Task-3

Visualization using Histogram

Create a histogram or bar chart to visualize the distribution of data in a dataset

In [5]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [6]:
```
path=r"C:\Users\Sruth\Documents\Naresh it\EDA\Datafiles\Loan_prediction_data.csv
df=pd.read_csv(path)
df
```

Out[6]:

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantInco
0	LP001002	Male	No	0	Graduate	No	5
1	LP001003	Male	Yes	1	Graduate	No	4
2	LP001005	Male	Yes	0	Graduate	Yes	3
3	LP001006	Male	Yes	0	Not Graduate	No	2
4	LP001008	Male	No	0	Graduate	No	6
...	...	...	...	...	...	...	
609	LP002978	Female	No	0	Graduate	No	2
610	LP002979	Male	Yes	3+	Graduate	No	4
611	LP002983	Male	Yes	1	Graduate	No	8
612	LP002984	Male	Yes	2	Graduate	No	7
613	LP002990	Female	No	0	Graduate	Yes	4

614 rows × 13 columns

In [7]:
```
cat_cols=df.select_dtypes(include='object').columns
cat_cols
```

Out[7]:
```
Index(['Loan_ID', 'Gender', 'Married', 'Dependents', 'Education',
       'Self_Employed', 'Property_Area', 'Loan_Status'],
      dtype='object')
```

In [9]:
```
num_cols=df.select_dtypes(exclude='object').columns
num_cols
```

Out[9]:
```
Index(['ApplicantIncome', 'CoapplicantIncome', 'LoanAmount',
       'Loan_Amount_Term', 'Credit_History'],
      dtype='object')
```

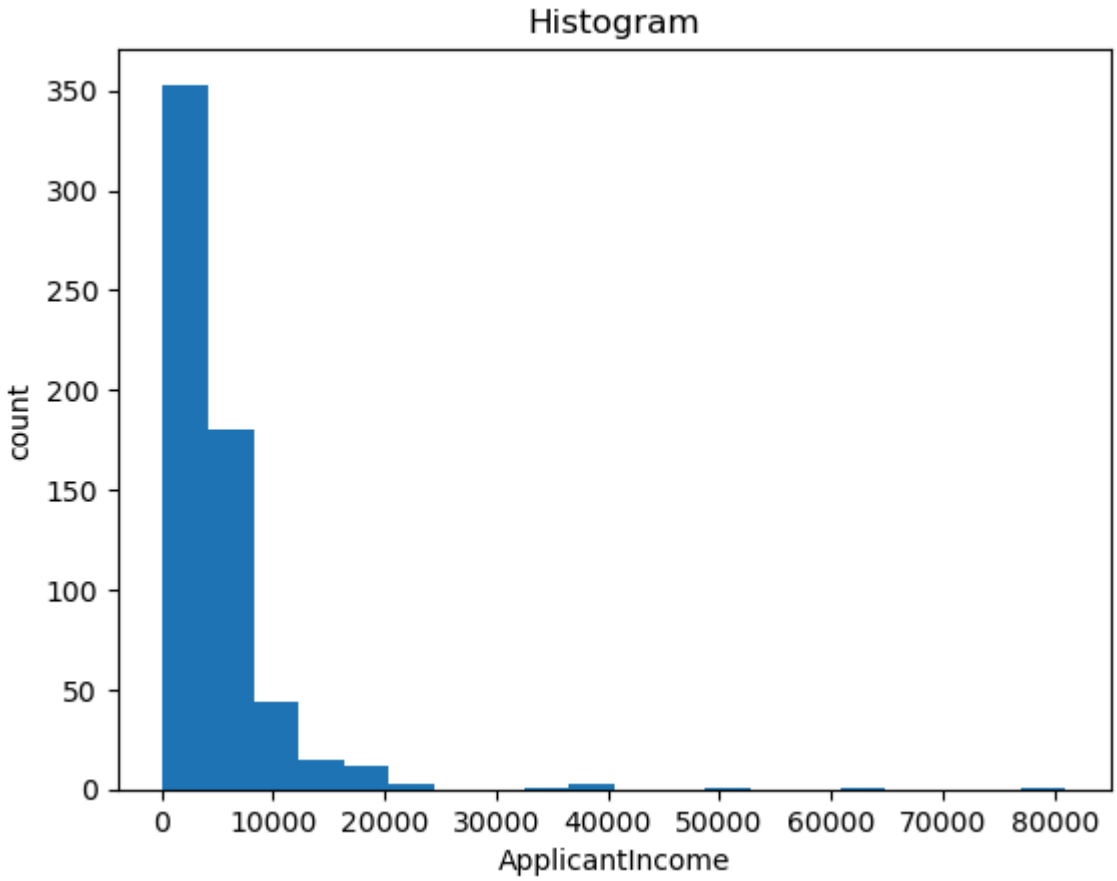
For numerical columns

In [10]:
```
num_cols
```

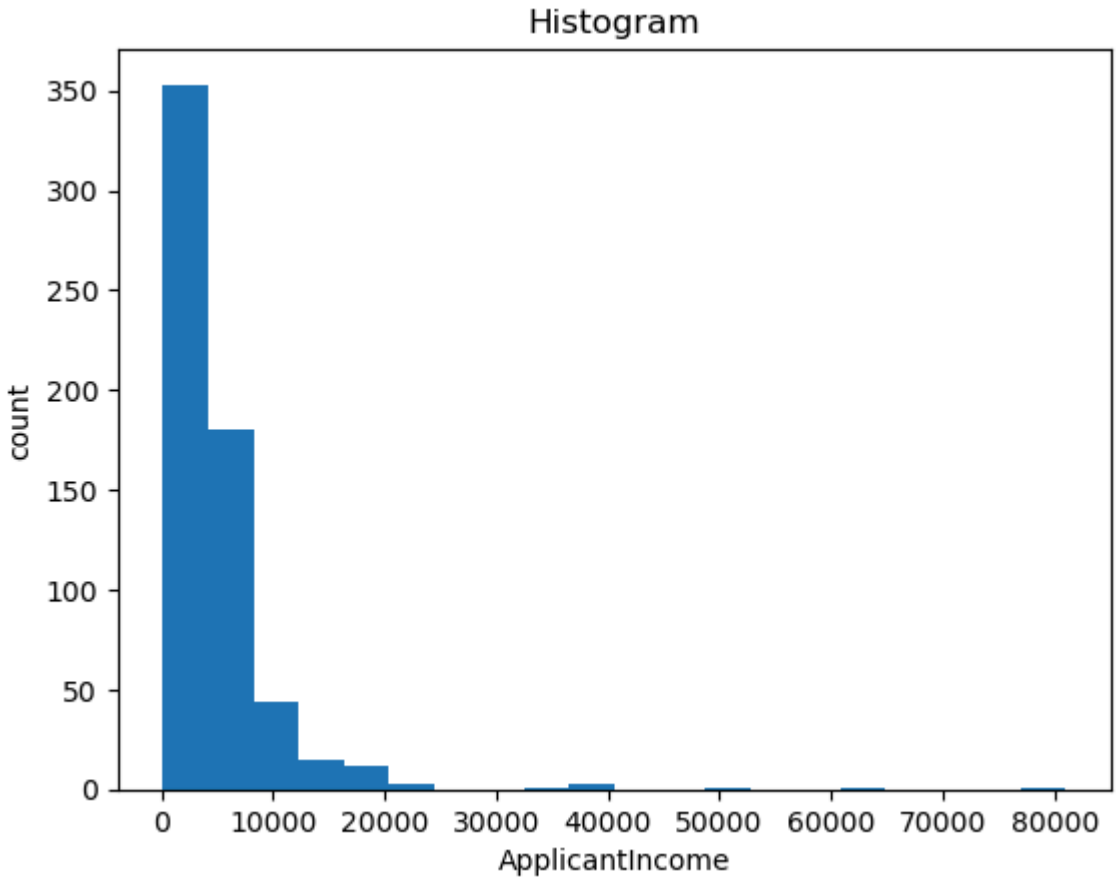
Out[10]:
```
Index(['ApplicantIncome', 'CoapplicantIncome', 'LoanAmount',
       'Loan_Amount_Term', 'Credit_History'],
      dtype='object')
```

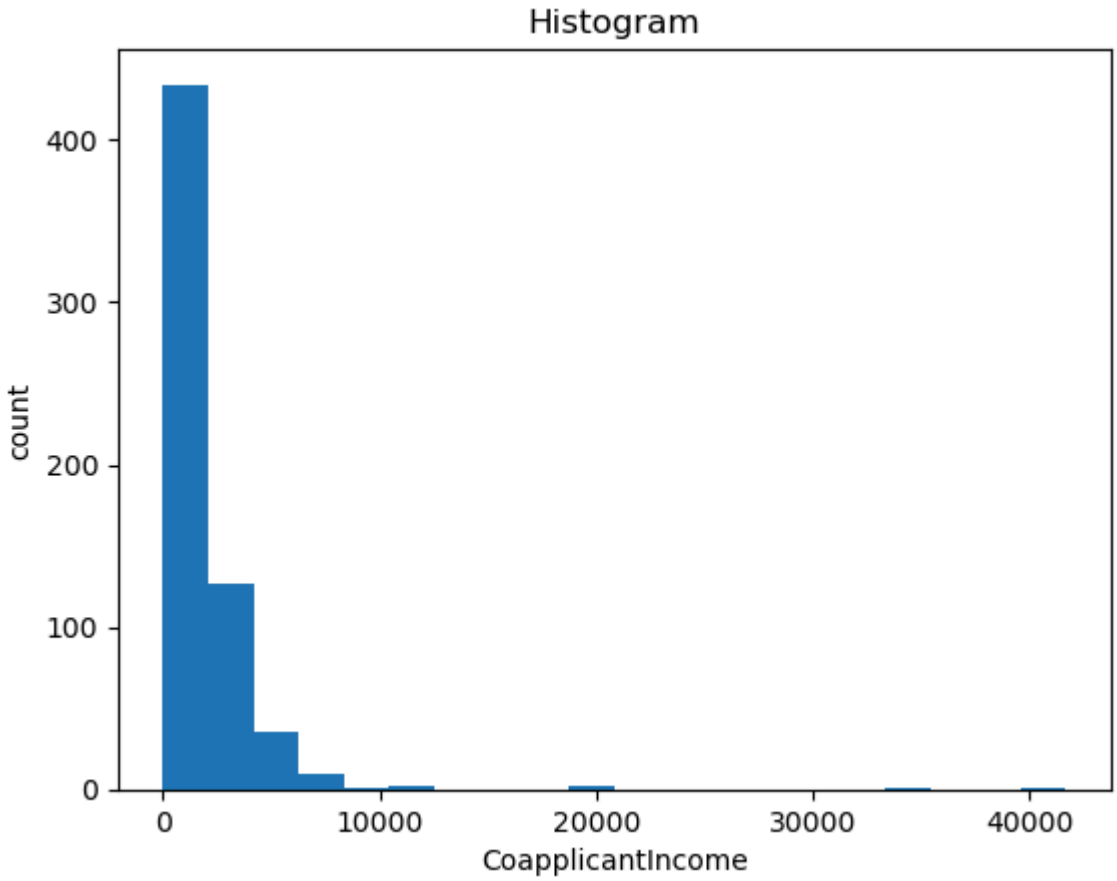
visualize Histogram

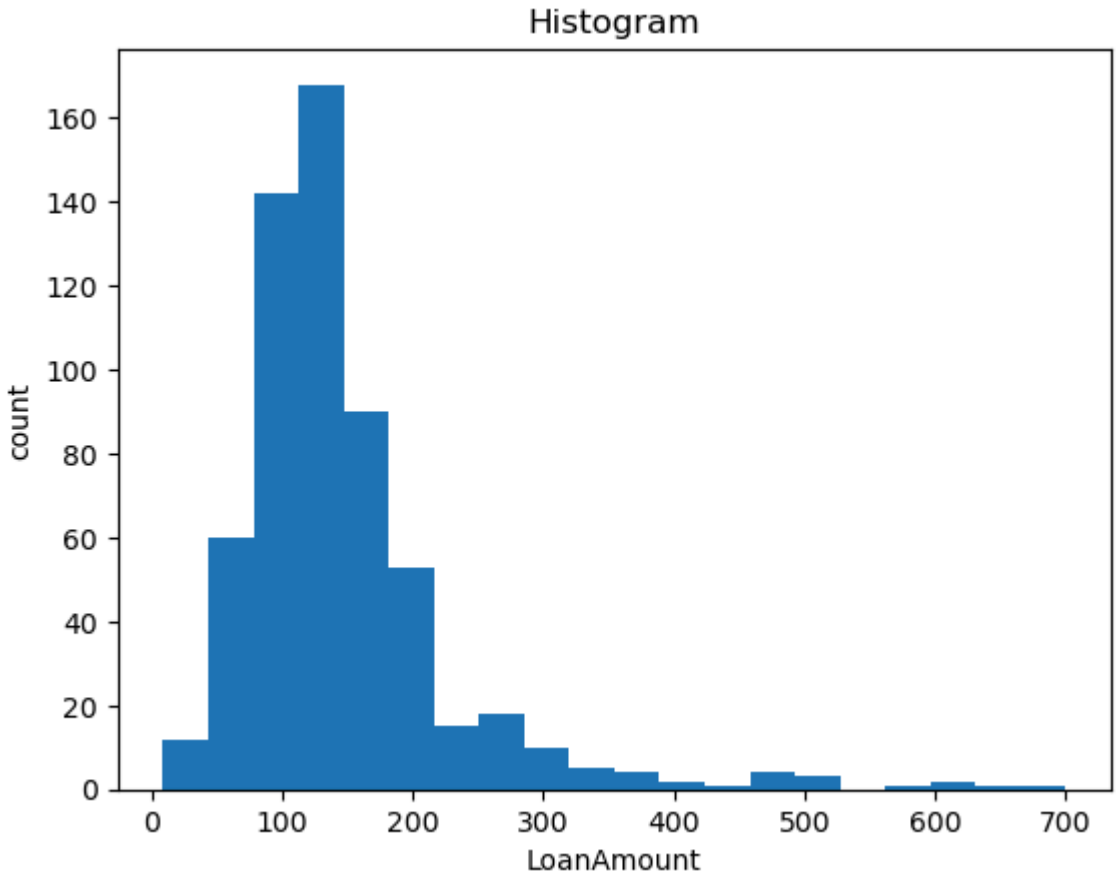
In [16]:
```
plt.hist(df['ApplicantIncome'],bins=20)
plt.title('Histogram')
plt.xlabel('ApplicantIncome')
plt.ylabel('count')
plt.show()
```

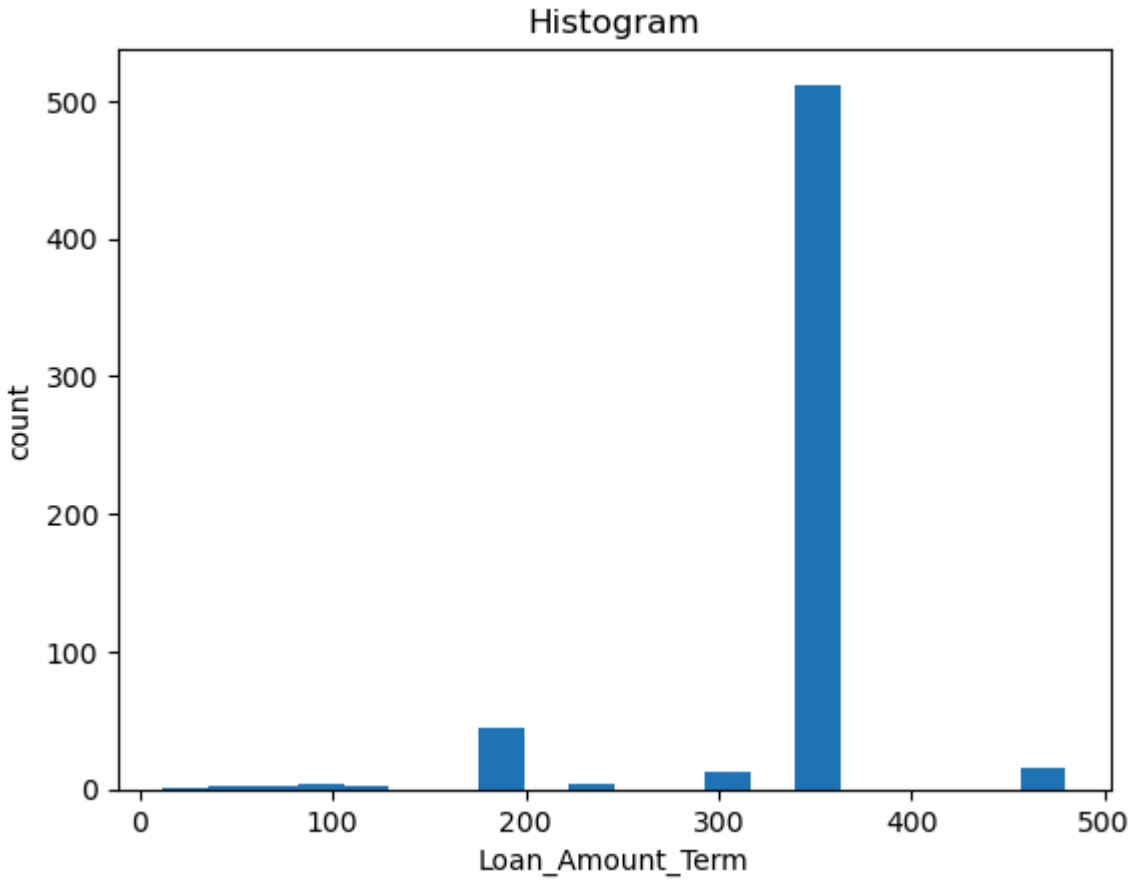


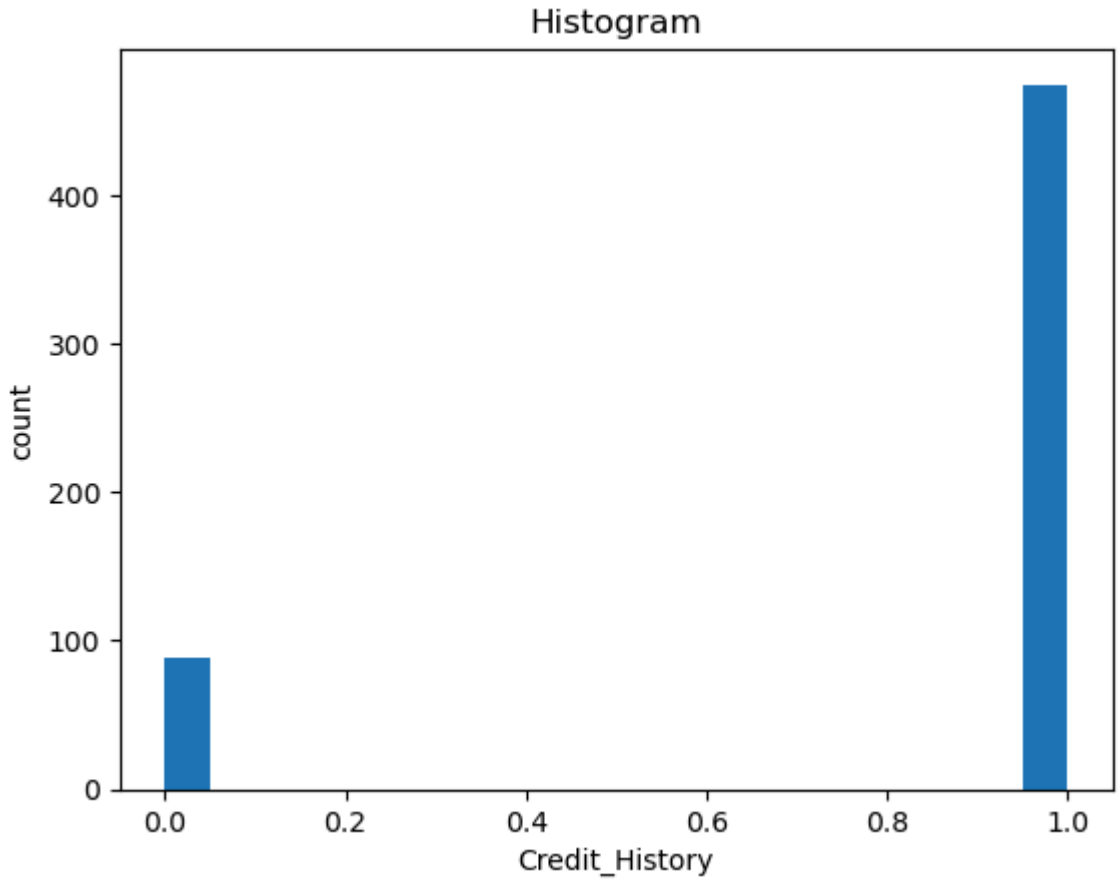
In [17]:
```
for i in num_cols:
    plt.hist(df[i],bins=20)
    plt.title('Histogram')
    plt.xlabel(i)
    plt.ylabel('count')
    plt.show()
```


Histogram
count
ApplicantIncome
0
50
100
150
200
250
300
350
0
10000
20000
30000
40000
50000
60000
70000
80000


Histogram
count
CoapplicantIncome
0
100
200
300
400
0
10000
20000
30000
40000


Histogram
count
LoanAmount
0
20
40
60
80
100
120
140
160
0
100
200
300
400
500
600
700


Histogram
count
Loan_Amount_Term
0
100
200
300
400
500
0
100
200
300
400
500


Histogram
400
300
count
200
100
0
0.0
0.2
0.4
0.6
0.8
1.0
Credit_History

In [ ]: