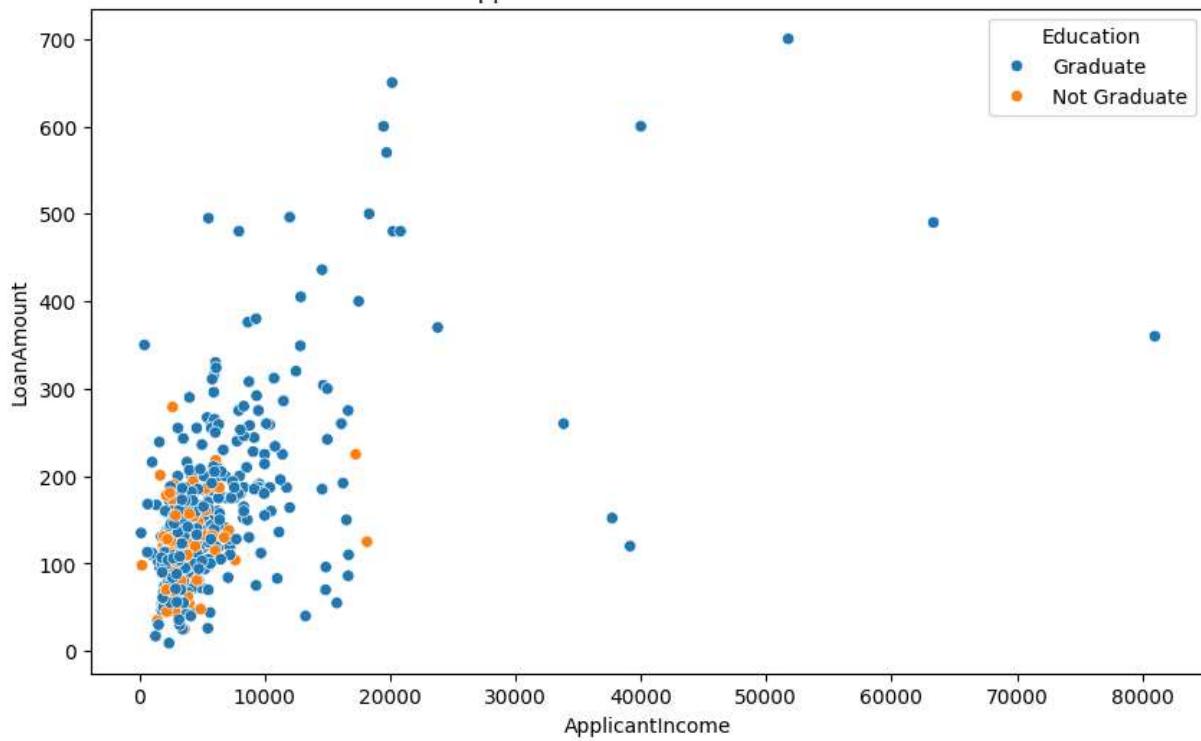
Scatter Plot of ApplicantIncome Vs LoanAmount on Married



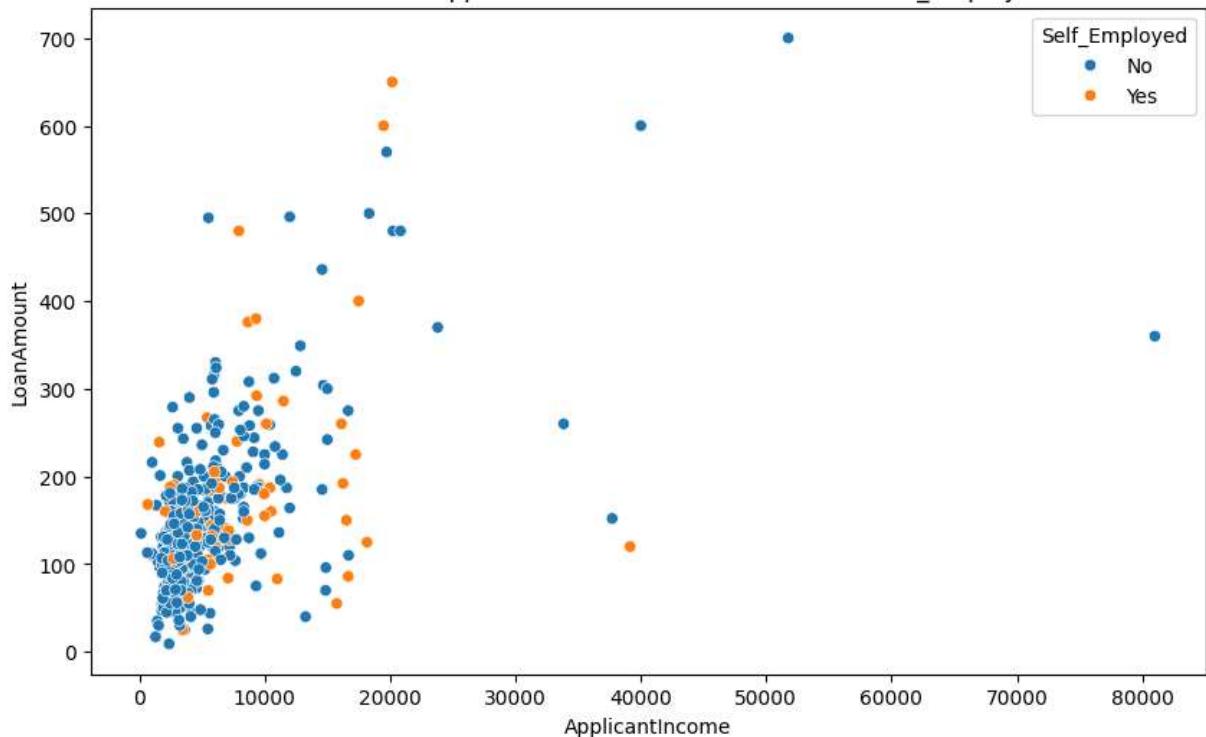
Scatter Plot of ApplicantIncome Vs LoanAmount on Dependents



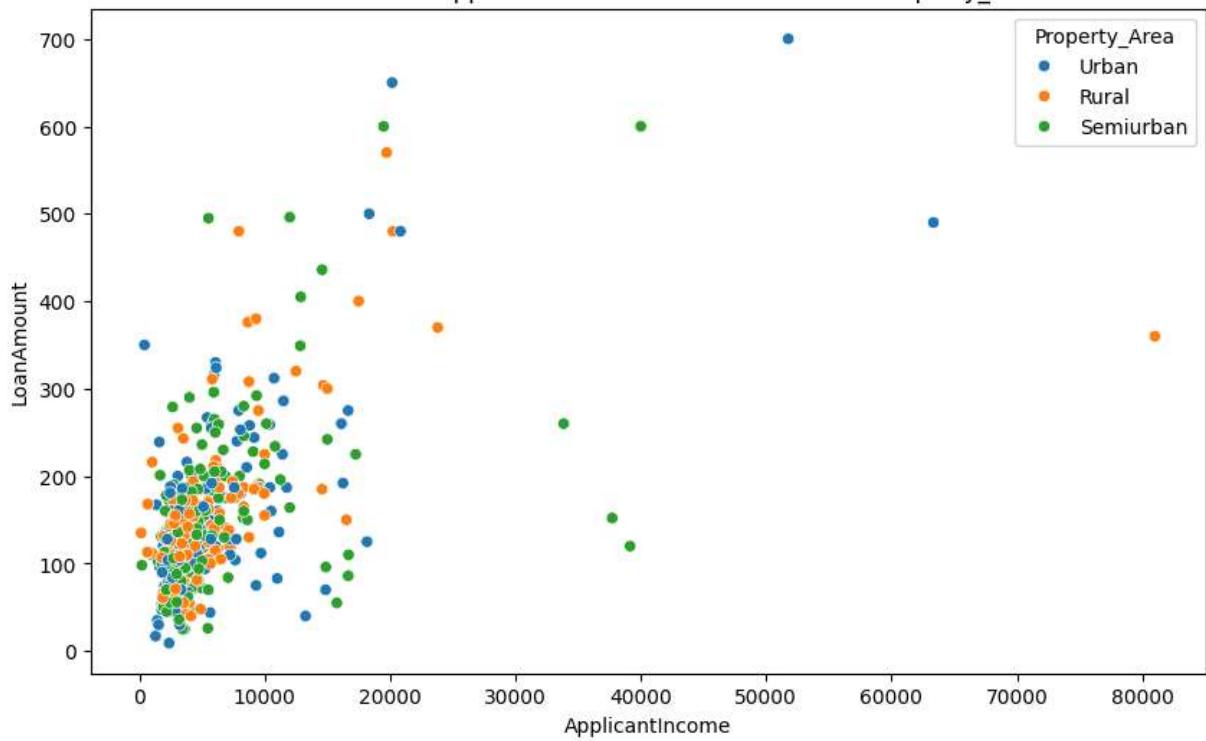
Scatter Plot of ApplicantIncome Vs LoanAmount on Education

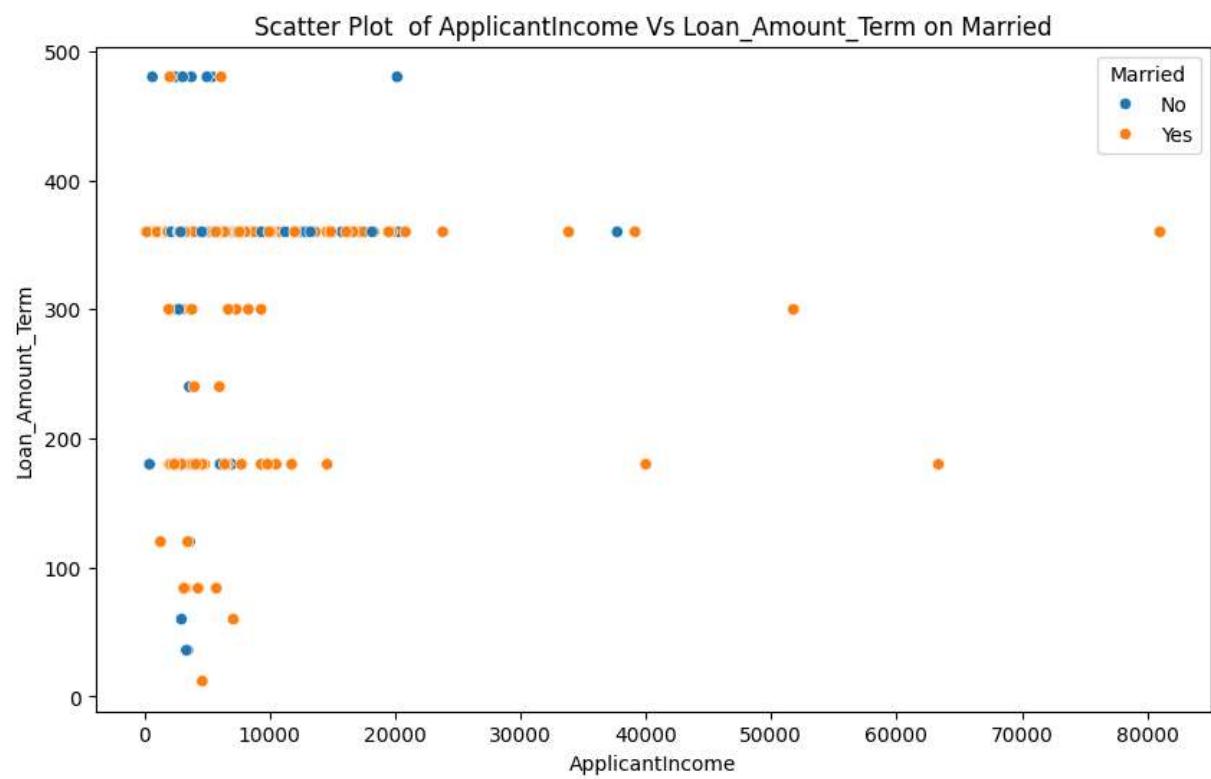
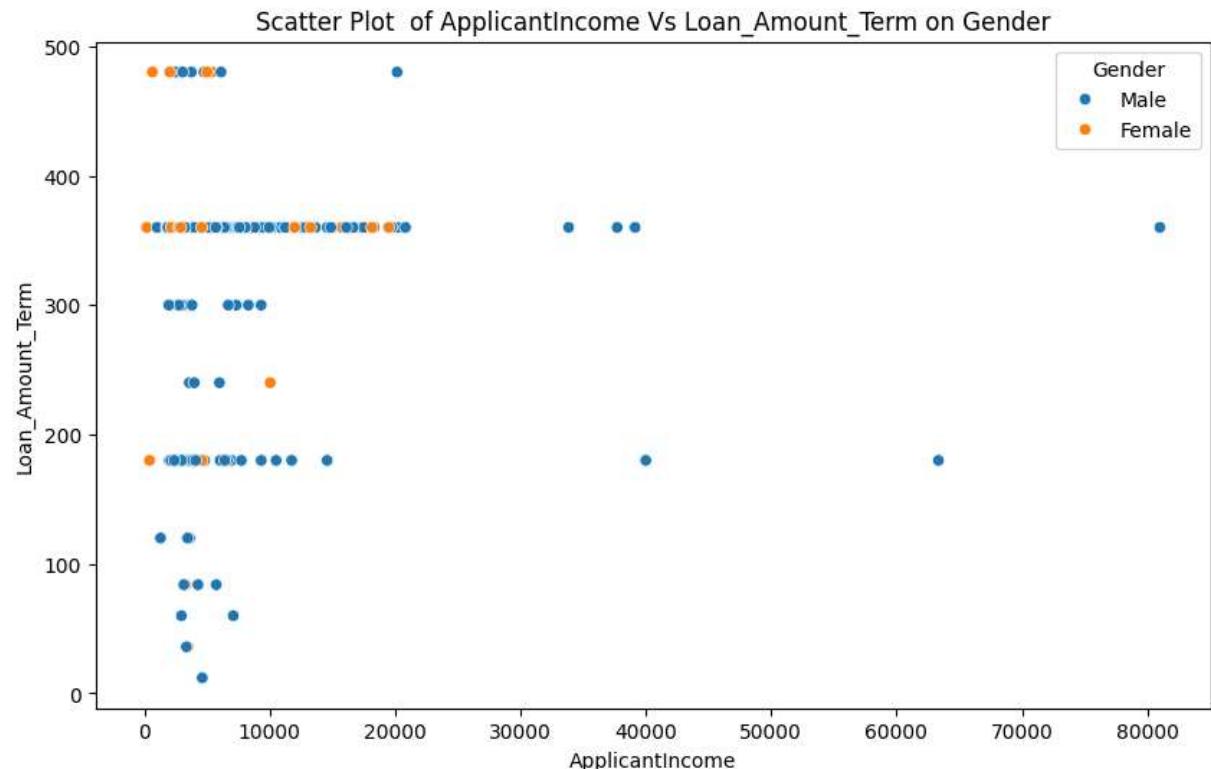


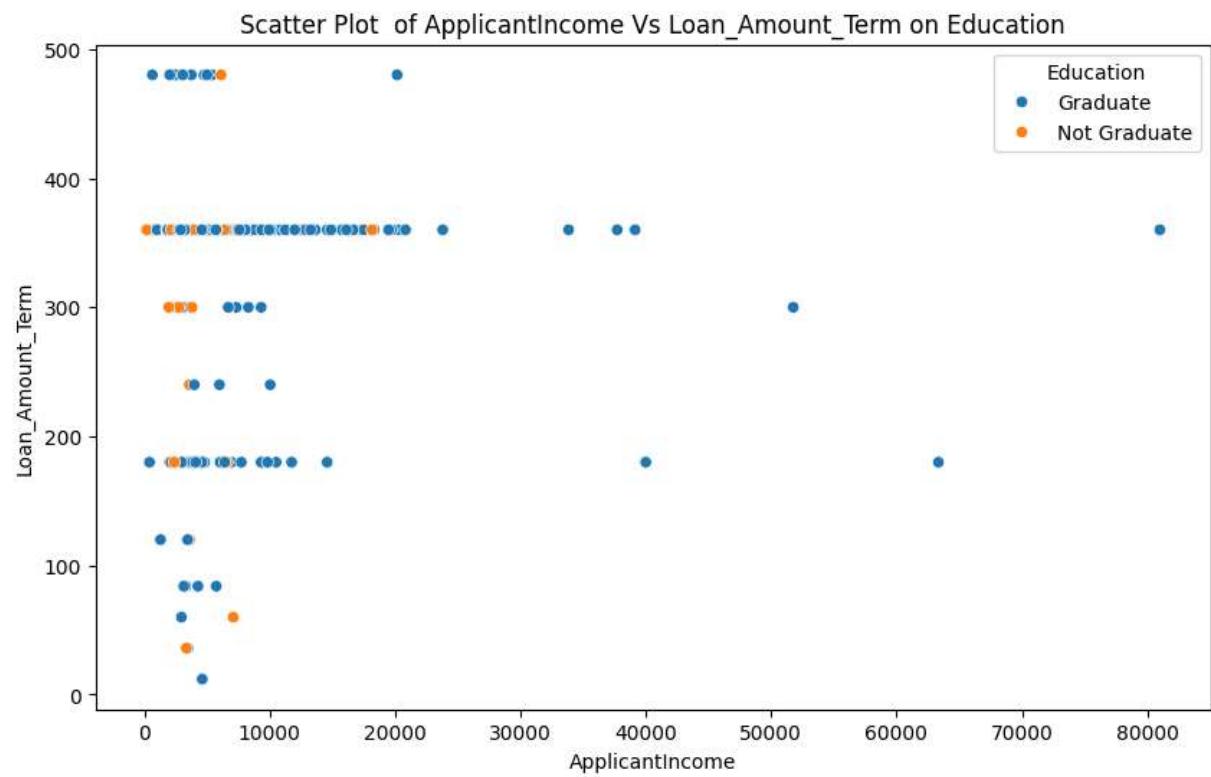
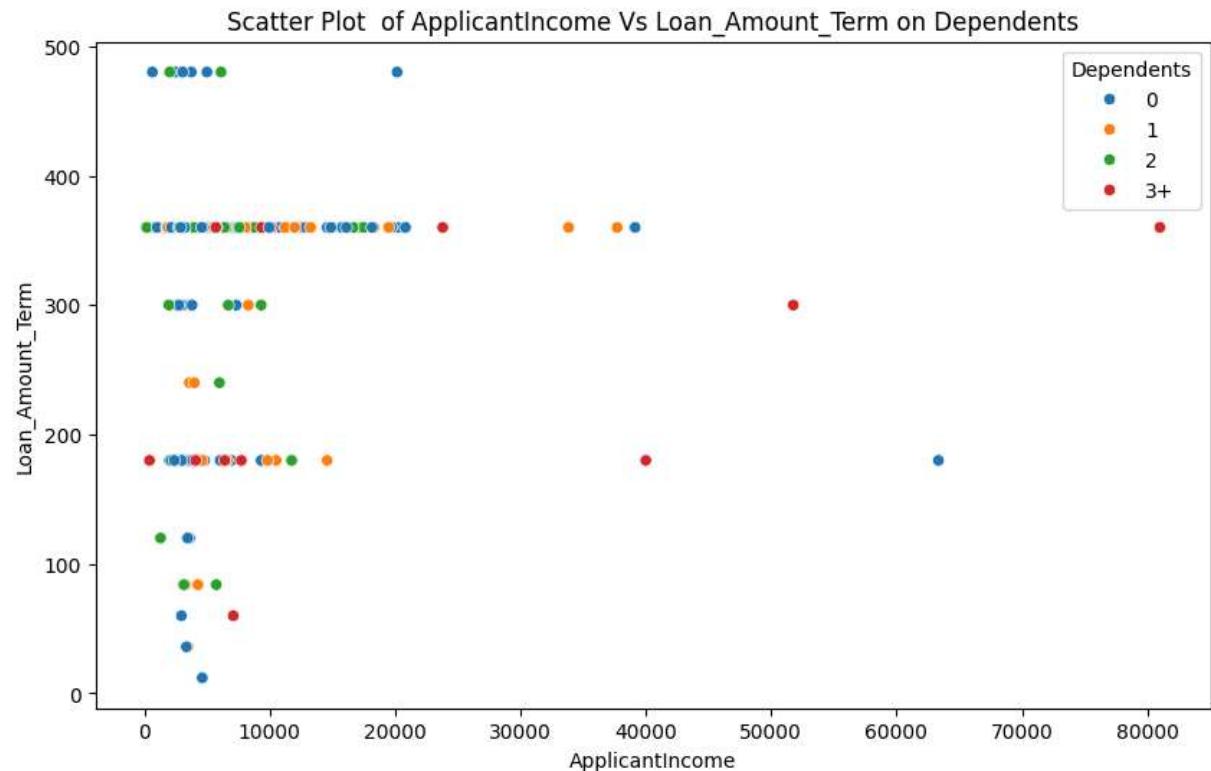
Scatter Plot of ApplicantIncome Vs LoanAmount on Self_Employed

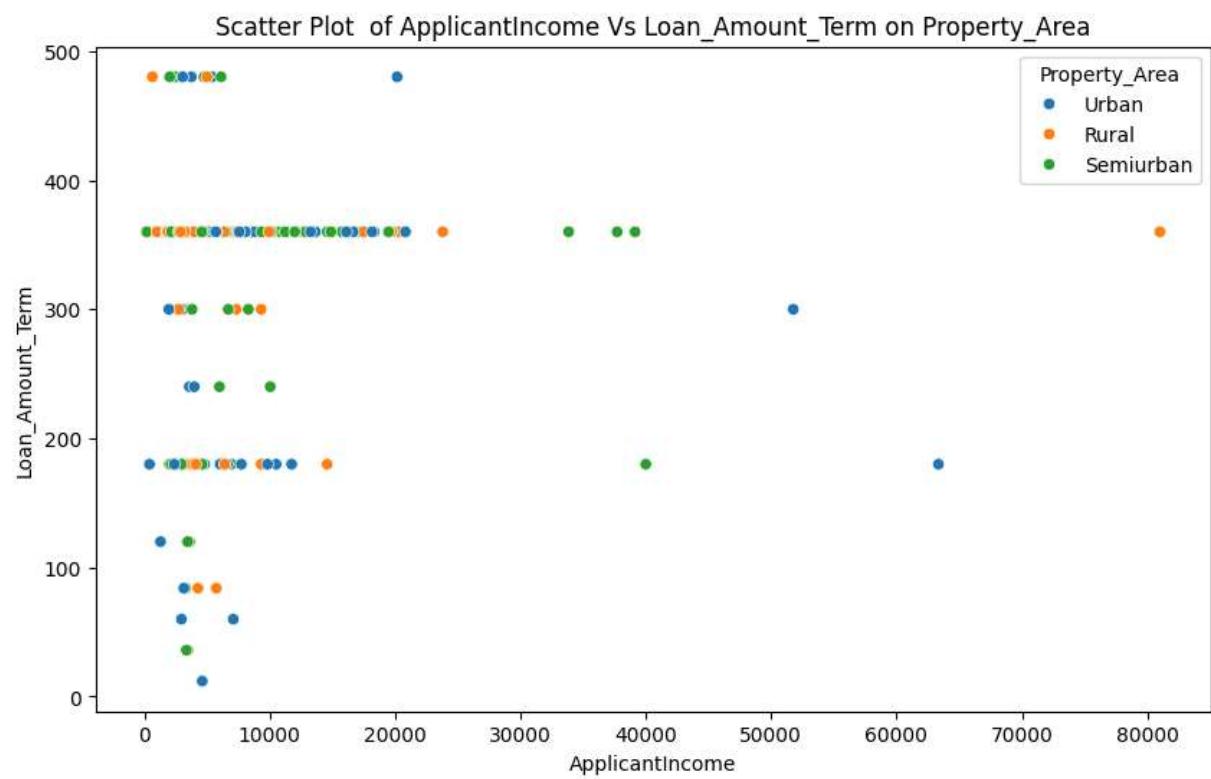
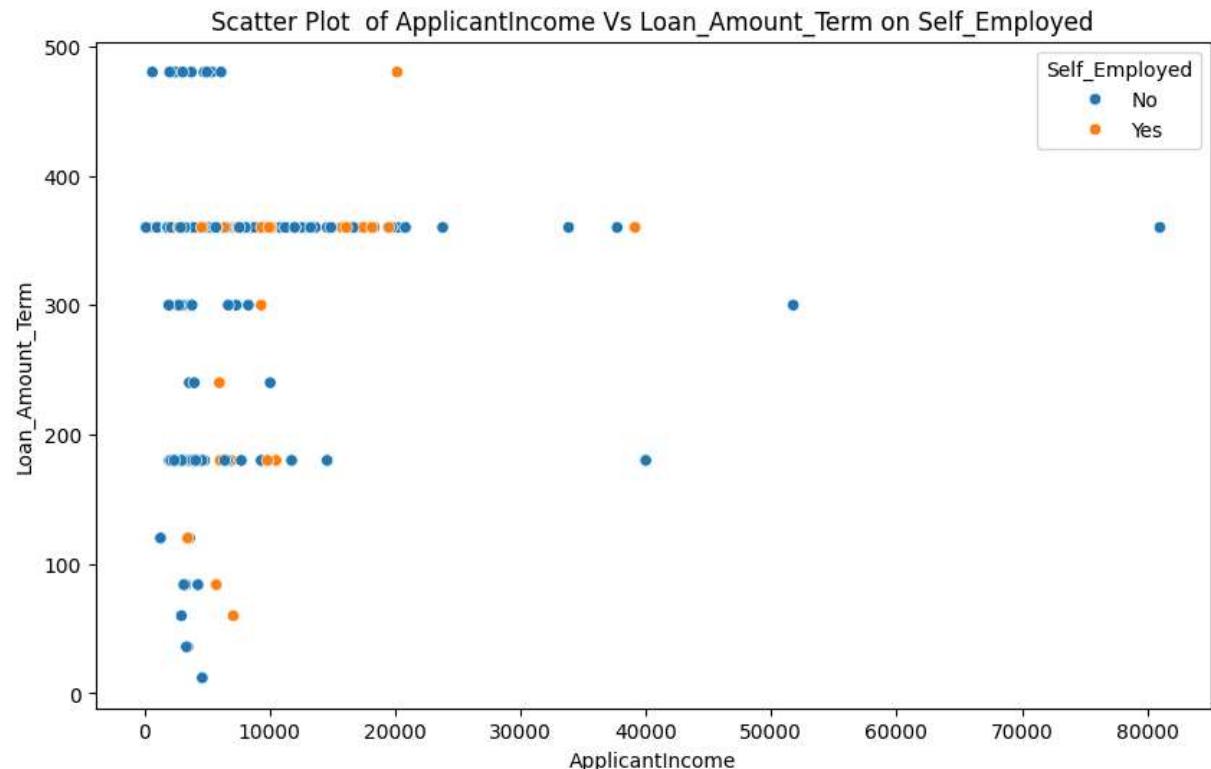


Scatter Plot of ApplicantIncome Vs LoanAmount on Property_Area

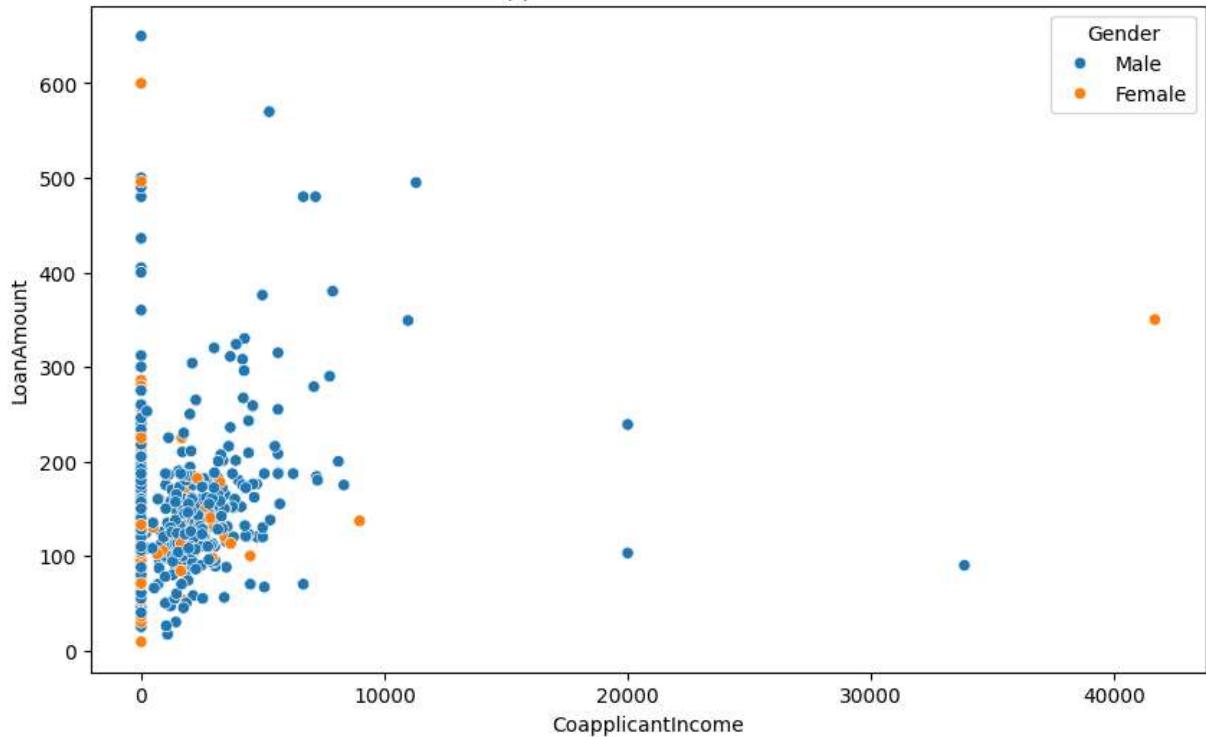




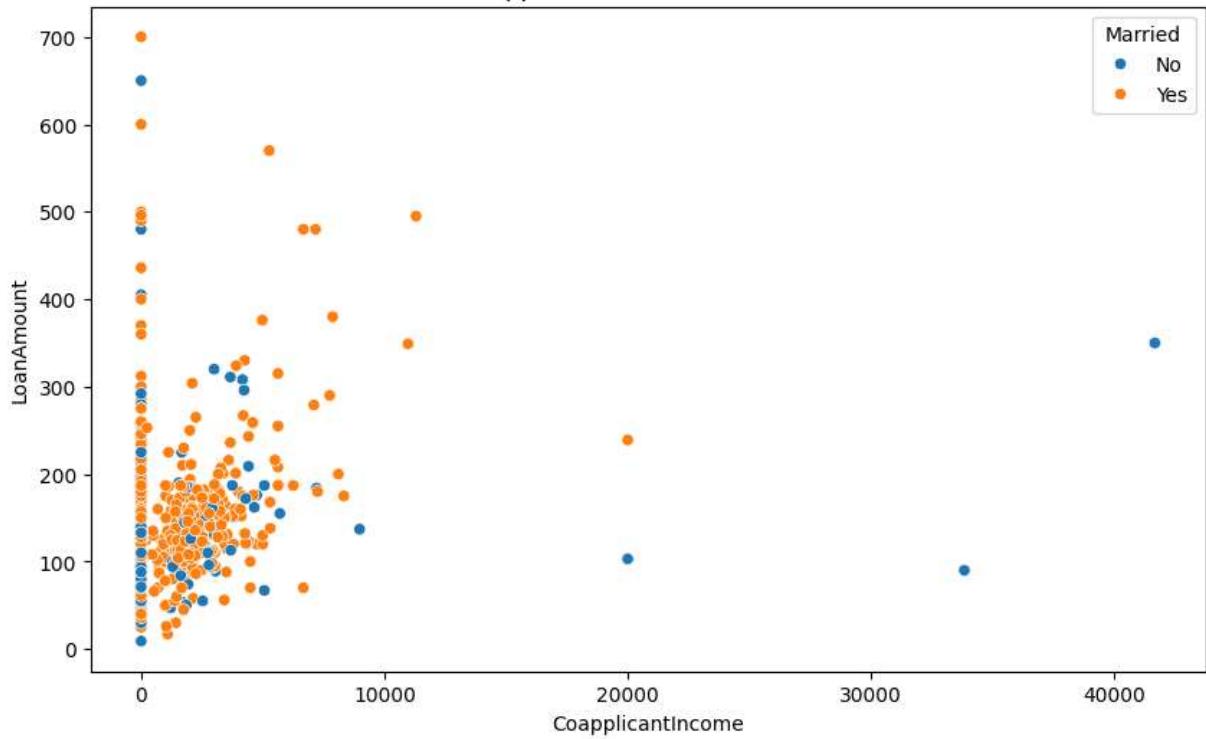




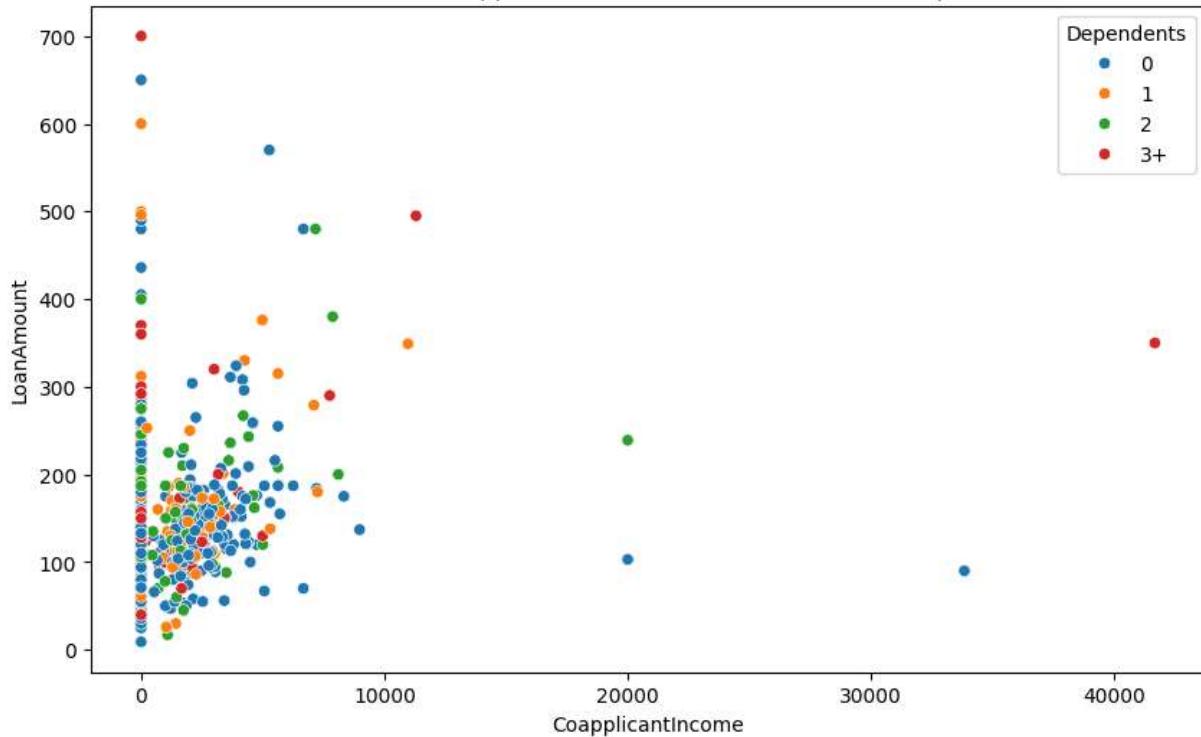
Scatter Plot of CoapplicantIncome Vs LoanAmount on Gender



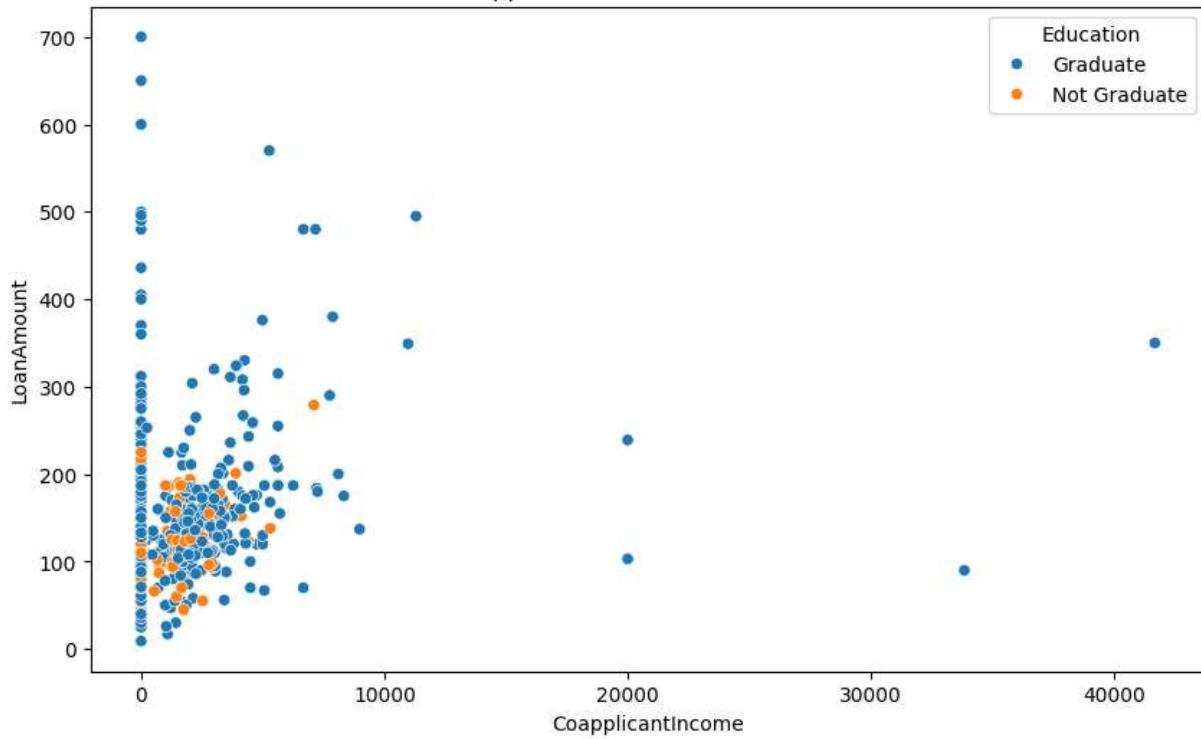
Scatter Plot of CoapplicantIncome Vs LoanAmount on Married



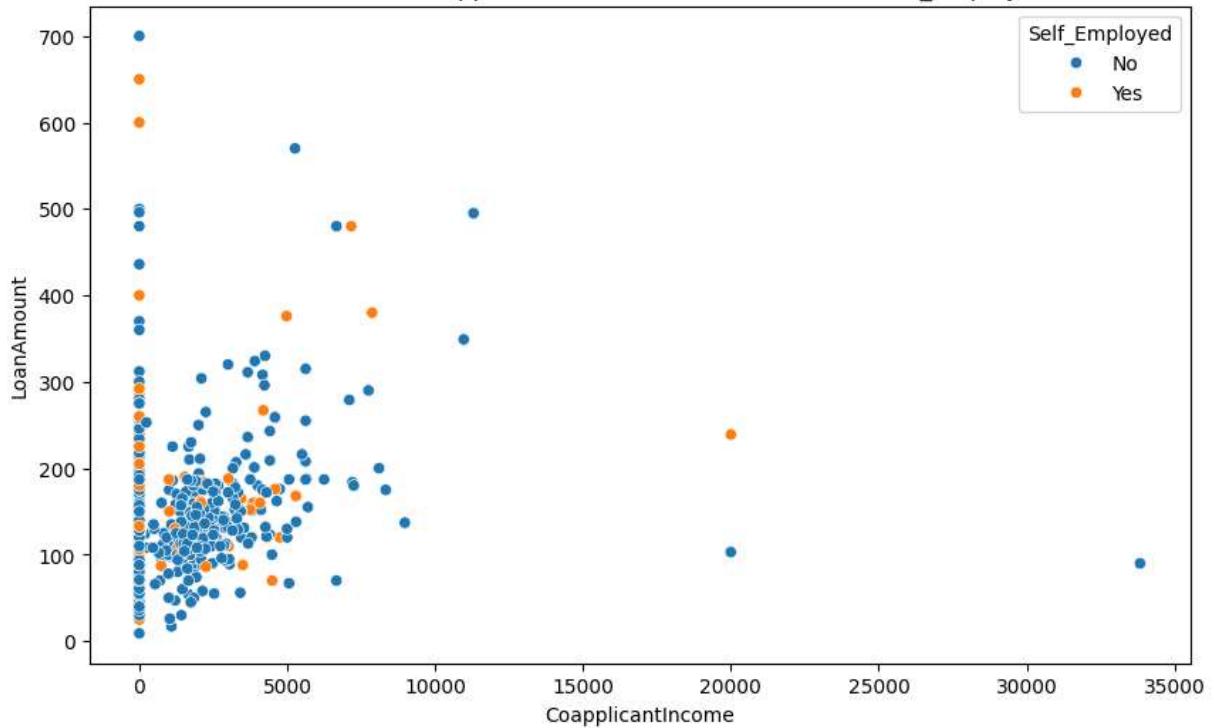
Scatter Plot of CoapplicantIncome Vs LoanAmount on Dependents



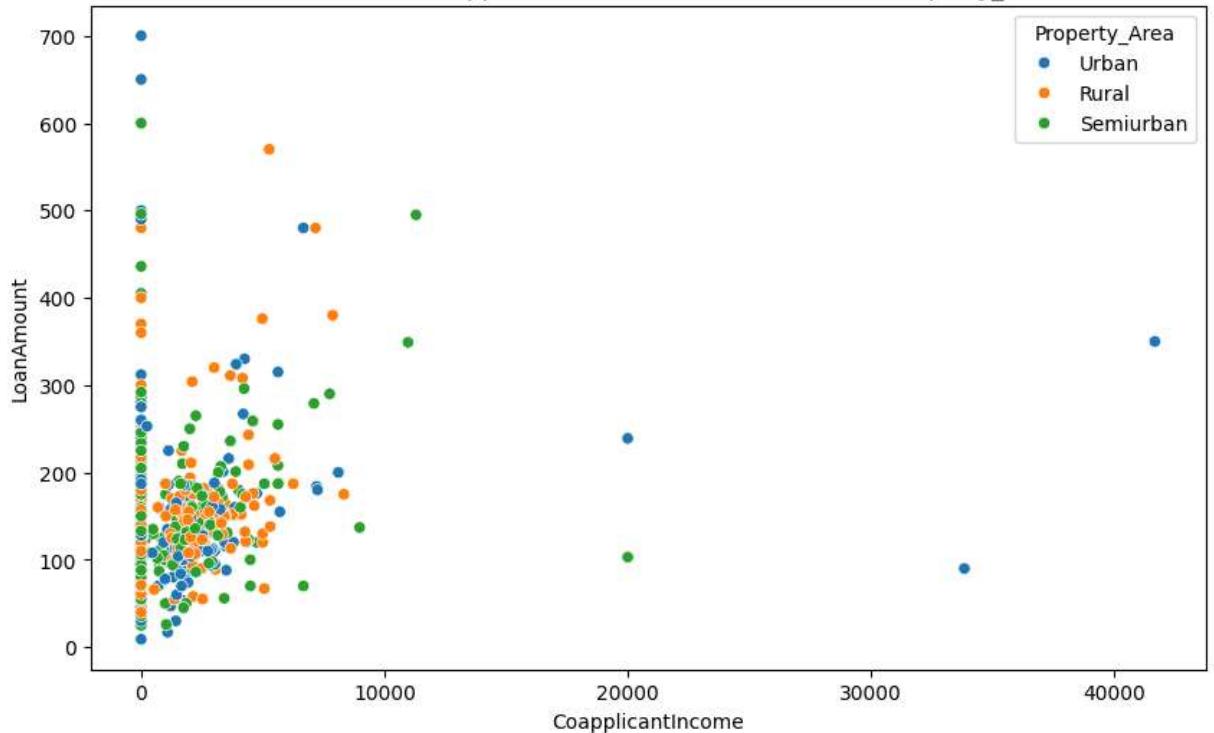
Scatter Plot of CoapplicantIncome Vs LoanAmount on Education

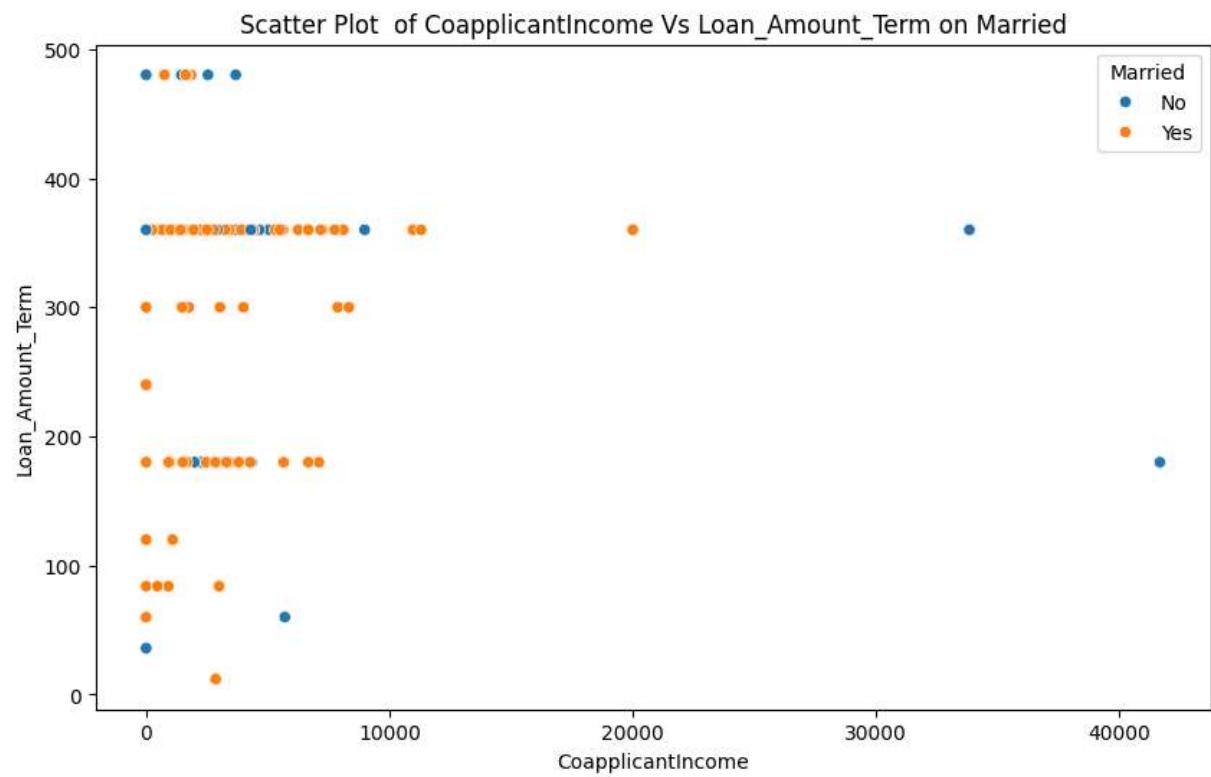
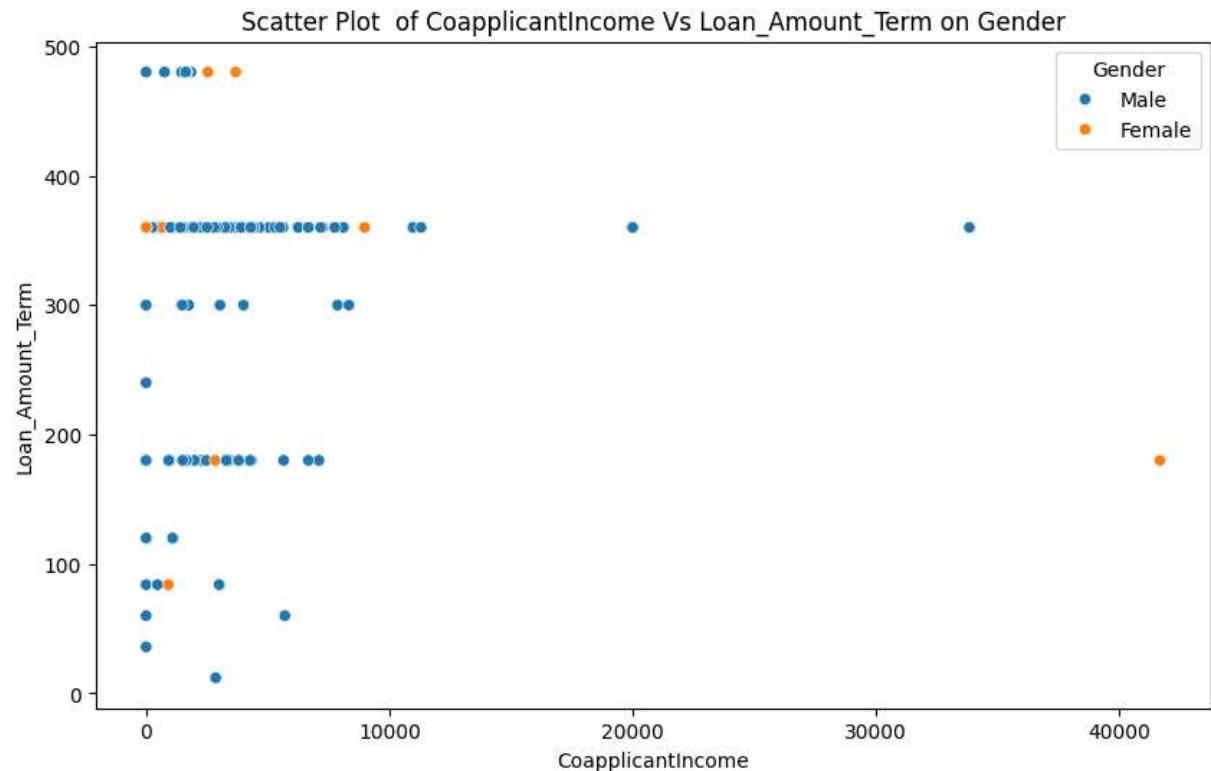


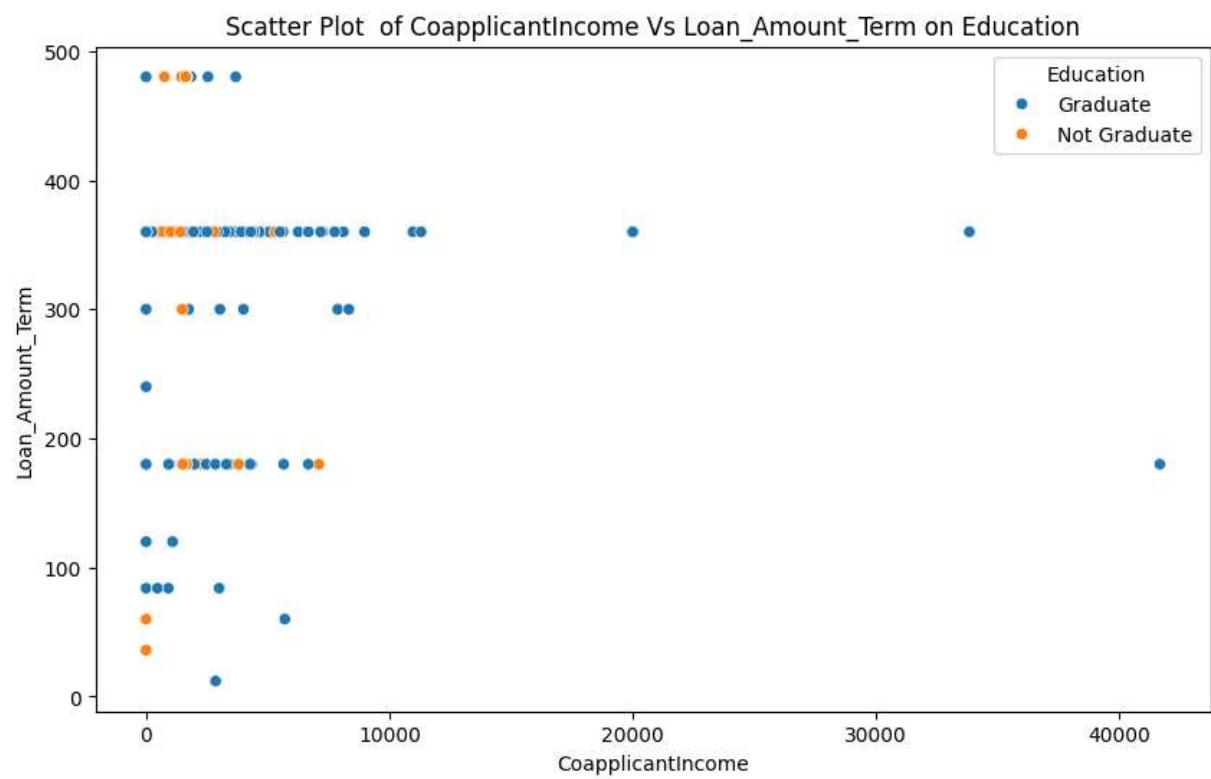
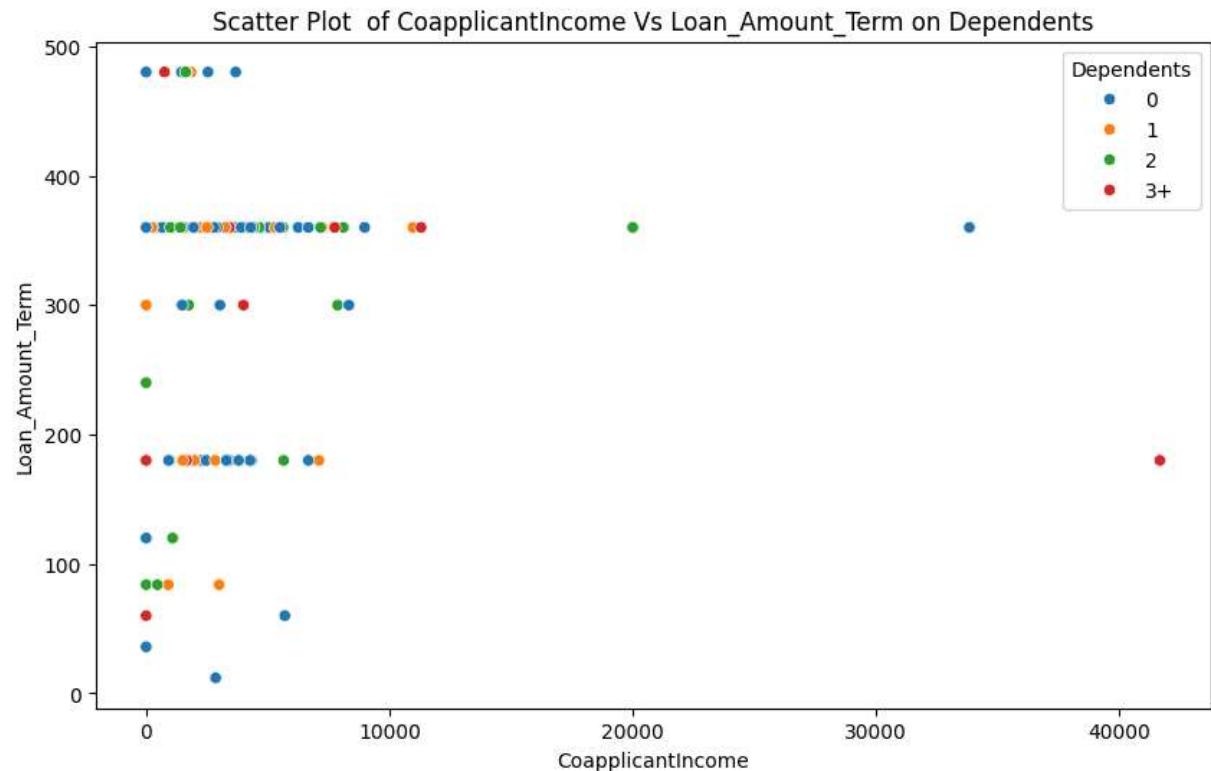
Scatter Plot of CoapplicantIncome Vs LoanAmount on Self_Employed

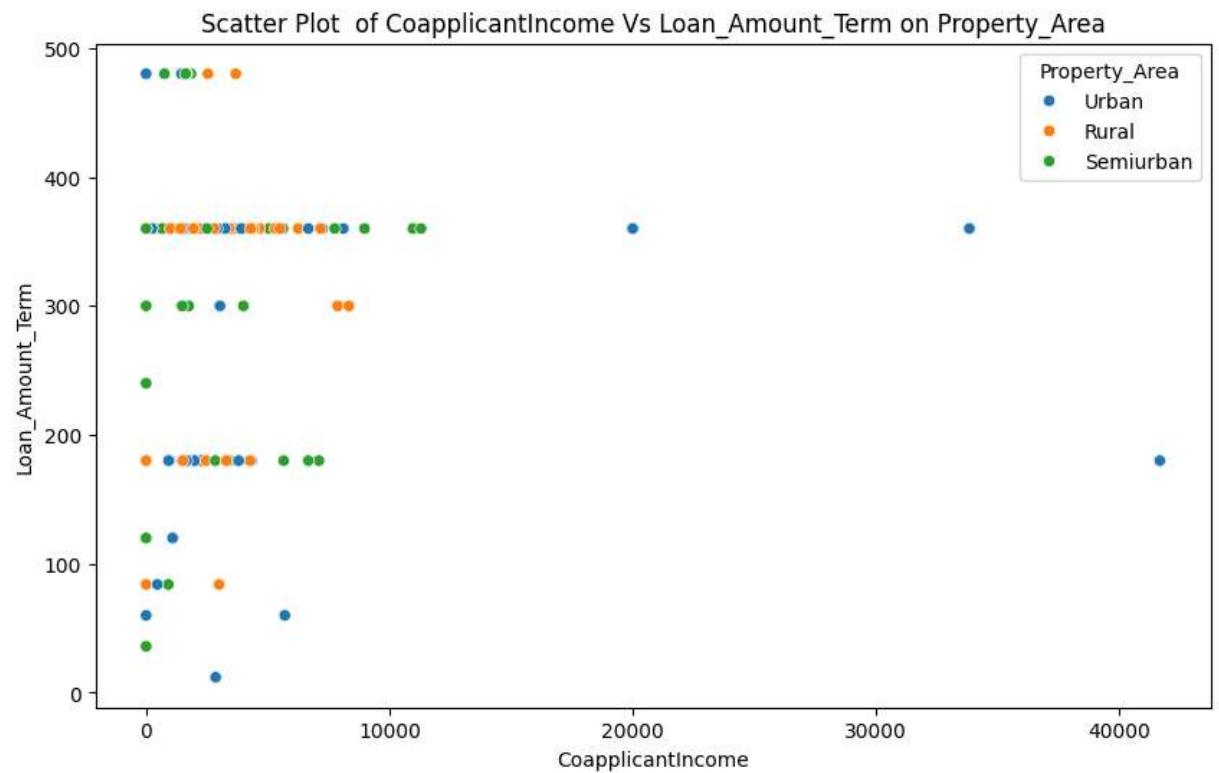
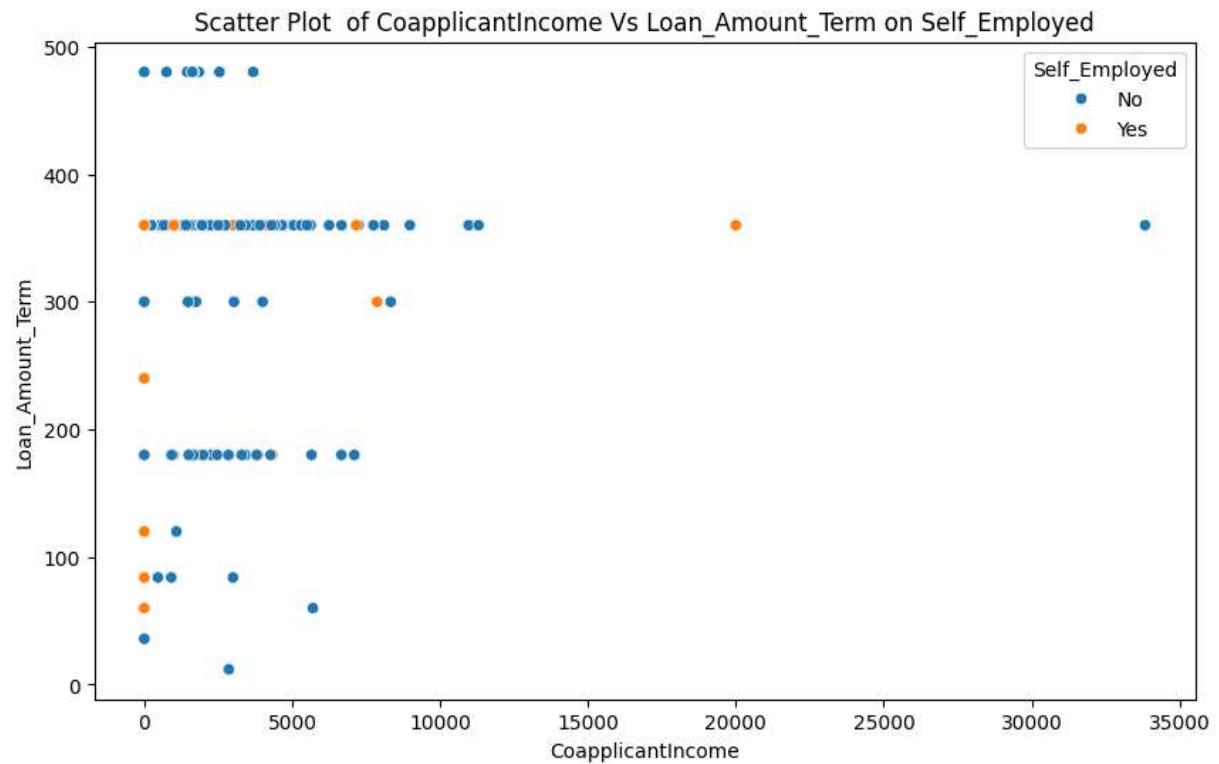


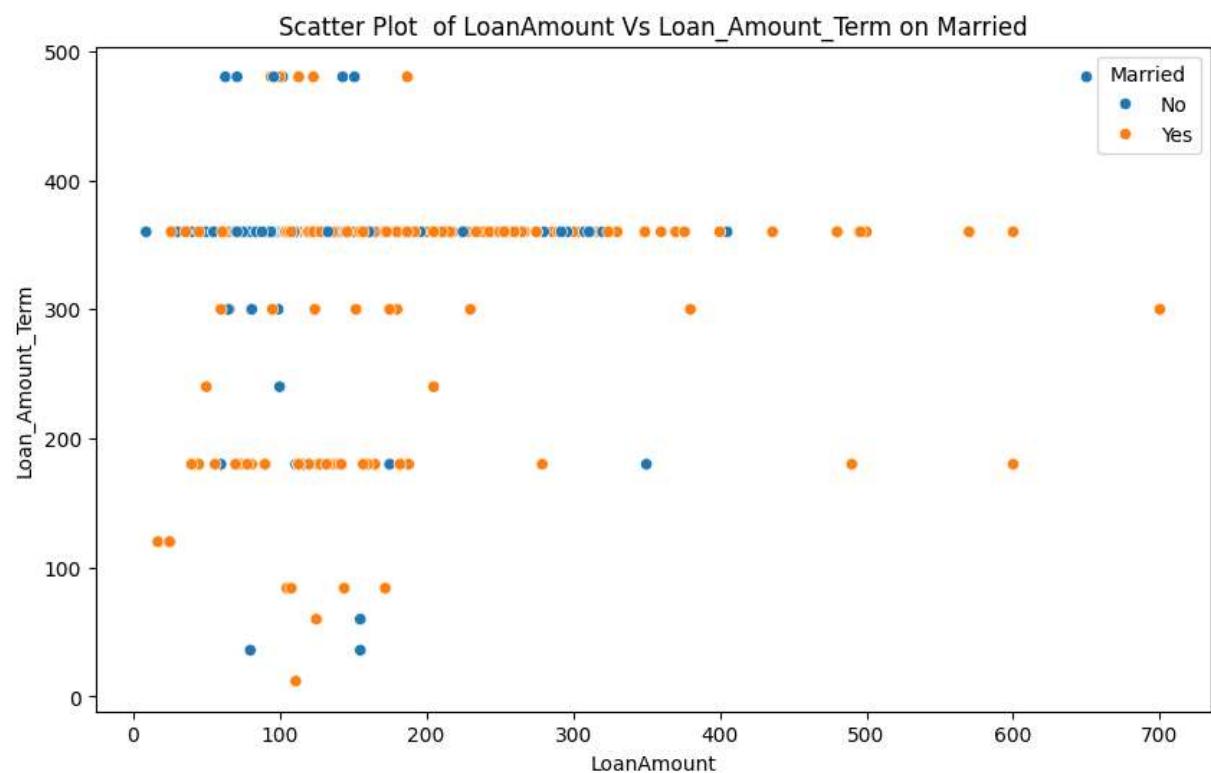
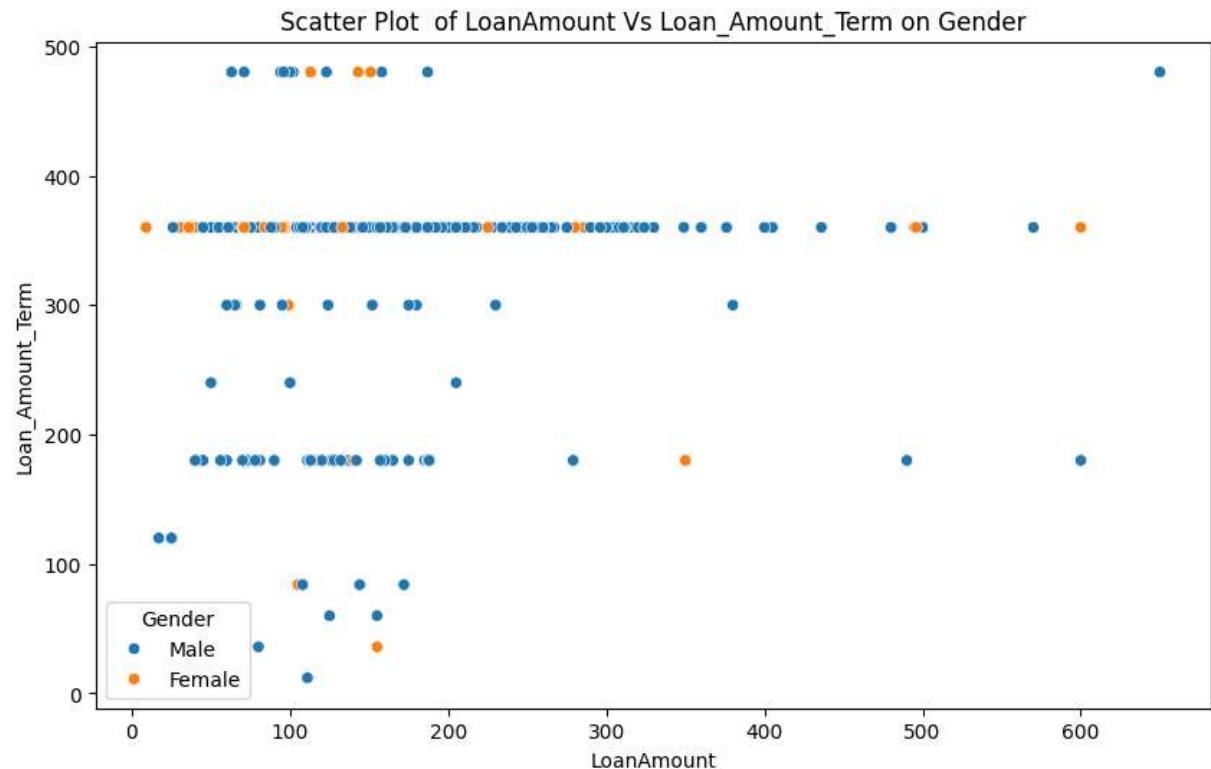
Scatter Plot of CoapplicantIncome Vs LoanAmount on Property_Area

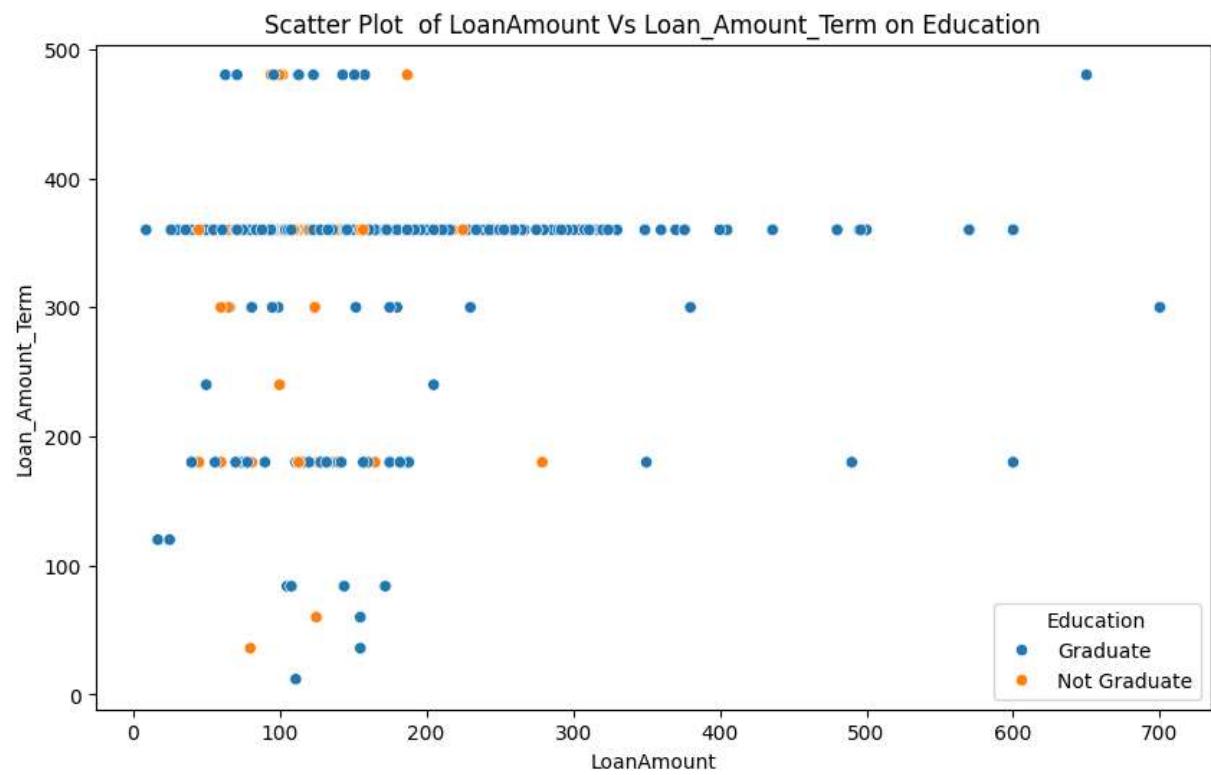
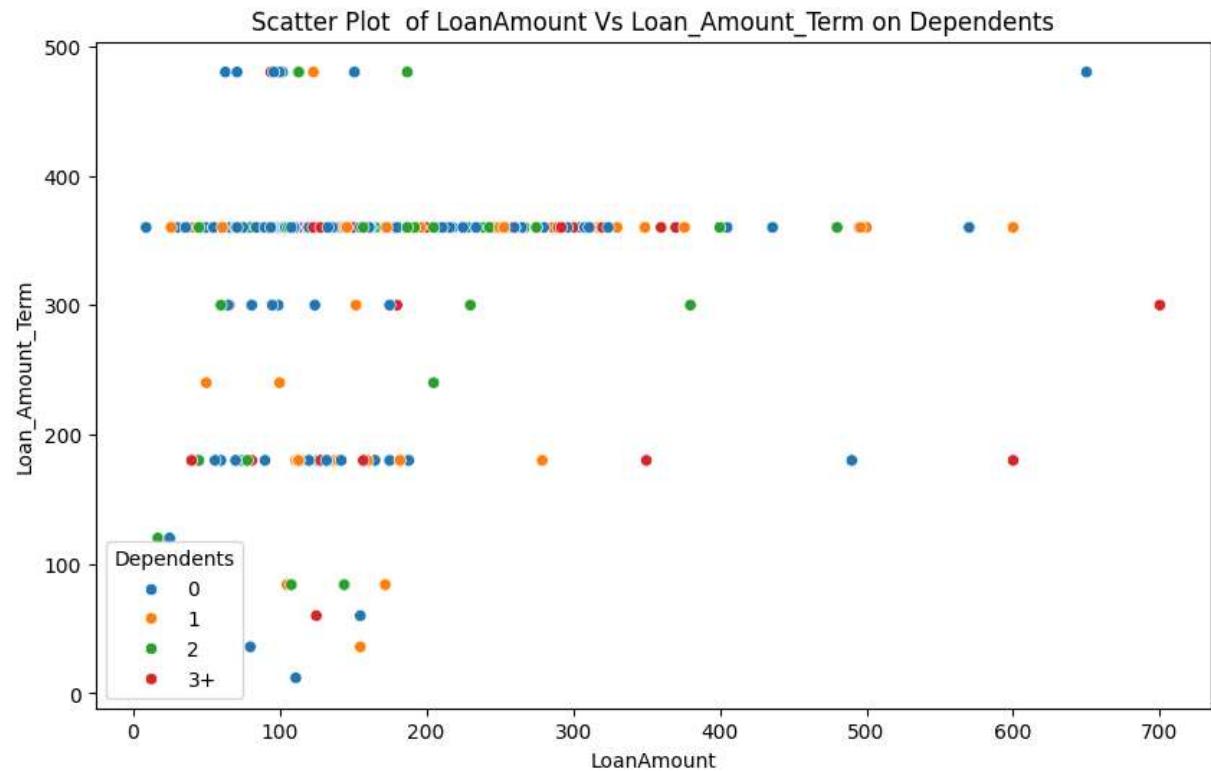


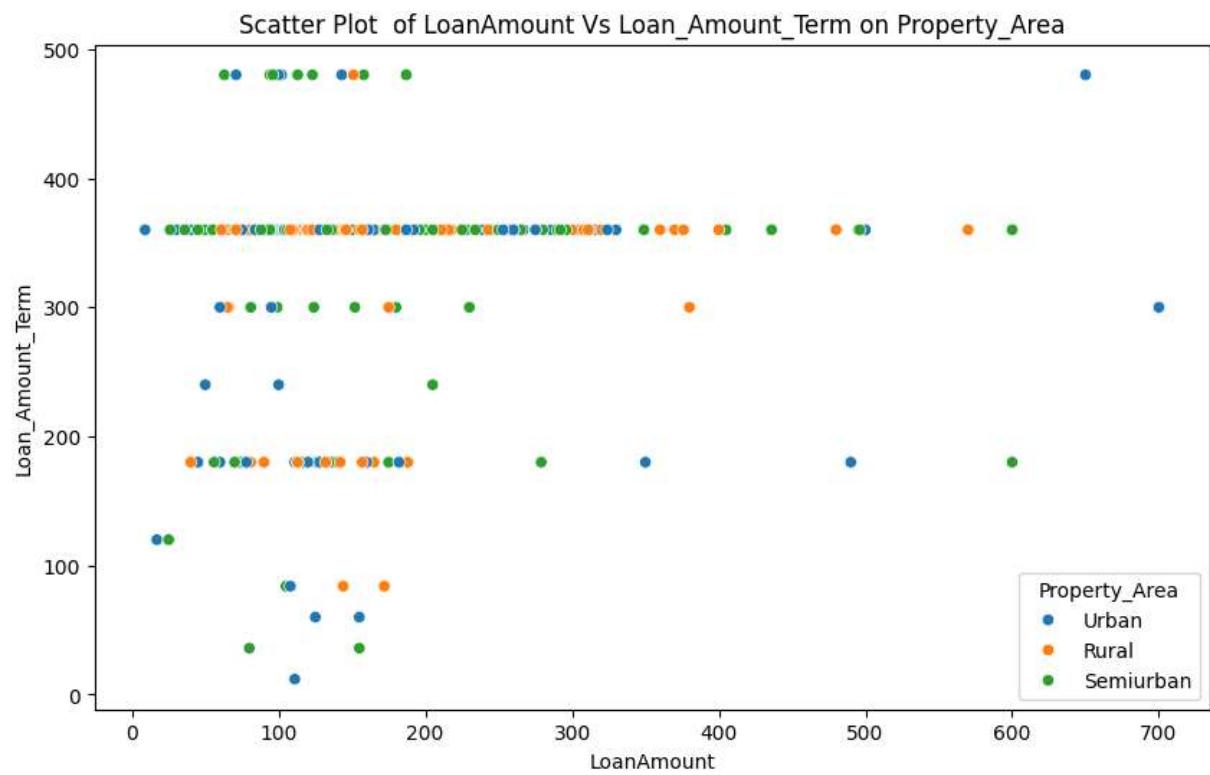
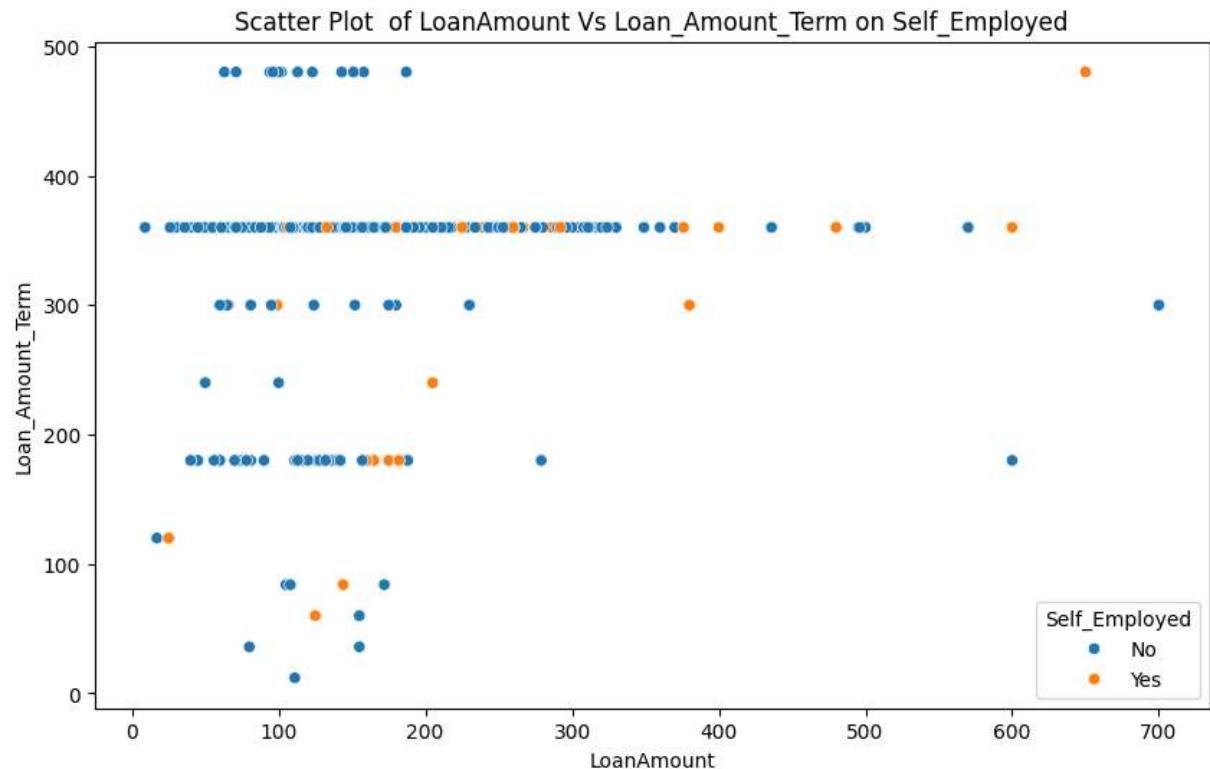




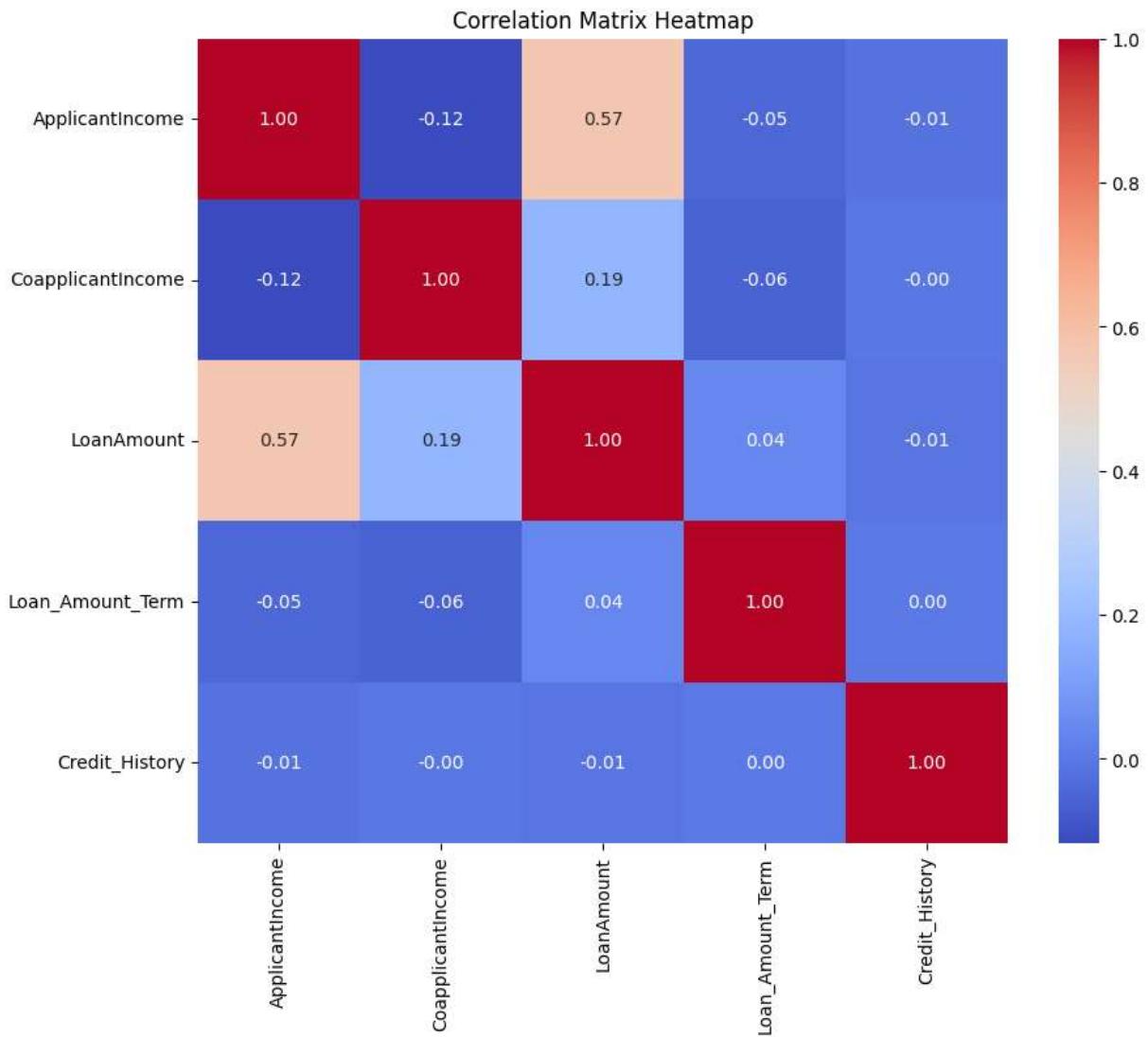




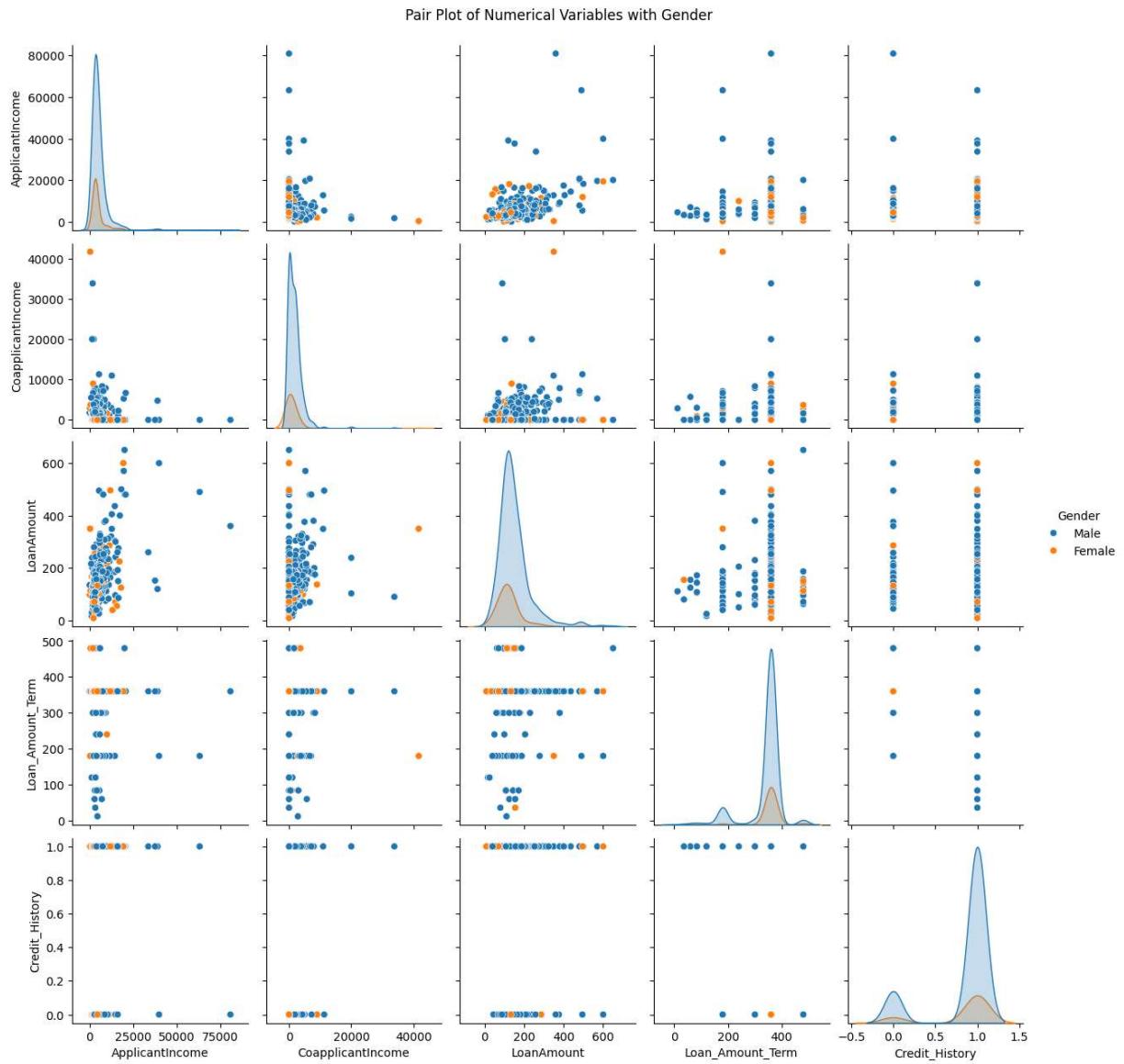




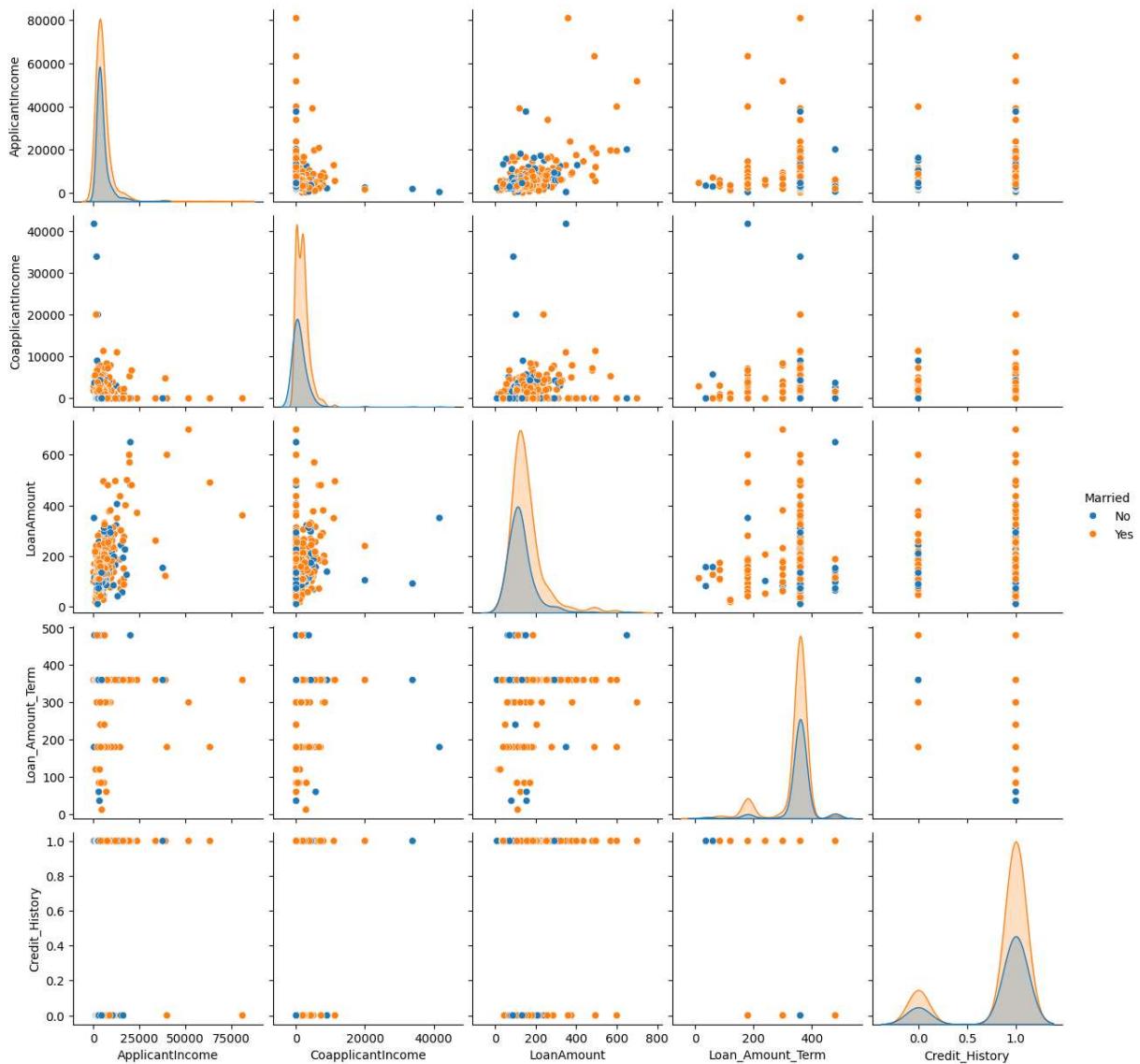
```
In [22]: # Multivariate Visualization
# Example 5: Heatmap of Correlation Matrix
correlation_matrix = df[numerical_variables].corr()
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Matrix Heatmap')
plt.show()
```



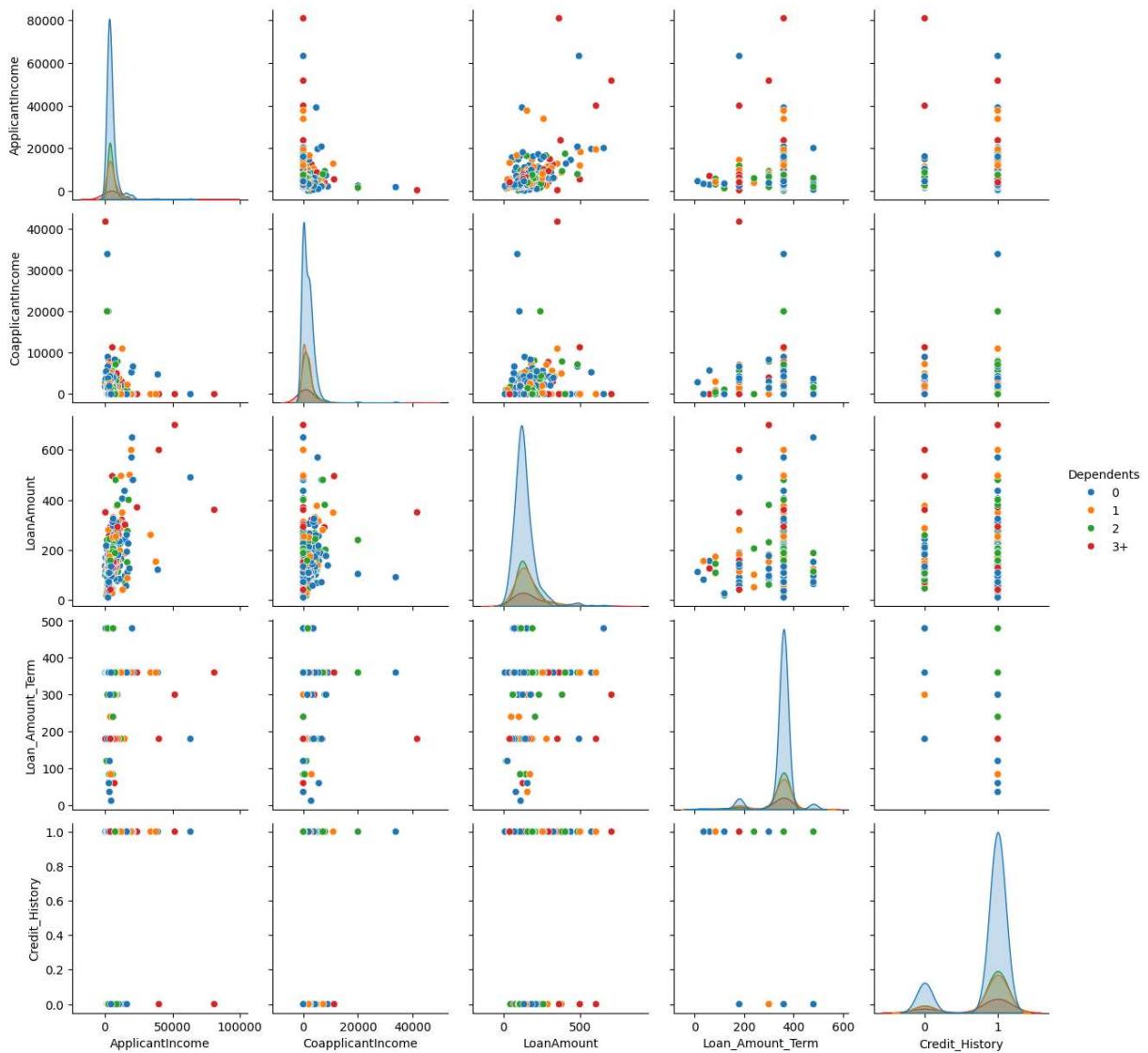
```
In [23]: # Create a pair plot with hue
for variables in categorical_variables:
    sns.pairplot(df, vars=numerical_variables, hue=variables)
    plt.suptitle(f'Pair Plot of Numerical Variables with {variables}', y=1.02)
    plt.show()
```



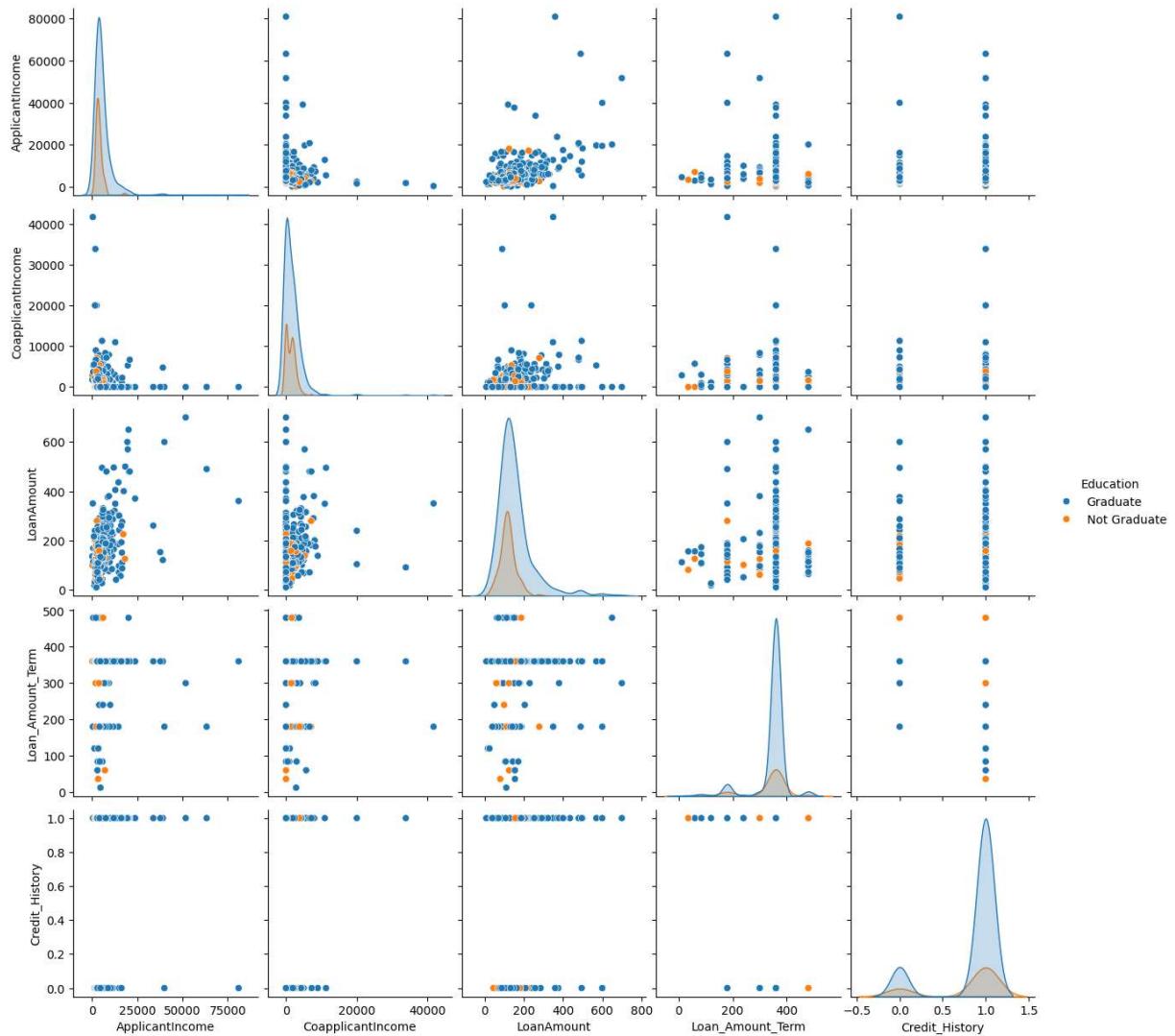
Pair Plot of Numerical Variables with Married



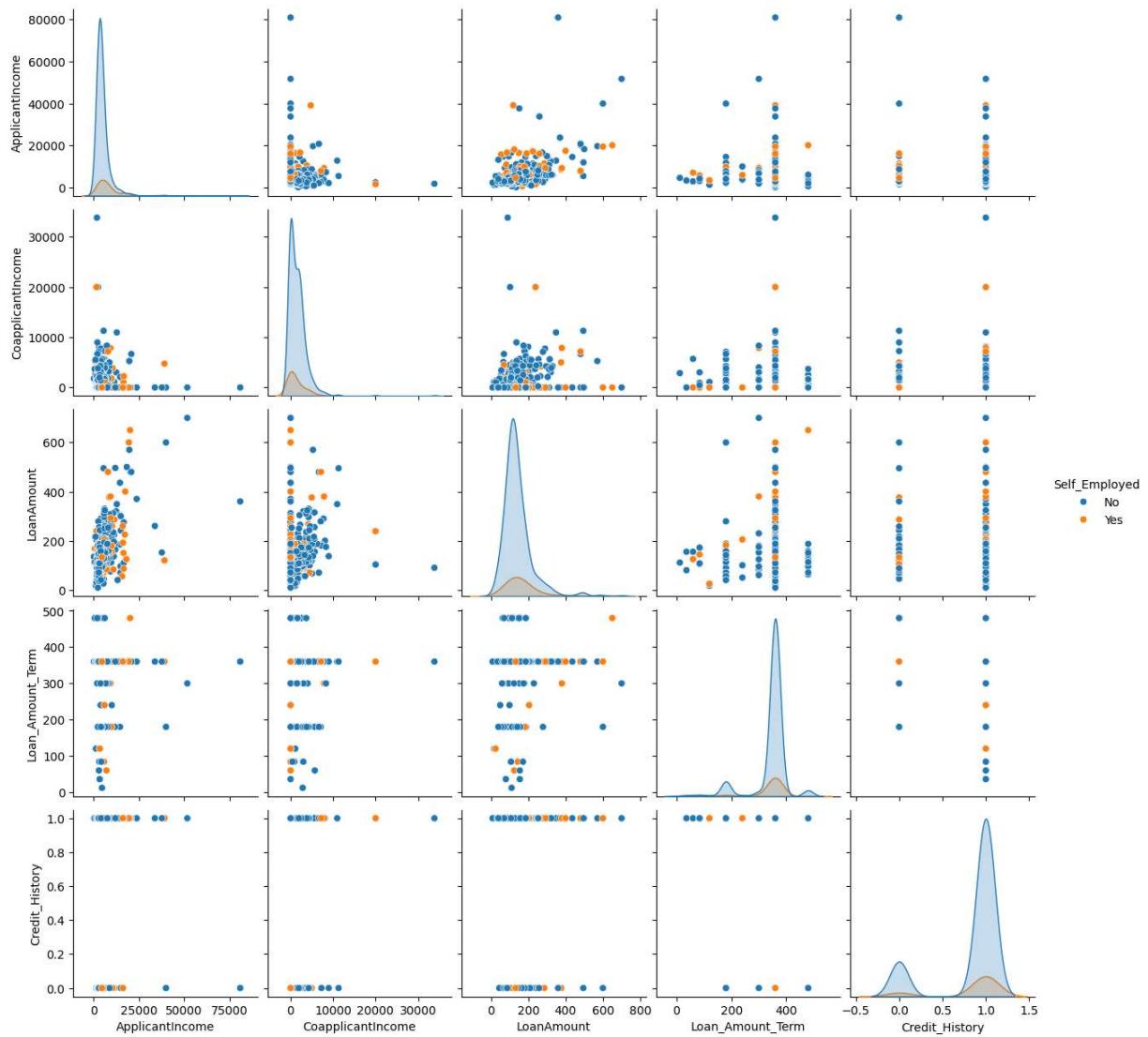
Pair Plot of Numerical Variables with Dependents



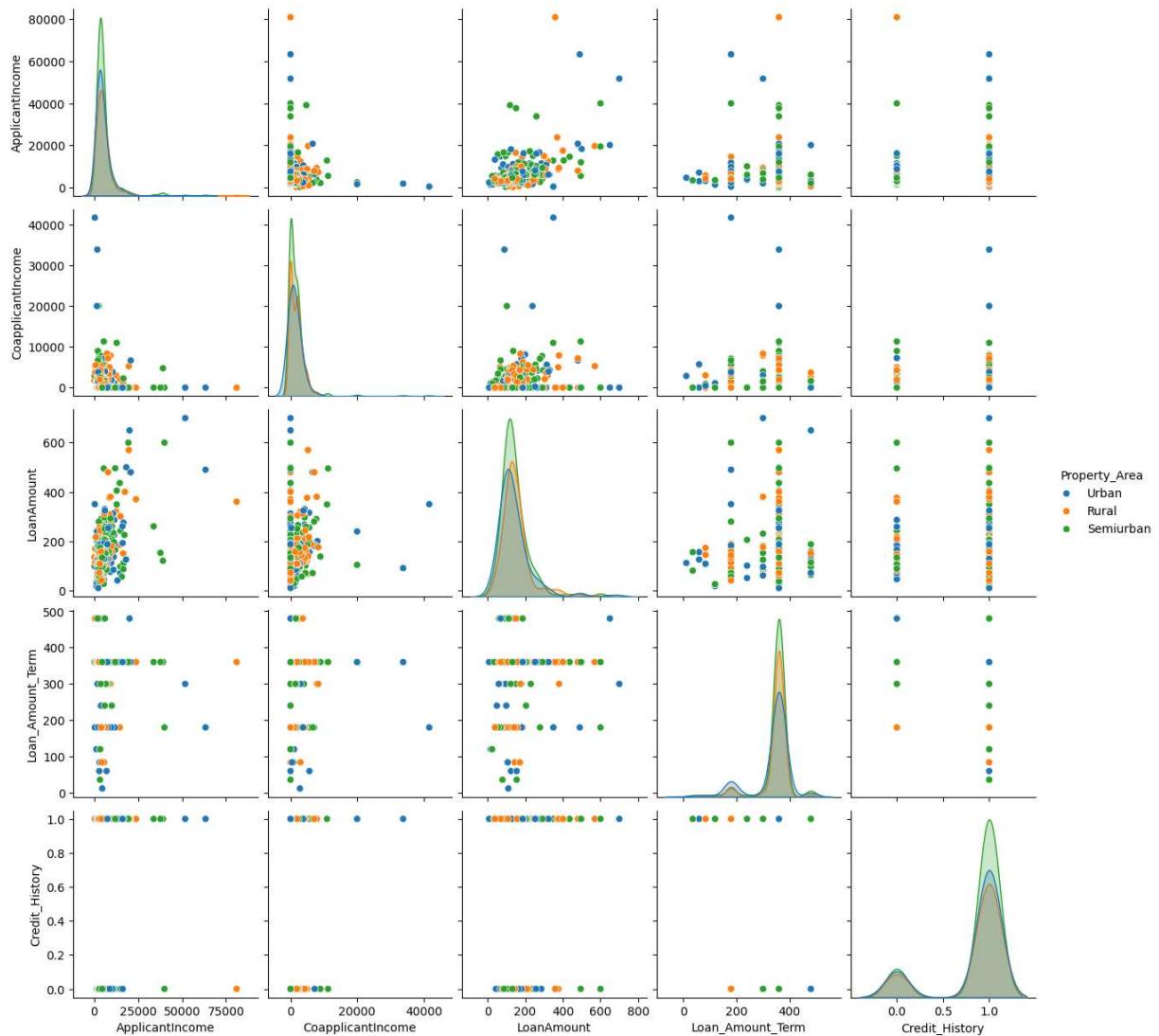
Pair Plot of Numerical Variables with Education

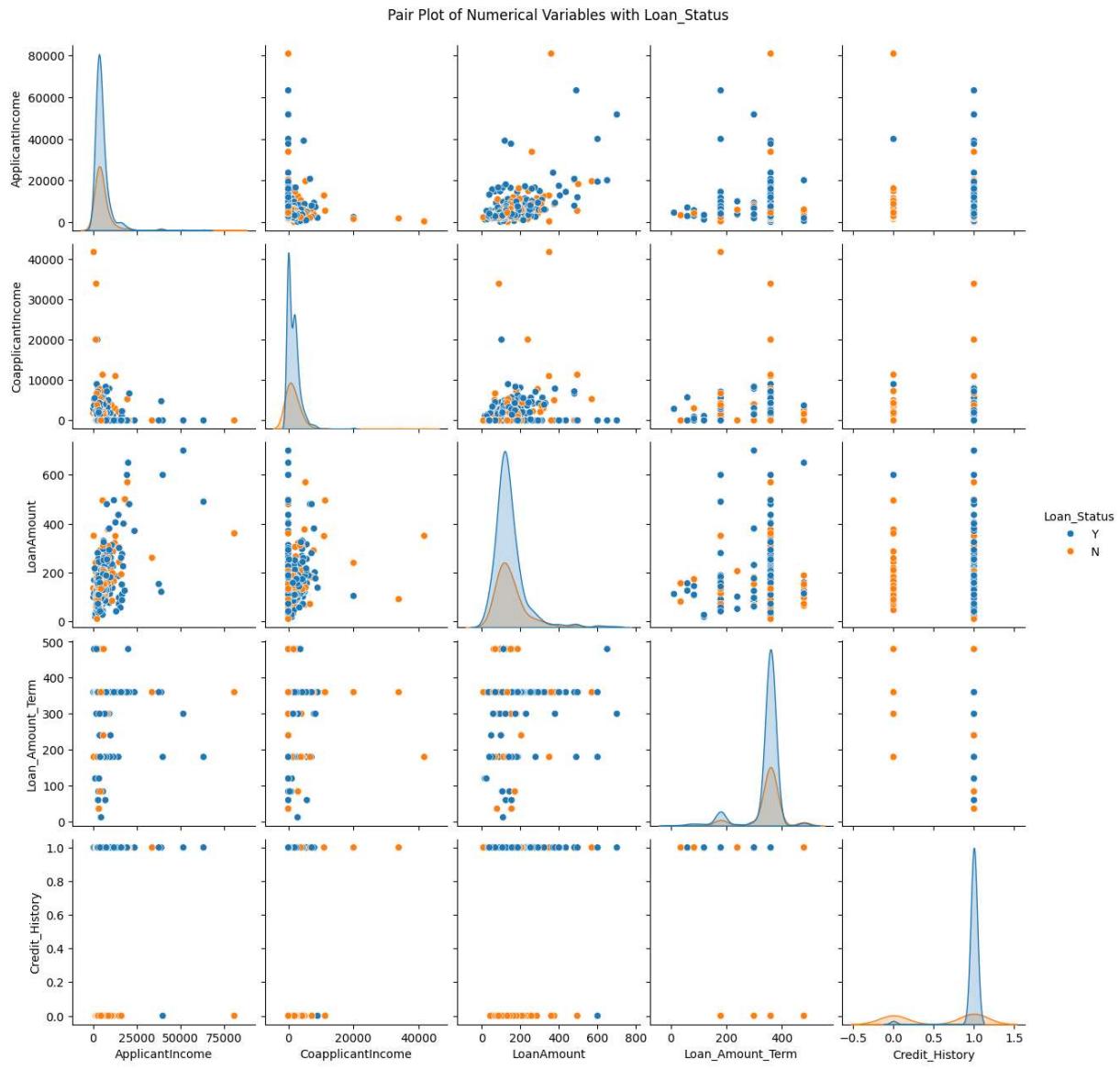


Pair Plot of Numerical Variables with Self_Employed



Pair Plot of Numerical Variables with Property_Area





In []: