Multivariate Analysis

The dataset is already download in .csv format

IMPORTING THE PACKAGE

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

import warnings
warnings.filterwarnings('ignore')
```

Load the dataset

In [2]:
```
df=pd.read_csv("C:\loan_prediction.csv")
```

In [3]:
```
df.head()
```

Out[3]:

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	Coap
0	LP001002	Male	No	0	Graduate	No	5849	
1	LP001003	Male	Yes	1	Graduate	No	4583	
2	LP001005	Male	Yes	0	Graduate	Yes	3000	
3	LP001006	Male	Yes	0	Not Graduate	No	2583	
4	LP001008	Male	No	0	Graduate	No	6000	

In [4]:
```
numerical_features = df.select_dtypes(include = [np.number]).columns
categorical_features = df.select_dtypes(include = [np.object]).columns
```

In [5]:
```
numerical_features
```

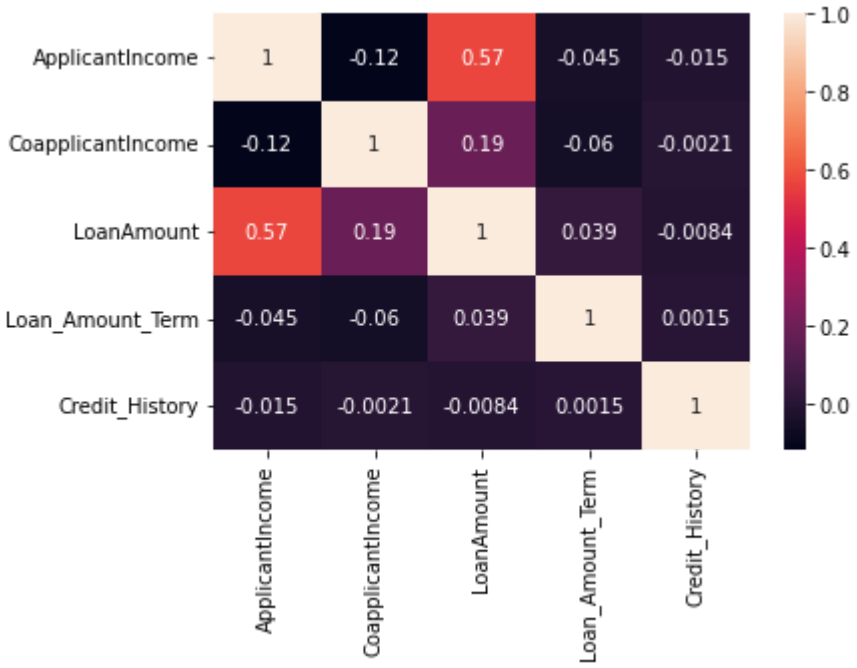
Out[5]:
```
Index(['ApplicantIncome', 'CoapplicantIncome', 'LoanAmount',
       'Loan_Amount_Term', 'Credit_History'],
      dtype='object')
```

In [6]:
```
categorical_features
```

Out[6]:
```
Index(['Loan_ID', 'Gender', 'Married', 'Dependents', 'Education',
       'Self_Employed', 'Property_Area', 'Loan_Status'],
      dtype='object')
```

In [7]:
```
df.corr()
#plotting the correlation
plt.figure(1)
sns.heatmap(df.corr(), annot = True)
```

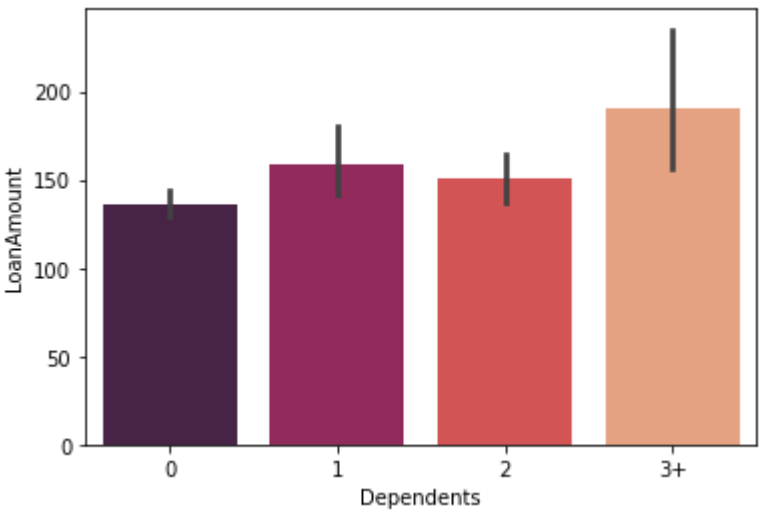
Out[7]:



Perform Visualizations

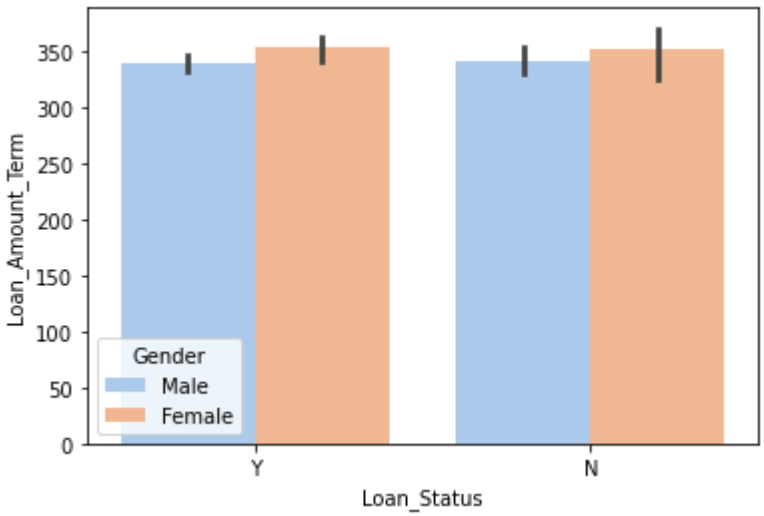
In [8]:
```
#Visualizations Dependents vs LoanAmount vs Gender
sns.barplot(df.Dependents,df.LoanAmount, hue_order=df.Gender ,palette="rocket")
```

Out[8]:



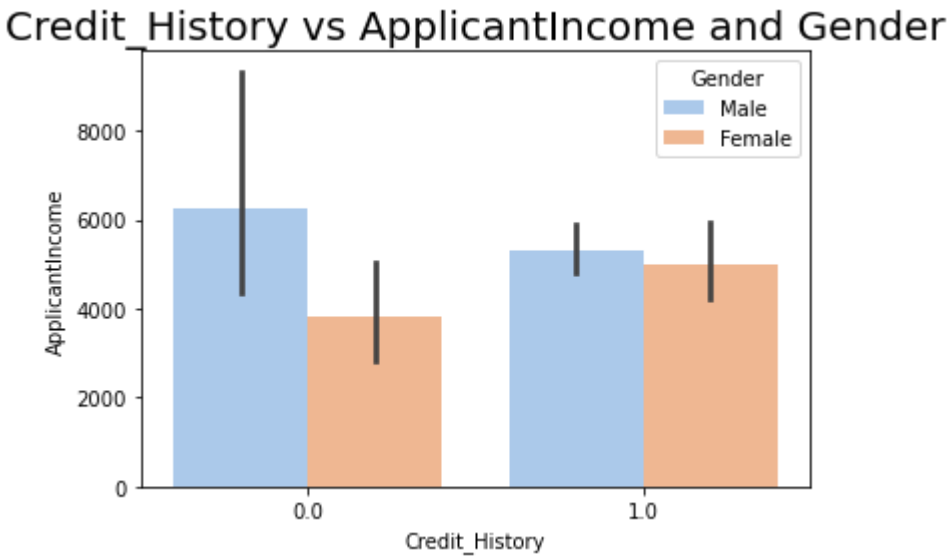
In [9]:
```
#Visualizations Loan_Status vs Loan_Amount_Term vs Gender
sns.barplot(df.Loan_Status,df.Loan_Amount_Term,hue=df.Gender, palette="pastel")
```

Out[9]:



In [10]:
```
# Credit_History vs ApplicantIncome and Gender
sns.barplot(df['Credit_History'], df['ApplicantIncome'], hue = df['Gender'], pa
plt.title('Credit_History vs ApplicantIncome and Gender', fontsize = 20)
```

Out[10]: Text(0.5, 1.0, 'Credit_History vs ApplicantIncome and Gender')



3-D Plotting

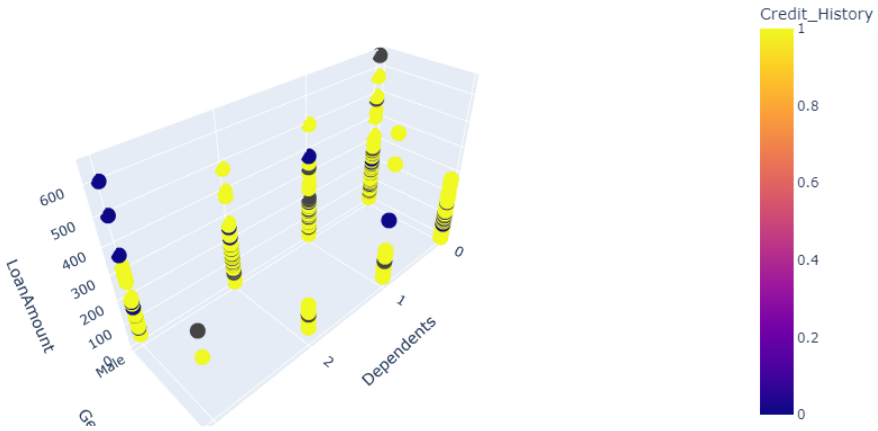
In [11]:
```
plt.figure(figsize=(10,10))

import plotly.express as px

fig = px.scatter_3d(df, x='Dependents', y='Gender', z='LoanAmount',#hue='Gender

color='Credit_History')
```

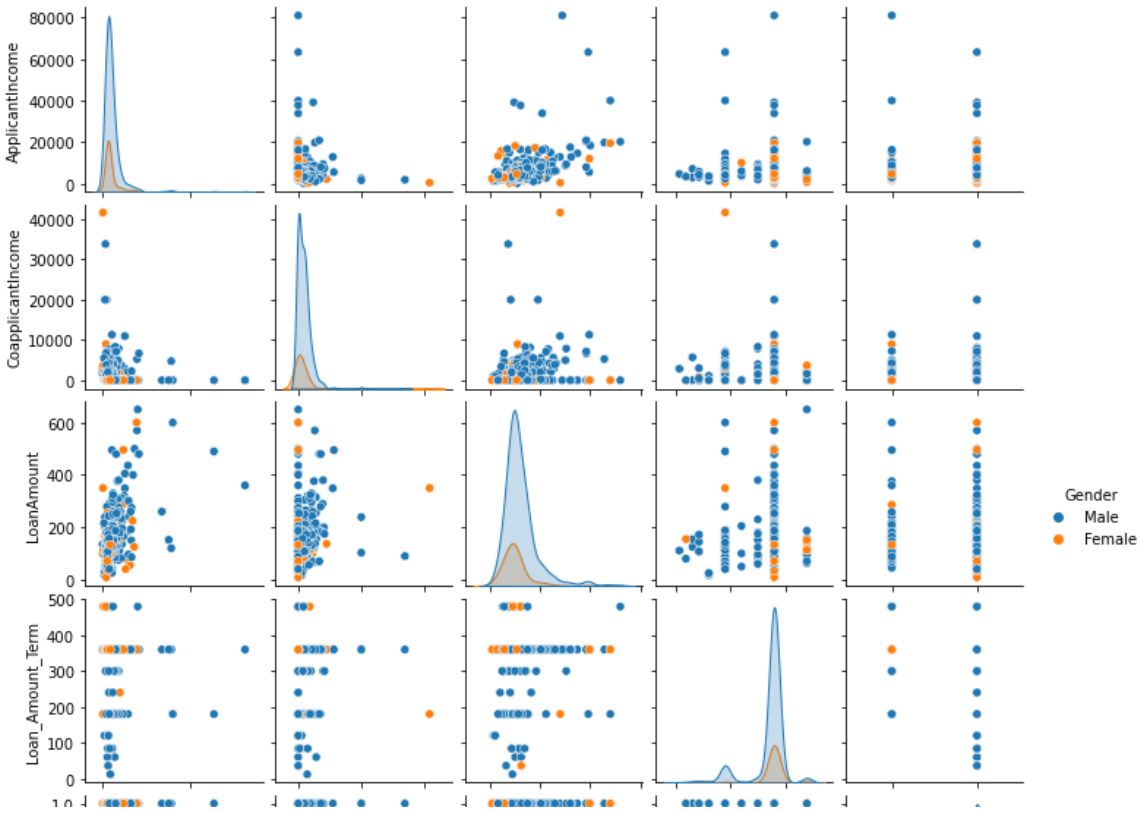
```
fig.show()
```

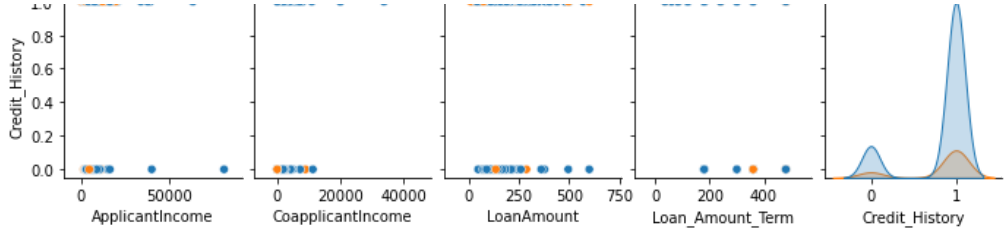


Pairplot Analysis

In [12]:
```
#Pairplot Analysis for Gender
sns.pairplot(df,hue='Gender',size=2)
```

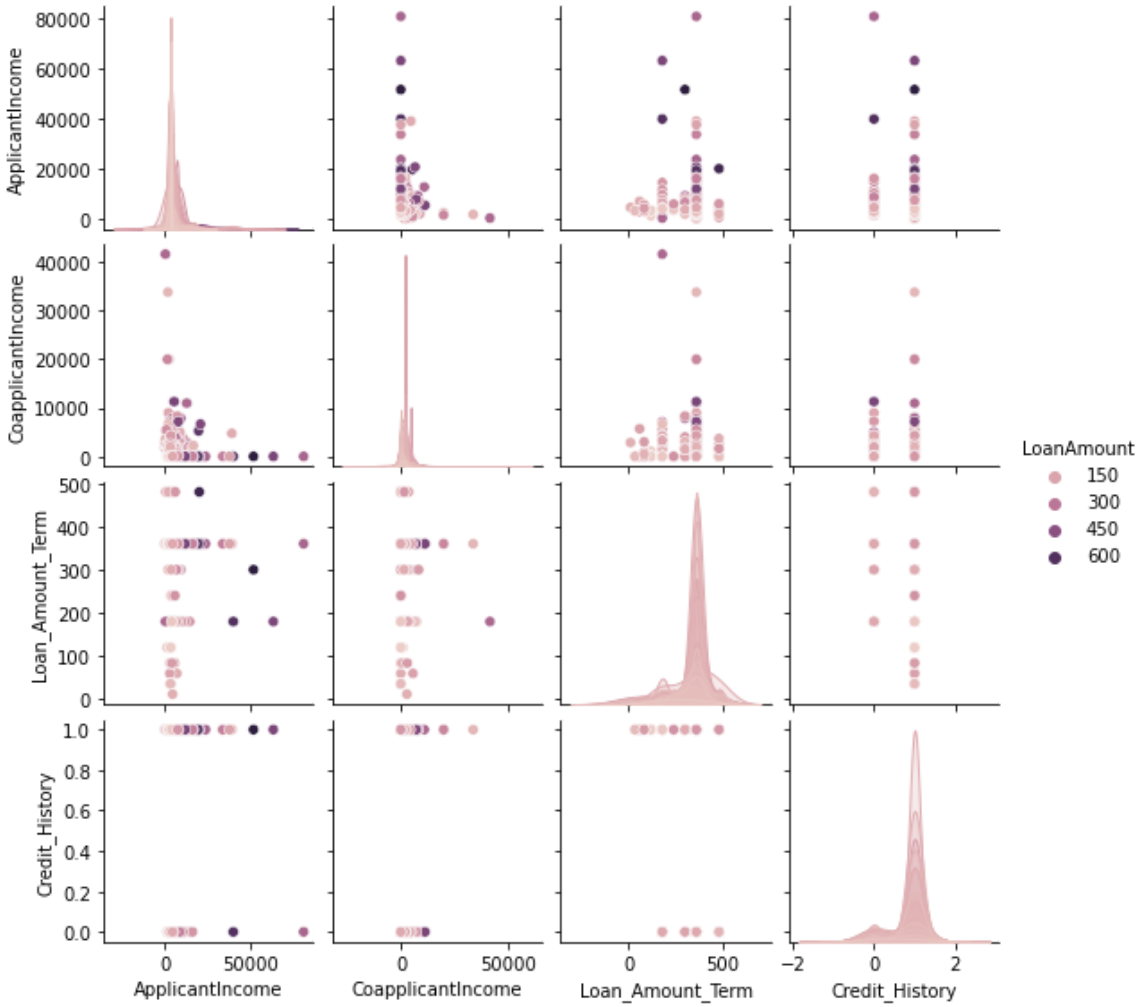
Out[12]:





In [13]:
```
#Pairplot Analysis for LoanAmount
sns.pairplot(df,hue='LoanAmount',size=2)
```

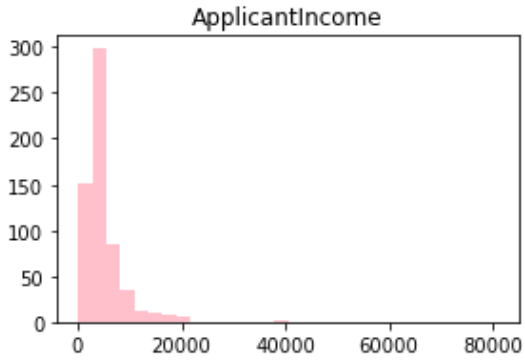
Out[13]:

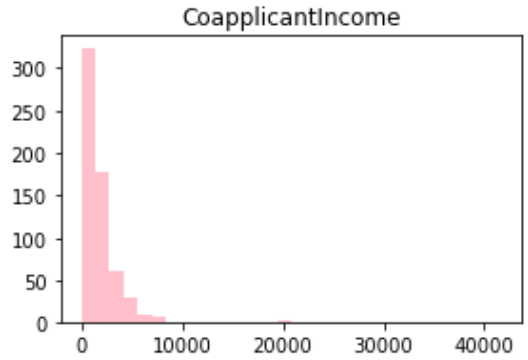


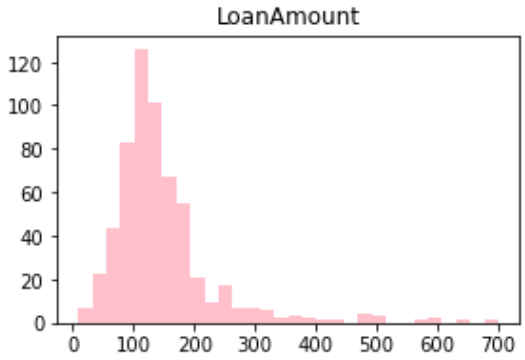
Histogram visualisation

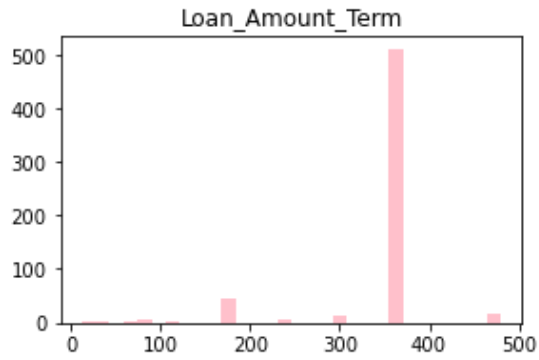
In [14]:
```
# Histogram visualisation for each attribute to know what kind of distribution
df.hist(figsize=(10,10), grid=False, layout=(3,2), bins = 30 , color = "pink" )
```

Out[14]:
```
array([[,
        ],
       [,
        ],
       [, ]],
      dtype=object)
```


ApplicantIncome
300
250
200
150
100
50
0
0
20000
40000
60000
80000


CoapplicantIncome
300
250
200
150
100
50
0
0
10000
20000
30000
40000


LoanAmount
120
100
80
60
40
20
0
0
100
200
300
400
500
600
700


Loan_Amount_Term
500
400
300
200
100
0
0
100
200
300
400
500

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

In [ ]: