

```
[1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[2]: df=pd.read_csv('/content/loan_test.csv')
df
```

```
[2]:
```

	Gender	Married	Dependents	Education	Self_Employed	Applicant_Income \
0	Male	Yes	0	Graduate	No	572000
1	Male	Yes	1	Graduate	No	307600
2	Male	Yes	2	Graduate	No	500000
3	Male	Yes	2	Graduate	No	234000
4	Male	No	0	Not Graduate	No	327600
..	...	...	...			...
362	Male	Yes	3+	Not Graduate	Yes	400900
363	Male	Yes	0	Graduate	No	415800
364	Male	No	0	Graduate	No	325000
365	Male	Yes	0	Graduate	No	500000
366	Male	No	0	Graduate	Yes	920000

	Coapplicant_Income	Loan_Amount	Term	Credit_History	Area
0	0	11000000	360.0	1.0	Urban
1	150000	12600000	360.0	1.0	Urban
2	180000	20800000	360.0	1.0	Urban
3	254600	10000000	360.0	NaN	Urban
4	0	7800000	360.0	1.0	Urban
..	...	...	...	...	...
362	177700	11300000	360.0	1.0	Urban
363	70900	11500000	360.0	1.0	Urban
364	199300	12600000	360.0	NaN	Semiurban
365	239300	15800000	360.0	1.0	Rural
366	0	9800000	180.0	1.0	Rural

[367 rows x 11 columns]

```
[3]: df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 367 entries, 0 to 366
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Gender                 356 non-null   object
1   Married                367 non-null   object
2   Dependents             357 non-null   object
3   Education              367 non-null   object
4   Self_Employed          344 non-null   object
5   Applicant_Income       367 non-null   int64
6   Coapplicant_Income     367 non-null   int64
7   Loan_Amount            367 non-null   int64
8   Term                   361 non-null   float64
9   Credit_History         338 non-null   float64
10  Area                   367 non-null   object
dtypes: float64(2), int64(3), object(6)
memory usage: 31.7+ KB

```

```
[4]: (df.isnull().sum()/df.shape[0])*100
```

```

[4]: Gender                2.997275
Married                0.000000
Dependents             2.724796
Education              0.000000
Self_Employed          6.267030
Applicant_Income       0.000000
Coapplicant_Income     0.000000
Loan_Amount            0.000000
Term                   1.634877
Credit_History         7.901907
Area                   0.000000
dtype: float64

```

```
[5]: df["Gender"].value_counts()
```

```

[5]: Gender
Male      286
Female    70
Name: count, dtype: int64

```

```
[6]: df["Gender"].isnull().sum()
```

```
[6]: 11
```

```
[7]: df["Gender"].mode()[0]
```

```
[7]: 'Male'
```

```
[8]: df["Gender"].fillna(df["Gender"].mode()[0],inplace=True)
df
```

```
[8]:      Gender Married Dependents      Education Self_Employed Applicant_Income \
0      Male      Yes           0      Graduate           No      572000
1      Male      Yes           1      Graduate           No      307600
2      Male      Yes           2      Graduate           No      500000
3      Male      Yes           2      Graduate           No      234000
4      Male      No            0  Not Graduate           No      327600
..      ...      ...           ...           ...           ...
362     Male      Yes          3+  Not Graduate          Yes      400900
363     Male      Yes           0      Graduate           No      415800
364     Male      No            0      Graduate           No      325000
365     Male      Yes           0      Graduate           No      500000
366     Male      No            0      Graduate          Yes      920000
```

```
      Coapplicant_Income Loan_Amount   Term Credit_History      Area
0              0      11000000  360.0           1.0      Urban
1          150000      12600000  360.0           1.0      Urban
2          180000      20800000  360.0           1.0      Urban
3          254600      10000000  360.0           NaN      Urban
4              0       78000000  360.0           1.0      Urban
..      ...           ...           ...           ...
362          177700      11300000  360.0           1.0      Urban
363           70900      11500000  360.0           1.0      Urban
364          199300      12600000  360.0           NaN  Semiurban
365          239300      15800000  360.0           1.0      Rural
366              0       9800000  180.0           1.0      Rural
```

```
[367 rows x 11 columns]
```

```
[9]: df.dtypes
```

```
[9]: Gender           object
Married           object
Dependents        object
Education         object
Self_Employed     object
Applicant_Income  int64
Coapplicant_Income int64
Loan_Amount       int64
Term             float64
Credit_History   float64
Area             object
dtype: object
```

```
[10]: df['Dependents'].fillna(df['Dependents'].mode()[0], inplace=True)
df['Dependents'].isnull().sum()
df
```

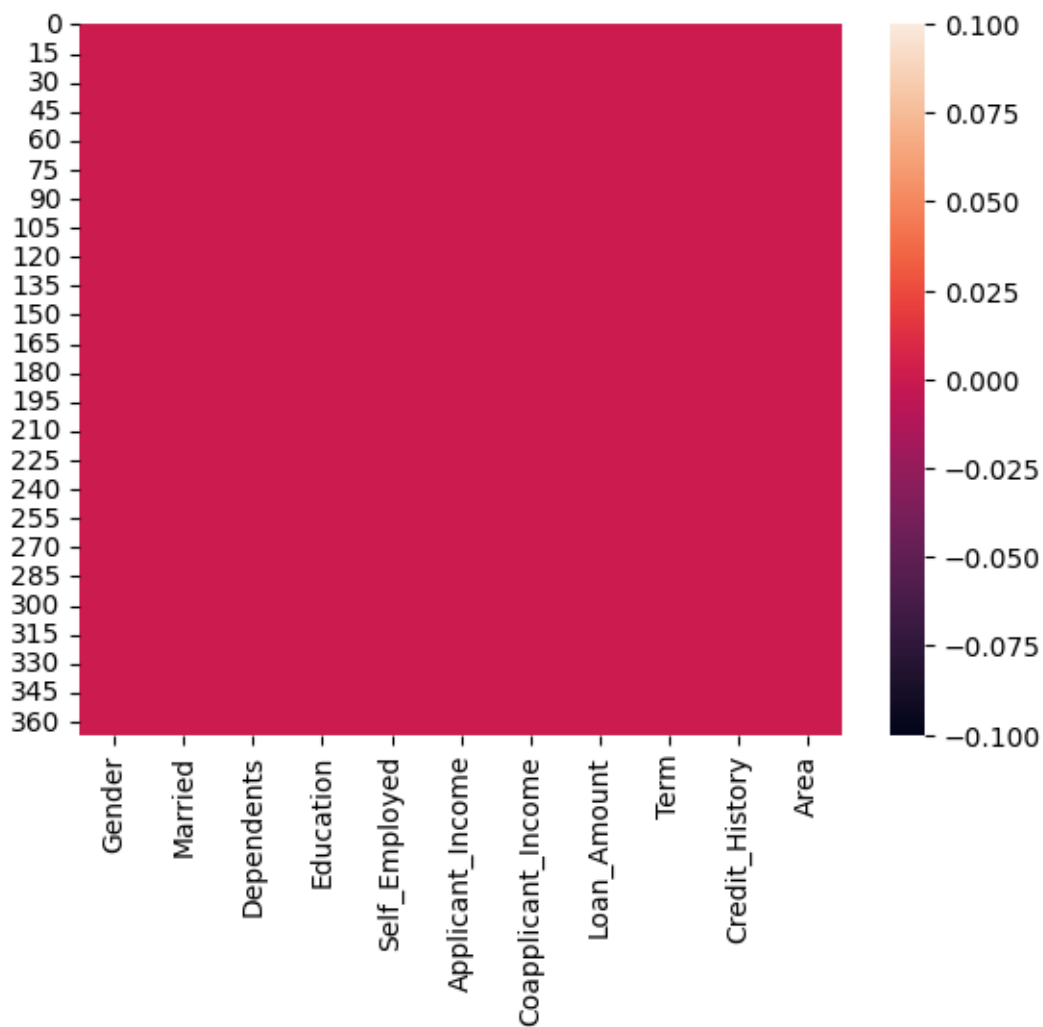
```
[10]:
```

	Gender	Married	Dependents	Education	Self_Employed	Applicant_Income	\
0	Male	Yes	0	Graduate	No	572000	
1	Male	Yes	1	Graduate	No	307600	
2	Male	Yes	2	Graduate	No	500000	
3	Male	Yes	2	Graduate	No	234000	
4	Male	No	0	Not Graduate	No	327600	
..	...	...	...	...	...	...	
362	Male	Yes	3+	Not Graduate	Yes	400900	
363	Male	Yes	0	Graduate	No	415800	
364	Male	No	0	Graduate	No	325000	
365	Male	Yes	0	Graduate	No	500000	
366	Male	No	0	Graduate	Yes	920000	

	Coapplicant_Income	Loan_Amount	Term	Credit_History	Area
0	0	11000000	360.0	1.0	Urban
1	150000	12600000	360.0	1.0	Urban
2	180000	20800000	360.0	1.0	Urban
3	254600	10000000	360.0	NaN	Urban
4	0	7800000	360.0	1.0	Urban
..	...	...	...	...	...
362	177700	11300000	360.0	1.0	Urban
363	70900	11500000	360.0	1.0	Urban
364	199300	12600000	360.0	NaN	Semiurban
365	239300	15800000	360.0	1.0	Rural
366	0	9800000	180.0	1.0	Rural

[367 rows x 11 columns]

```
[21]: sns.heatmap(df.isnull())
plt.show()
```



```
[12]: df['Credit_History'].fillna(df['Credit_History'].mode()[0],inplace=True)
df['Credit_History'].isnull().sum()
```

[12]: 0

```
[13]: df['Self_Employed'].fillna(df['Self_Employed'].mode()[0],inplace=True)
df['Self_Employed'].isnull().sum()
```

[13]: 0

```
[14]: df['Term'].isnull().sum()
```

[14]: 6

```
[15]: df['Term'].fillna(method="bfill",inplace=True)
df['Term'].isnull().sum()
df
```

```
[15]:
```

	Gender	Married	Dependents	Education	Self_Employed	Applicant_Income	\
0	Male	Yes	0	Graduate	No	572000	
1	Male	Yes	1	Graduate	No	307600	
2	Male	Yes	2	Graduate	No	500000	
3	Male	Yes	2	Graduate	No	234000	
4	Male	No	0	Not Graduate	No	327600	
..	...	...	...	...	...	...	
362	Male	Yes	3+	Not Graduate	Yes	400900	
363	Male	Yes	0	Graduate	No	415800	
364	Male	No	0	Graduate	No	325000	
365	Male	Yes	0	Graduate	No	500000	
366	Male	No	0	Graduate	Yes	920000	

	Coapplicant_Income	Loan_Amount	Term	Credit_History	Area
0	0	11000000	360.0	1.0	Urban
1	150000	12600000	360.0	1.0	Urban
2	180000	20800000	360.0	1.0	Urban
3	254600	10000000	360.0	1.0	Urban
4	0	7800000	360.0	1.0	Urban
..	...	...	...	...	...
362	177700	11300000	360.0	1.0	Urban
363	70900	11500000	360.0	1.0	Urban
364	199300	12600000	360.0	1.0	Semiurban
365	239300	15800000	360.0	1.0	Rural
366	0	9800000	180.0	1.0	Rural

[367 rows x 11 columns]

```
[16]: df['Education'].replace({'Not Graduate':0,'Graduate':1})
df
```

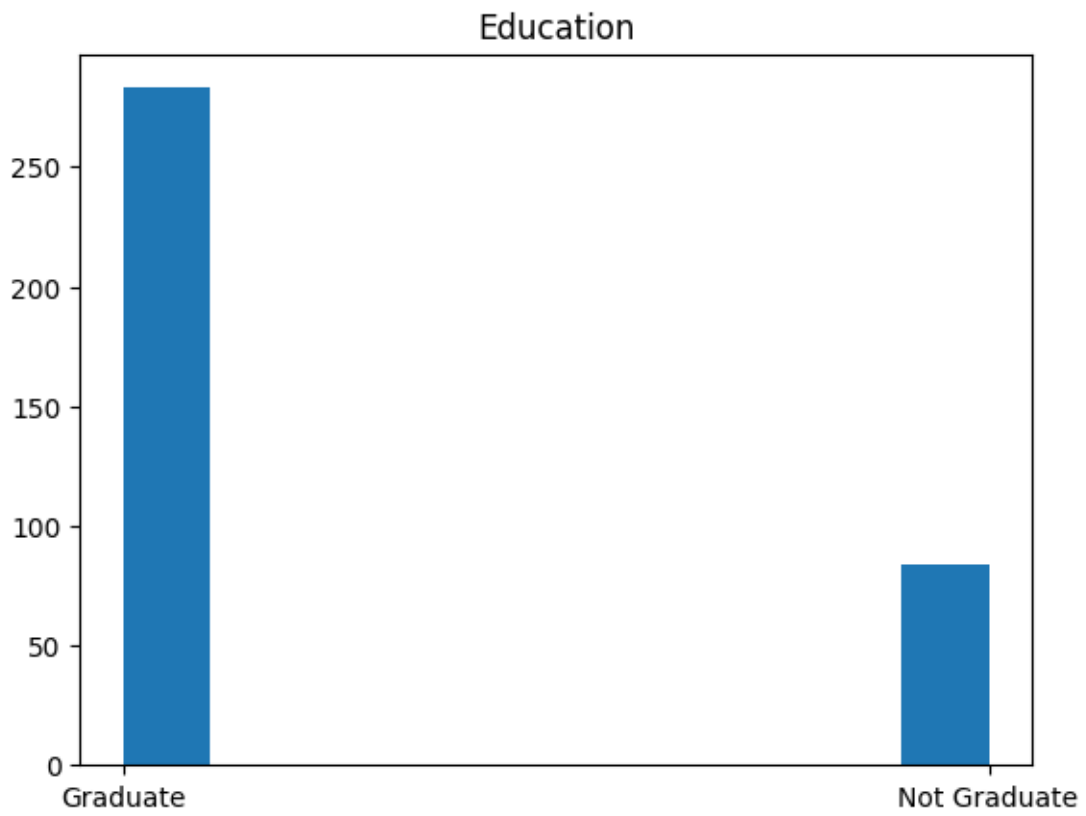
```
[16]:
```

	Gender	Married	Dependents	Education	Self_Employed	Applicant_Income	\
0	Male	Yes	0	Graduate	No	572000	
1	Male	Yes	1	Graduate	No	307600	
2	Male	Yes	2	Graduate	No	500000	
3	Male	Yes	2	Graduate	No	234000	
4	Male	No	0	Not Graduate	No	327600	
..	...	...	...	...	...	...	
362	Male	Yes	3+	Not Graduate	Yes	400900	
363	Male	Yes	0	Graduate	No	415800	
364	Male	No	0	Graduate	No	325000	
365	Male	Yes	0	Graduate	No	500000	
366	Male	No	0	Graduate	Yes	920000	

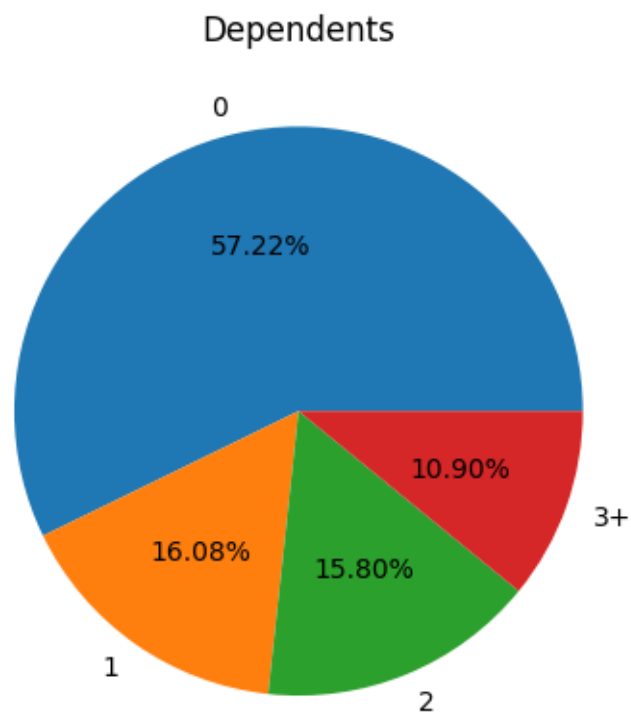
	Coapplicant_Income	Loan_Amount	Term	Credit_History	Area
0	0	11000000	360.0	1.0	Urban
1	150000	12600000	360.0	1.0	Urban
2	180000	20800000	360.0	1.0	Urban
3	254600	10000000	360.0	1.0	Urban
4	0	7800000	360.0	1.0	Urban
..	...	...	...	...	...
362	177700	11300000	360.0	1.0	Urban
363	70900	11500000	360.0	1.0	Urban
364	199300	12600000	360.0	1.0	Semiurban
365	239300	15800000	360.0	1.0	Rural
366	0	9800000	180.0	1.0	Rural

[367 rows x 11 columns]

```
[17]: #Education
plt.hist(x=df['Education']).value_counts(), labels=["Graduate", "Not_
↳Graduate"], autopct="%0.2f%%")
plt.title("Education")
plt.show()
plt.pie(x=df['Dependents'].value_counts(), labels=["0", "1", "2", "3+"], autopct="%0.
↳2f%%")
plt.title("Dependents")
plt.show()
```







```
[18]: #plt.scatter(df['Applicant_Income'],df['Loan_Amount'])
      #plt.show()
```

```
[19]: df.replace({'Self_Employed':{'No':0,'Yes':1}},{'Married':{'No':0,'Yes':1}})
      df
```

```
[19]:
```

	Gender	Married	Dependents	Education	Self_Employed	Applicant_Income	\
0	Male	Yes	0	Graduate	No	572000	
1	Male	Yes	1	Graduate	No	307600	
2	Male	Yes	2	Graduate	No	500000	
3	Male	Yes	2	Graduate	No	234000	
4	Male	No	0	Not Graduate	No	327600	
..	...	...	...	...	...	...	
362	Male	Yes	3+	Not Graduate	Yes	400900	
363	Male	Yes	0	Graduate	No	415800	
364	Male	No	0	Graduate	No	325000	
365	Male	Yes	0	Graduate	No	500000	
366	Male	No	0	Graduate	Yes	920000	

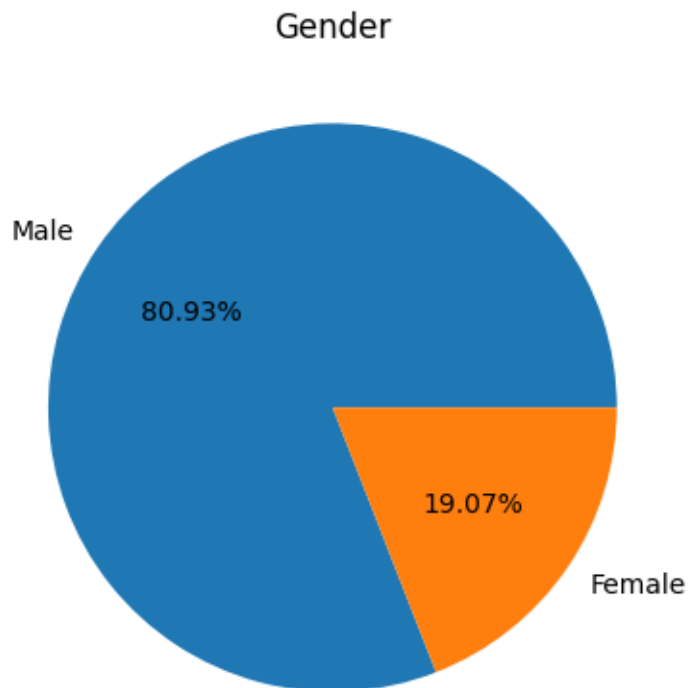
  

	Coapplicant_Income	Loan_Amount	Term	Credit_History	Area
0	0	11000000	360.0	1.0	Urban
1	150000	12600000	360.0	1.0	Urban

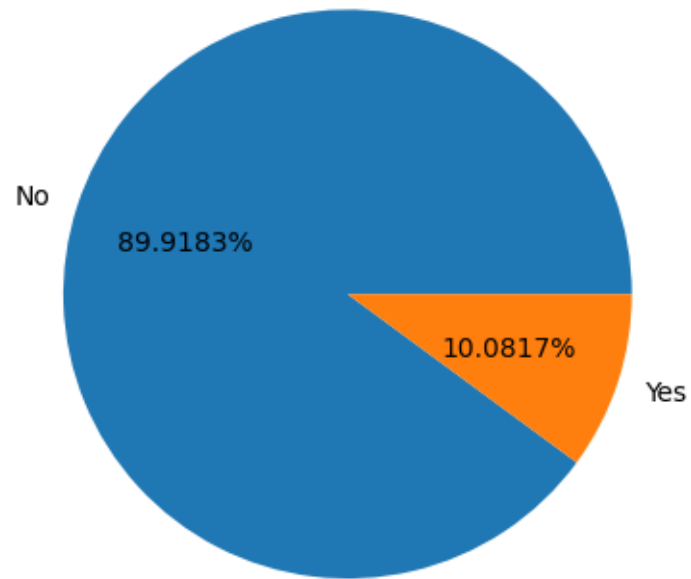
2	180000	20800000	360.0	1.0	Urban
3	254600	10000000	360.0	1.0	Urban
4	0	7800000	360.0	1.0	Urban
..	...	...	...	...	...
362	177700	11300000	360.0	1.0	Urban
363	70900	11500000	360.0	1.0	Urban
364	199300	12600000	360.0	1.0	Semiurban
365	239300	15800000	360.0	1.0	Rural
366	0	9800000	180.0	1.0	Rural

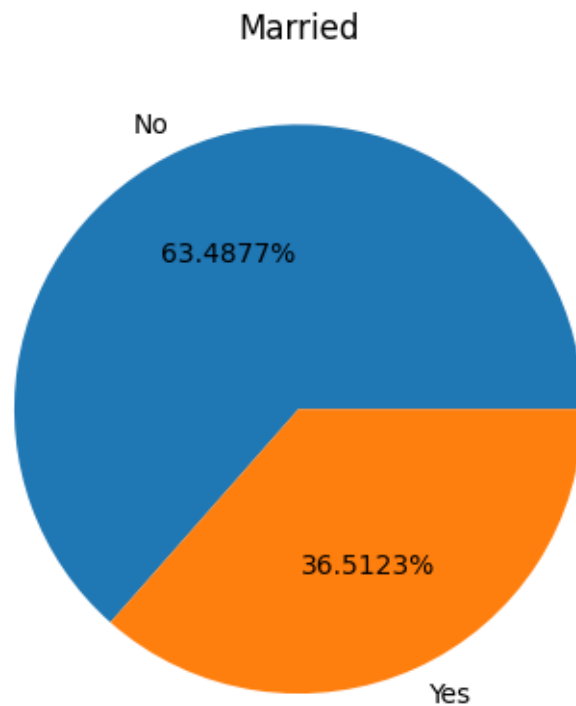
[367 rows x 11 columns]

```
[22]: plt.pie(x=df['Gender'].value_counts(),labels=["Male","Female"],autopct="%0.
      ↪2f%%")
plt.title("Gender")
plt.show()
plt.pie(x=df['Self_Employed'].value_counts(),labels=["No","Yes"],autopct="%0.
      ↪4f%%" )
plt.title("Self Employed")
plt.show()
plt.pie(x=df['Married'].value_counts(),labels=["No","Yes"],autopct="%0.4f%%")
plt.title("Married")
plt.show()
```



### Self Employed





```
[23]: df.describe()
```

```
[23]:
```

	Applicant_Income	Coapplicant_Income	Loan_Amount	Term \
count	3.670000e+02	3.670000e+02	3.670000e+02	367.000000
mean	4.805599e+05	1.569578e+05	1.342779e+07	342.659401
std	4.910685e+05	2.334232e+05	6.296143e+06	64.690698
min	0.000000e+00	0.000000e+00	0.000000e+00	6.000000
25%	2.864000e+05	0.000000e+00	1.000000e+07	360.000000
50%	3.786000e+05	1.025000e+05	1.250000e+07	360.000000
75%	5.060000e+05	2.430500e+05	1.575000e+07	360.000000
max	7.252900e+06	2.400000e+06	5.500000e+07	480.000000

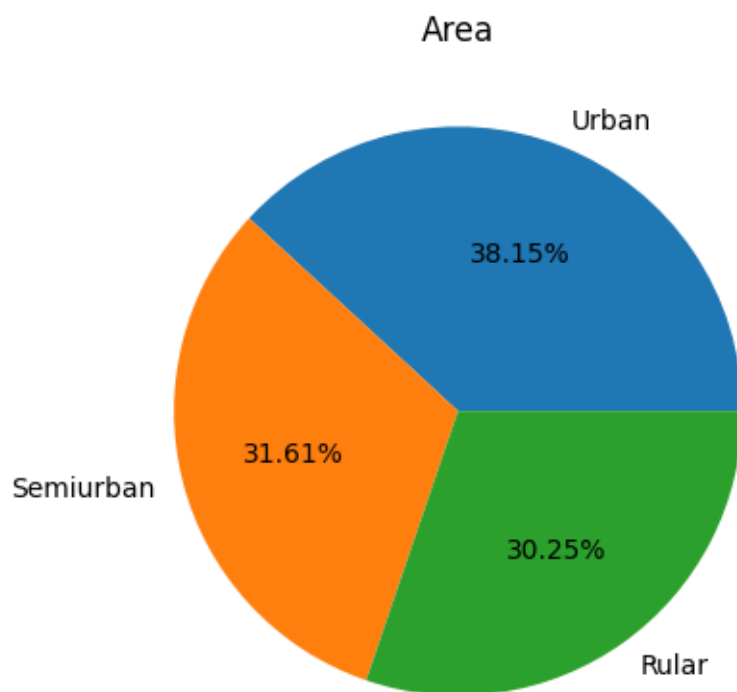
  

	Credit_History
count	367.000000
mean	0.839237
std	0.367814
min	0.000000
25%	1.000000
50%	1.000000
75%	1.000000
max	1.000000

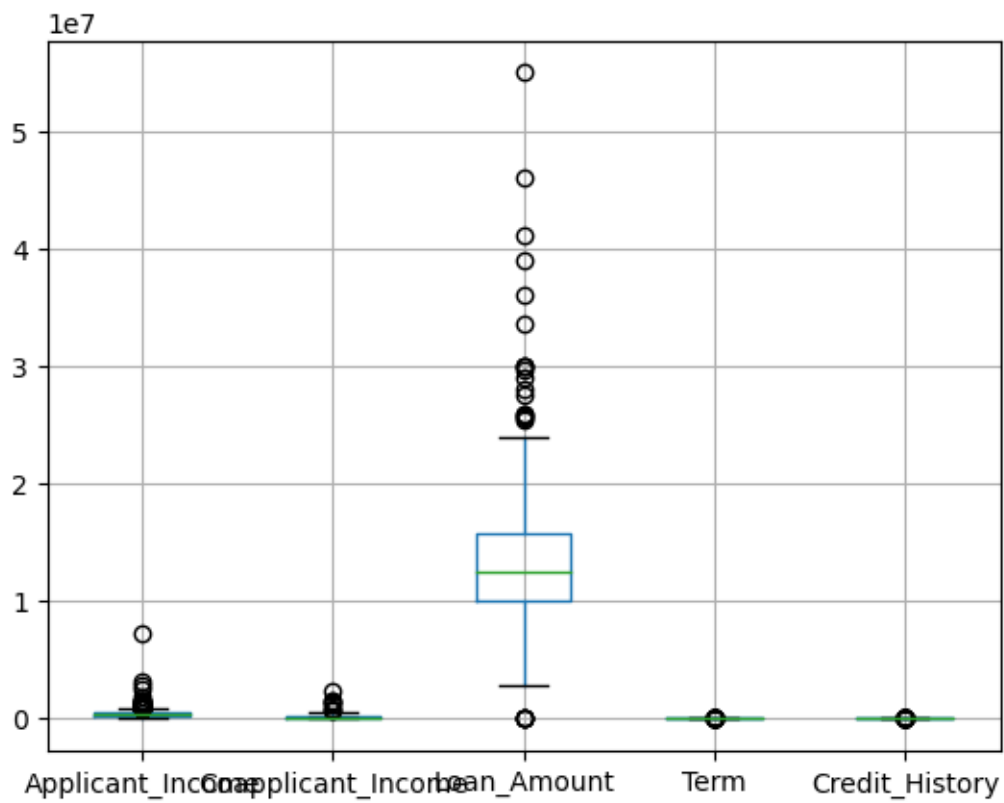
```
[24]: df["Area"].value_counts()
```

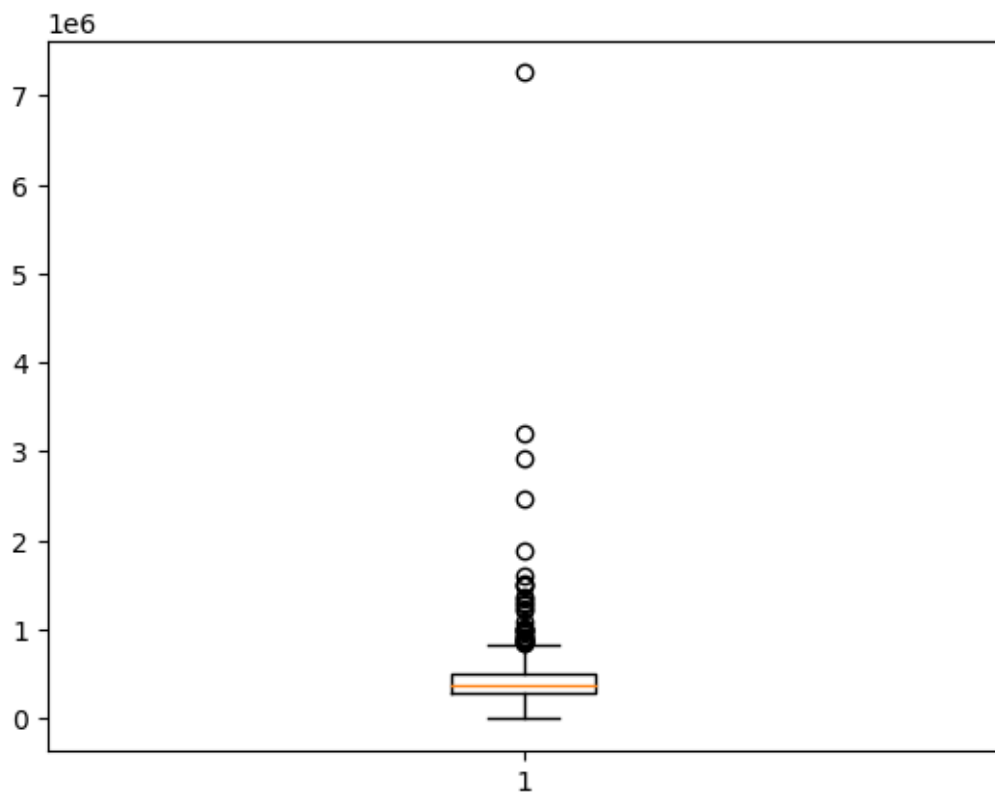
```
[24]: Area
Urban      140
Semiurban  116
Rural      111
Name: count, dtype: int64
```

```
[25]: df["Area"].replace({"Urban":0,"Semiurban":1,"Rular":2}) ,plt.pie(x=df['Area'].
    ↪value_counts(),labels=["Urban","Semiurban","Rular"],autopct="%0.2f%%")
plt.title("Area")
plt.show()
```



```
[34]: df.boxplot()
plt.show()
plt.boxplot(df['Applicant_Income'])
plt.show()
```





[ ]: